NET PAY EVALUATION: A COMPARISON OF METHODS TO ESTIMATE NET PAY AND NET-TO-GROSS RATIO USING SURROGATE VARIABLES

A Thesis

by

NICOLAS BOUFFIN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

August 2007

Major Subject: Petroleum Engineering

NET PAY EVALUATION: A COMPARISON OF METHODS TO ESTIMATE NET PAY AND NET-TO-GROSS RATIO USING SURROGATE VARIABLES

A Thesis

by

NICOLAS BOUFFIN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee, Jerry L. Jensen
Committee Members, Thomas A. Blasingame
Richard L. Gibson
Head of Department, Stephen A. Holditch

August 2007

Major Subject: Petroleum Engineering

ABSTRACT

Net Pay Evaluation: A Comparison of Methods to Estimate Net Pay and Net-to-Gross Ratio Using Surrogate Variables. (August 2007)

Nicolas Bouffin, M.Eng., Ecole Nationale Supérieure de Géologie, Nancy (France)

Chair of Advisory Committee: Dr. Jerry L. Jensen

Net pay (NP) and net-to-gross ratio (NGR) are often crucial quantities to characterize a reservoir and assess the amount of hydrocarbons in place. Numerous methods in the industry have been developed to evaluate NP and NGR, depending on the intended purposes. These methods usually involve the use of cut-off values of one or more surrogate variables to discriminate non-reservoir from reservoir rocks.

This study investigates statistical issues related to the selection of such cut-off values by considering the specific case of using porosity (ϕ) as the surrogate. Four methods are applied to permeability-porosity datasets to estimate porosity cut-off values. All the methods assume that a permeability cut-off value has been previously determined and each method is based on minimizing the prediction error when particular assumptions are satisfied.

The results show that delineating NP and evaluating NGR require different porosity cut-off values. In the case where porosity and the logarithm of permeability are joint normally distributed, NP delineation requires the use of the Y-on-X regression line to estimate the optimal porosity cut-off while the reduced major axis (RMA) line provides the optimal porosity cut-off value to evaluate NGR.

Alternatives to RMA and regression lines are also investigated, such as discriminant analysis and a data-oriented method using a probabilistic analysis of the porosity-permeability crossplots. Joint normal datasets are generated to test the ability of the methods to predict accurately the optimal porosity cut-off value for sampled sub datasets. These different methods have been compared to one another on the basis of the bias, standard error and robustness of the estimates.

A set of field data has been used from the Travis Peak formation to test the performance of the methods. The conclusions of the study have been confirmed when applied to field data: as long as the initial assumptions concerning the distribution of data are verified, it is recommended to use the Y-on-X regression line to delineate NP while either the RMA line or discriminant analysis should be used for evaluating NGR. In the case where the assumptions on data distribution are not verified, the quadrant method should be used.

ACKNOWLEDGMENTS

First, I would like to express my greatest respect and appreciation to my advisor, Dr. Jerry L. Jensen, for supervising this thesis project during all its development, for his excellent guidance, and for his dependability whenever help was needed.

I also wish to thank Drs. Thomas A. Blasingame and Richard L. Gibson for serving on my committee.

I am very thankful to the Tuck Foundation and the Texas A&M Petroleum Engineering Department for providing the funding for my studies here at Texas A&M University.

Finally, my sincere thanks go to my family, to the IFP School Aggies Mauro Marzi, Stephanie Lafage, Nicolas Decaillet and Coralie Genty for their constant support, and who have rejoiced with me in the good times and encouraged me to move forward through the bad.

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

Page

Page

Page

Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

BACKGROUND

Net pay (NP) may be defined as “any interval that contains producible hydrocarbon at economic rates given a specific production method”. It thus represents the portion of the reservoir that contains high storability (driven by porosity), high transmissivity (driven by the fluid mobility, which refers to as the ratio of permeability to fluid viscosity), and a significant hydrocarbon saturation (driven by water saturation, Sw). Net pay can be interpreted as an effective thickness that is pertinent to identification of flow units and target intervals for well completions and stimulation programs (Worthington and Cosentino, 2005). The associated net-to-gross ratio (NGR) corresponds to the ratio of the net pay thickness to the total (or gross) thickness of the reservoir under consideration.

Net pay and NGR are needed for several reservoir characterization activities. A major use of net pay is to compute volumetric hydrocarbons in-place. Another use of net pay is to determine the total energy of the reservoir i.e. both moveable and non-moveable hydrocarbons are taken into consideration. Net pay for this purpose may be therefore much greater than that for volumetrics calculation (George and Stiles, 1978). A third use of net pay is to evaluate the potential amount of hydrocarbon available for secondary recovery, meaning net pay with favorable relative permeability to the injected fluid, i.e. “floodable net pay” (Cobb and Marek, 1998). Net pay and NGR are crucial to quantify the hydrocarbon reserves and have a significant impact on the economic viability of hydrocarbon reservoir production (Worthington and Cosentino, 2005).

Net pay determination usually involves defining the threshold values (or cut-offs) of the characteristics of interest. These limiting values are designed to define those rock

---

This thesis follows the style and format of *Petrophysics*.

volumes that are not likely to contribute significantly to the hydrocarbon production. For example, regions with water saturations $S_w > 0.5$ may be considered non-pay. The establishment of these cut-off values i.e. 0.5, will vary according to the intended application and should be therefore fit for purpose, meaning that "the intended use of the net pay often determines how net pay is picked" (Snyder, 1971). Since the method to pick net pay (and to a larger extent NGR) depends on its usage, these uses determine also the method chosen for establishing cut-off values.

In conclusion, there is no systematic method for identifying cut-off variables and their values. The choice of variable and cut-off value depends strongly on the intended application of evaluated net pay and NGR.

SUMMARY OF THE PROBLEM

The permeability cut-off is very often considered to be the controlling parameter in net pay and NGR evaluation especially in cases involving the flow regime or the reservoir recovery mechanism. The permeability cut-off, k_c, is dependent on a limited number of parameters including the fluid mobility, the permeability distribution, the reservoir pressure differential, and the reservoir drive mechanism (primary or waterflood). Its range typically varies between 0.1 and 100 md depending mainly on the fluid mobility. Because of its low viscosity, gas mobility might remain significant in a tight reservoir so the reservoir is still producible: the mobility is therefore an "appropriate starting point" to determine net pay from permeability cut-off (Cobb and Marek, 1998).

Nonetheless there is no subsurface continuous permeability measurement, k, ("permeability log") and core permeability measurements are not available throughout all wells. As a consequence, surrogate variables usually derived from well log measurements, such as porosity (ϕ), amount of shale (V_{sh}) and water saturation (S_w), are generally used to infer the locations and amount of net pay. The selection of cut-off values for these surrogate variables needs to be carefully done in order to avoid introducing further errors into the net pay identification process. It is then necessary for this purpose to test the accuracy and robustness of the available methods providing cut-offs and determine the optimal ones when evaluating either net pay or NGR. In the case where it is already determined based on the mentioned engineering and geological considerations, the permeability cut-off k_c should be therefore related to those surrogate variables.

A common method to identify net pay using porosity (to a larger extent any surrogate variable such as water saturation S_w, shaliness V_{sh} or formation resistivity R_t) is to use semi logarithmic porosity vs. permeability crossplots and a least-squares regression line to obtain the porosity cut-off (Worthington and Cosentino, 2005). A porosity cut-off ϕ_c may be obtained from the regression line (Figure 1-1).

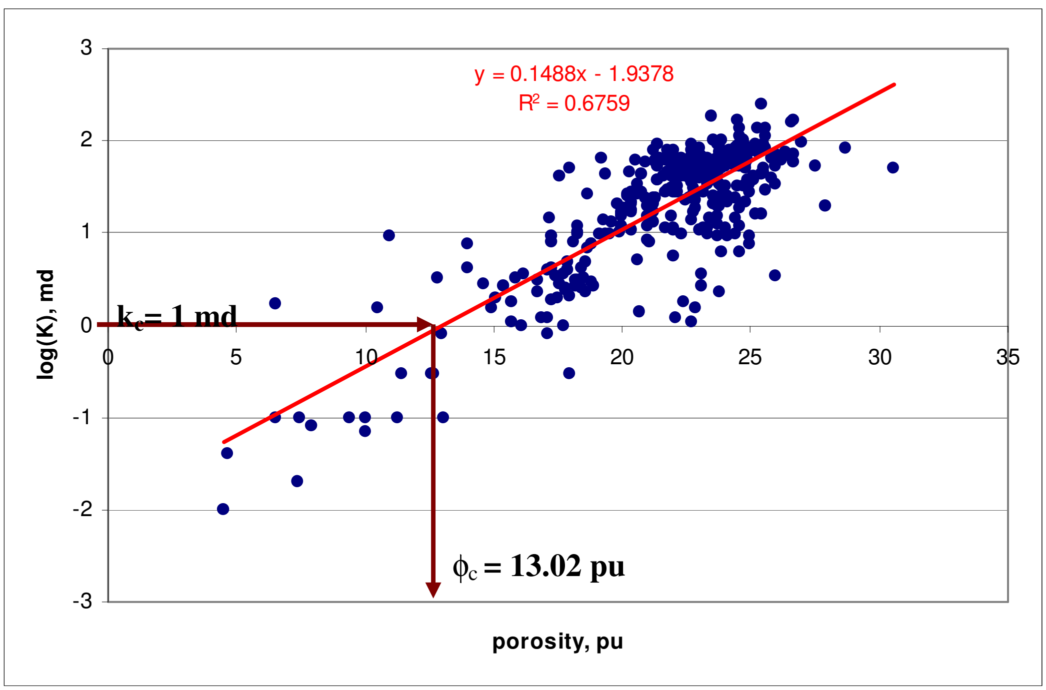


Fig. 1-1 Use of the Y-on-X line determined by least squares regression with porosity and log-permeability values. The porosity cut-off ϕ_c is obtained from a permeability cut-off of 1 md (data from Dutton et al., 2003; Nelson and Kibler, 2003).

The use of the Y-on-X regression line is an example of methods which may provide porosity cut-off values. These methods provide estimates of the "best" cut-off value with associated statistical characteristics. The best value is the value which, when used, gives the smallest likelihood error of prediction. This study will investigate which of these several porosity cut-off methods gives cut-off values which are optimal in term of bias, efficiency, and robustness when applied to evaluate net pay and NGR.

OBJECTIVES OF THE STUDY

The main objective is to evaluate net pay and net-to-gross estimators of the porosity cut-off value for their bias, efficiency, and robustness. The study has three component objectives.

The first component will be to provide an analytical justification of the observations made by Jensen and Menke (2006). Assuming that the log permeability and porosity are joint normally distributed, they found that the Y-on-X regression line is required to obtain optimal cut-off values for identifying net pay whereas the optimal porosity cut-off for evaluating NGR is obtained from the reduced major axis (RMA) line. A derivative analysis will be conducted using the probability density function of a joint normal bivariate population.

The second component compares the performance of estimators for predicting cut-off values for net pay and net-to-gross using the Monte Carlo method. Log (k)-ϕ datasets that are joint normal distributed will be generated and the different porosity cut-off estimators will be compared on these datasets. In order to study the optimal porosity cut-off for evaluating net pay, several methods will be investigated such as Y-on-X regression and RMA lines, discriminant analysis and purely data-oriented methods.

The second component results will include an assessment of the estimators when noise is present. These results will lead to recommendations as to which method is optimal depending on the purpose and the statistical properties of the studied data.

CHAPTER II

PREVIOUS WORK

There is no generally accepted protocol to delineate NP and NGR on the basis of cut-off values: it can be done by using surrogate variables tied back to a permeability cut-off value, by using capillary pressure, and/or by analyzing the respective distributions of pay and non-pay fractions of the dataset, i.e. discriminant analysis. It results in numerous studies and recommendations in the literature.

USE OF SURROGATES TO PREDICT NP AND NGR

A common approach is to define fixed permeability cut-off values according to the "Rule of Thumb": gas-bearing rocks for which $k \geq 0.1md$ are admitted as net pay whereas oil-bearing rocks for which $k \geq 1md$ are pay. This approach is arbitrary since the rule of thumb is not taking into consideration the reservoir fluid characteristics. For instance a 1.0 md permeability cut-off may be appropriate in the case of medium-gravity oils whereas a 0.1 md permeability cut-off is adequate for light, low-viscosity oils (George and Stiles, 1978).

Since there is no continuous measurement of permeability, the practice has been therefore to relate core permeability to porosity and/or other log-derivable measurements such as V_{sh}, S_w, and R_t. The cut-off values should be "dynamically conditioned", i.e. they should be tied back to a hydraulic parameter, such as absolute permeability, pore throat radius or fluid mobility (Worthington and Cosentino, 2005).

Pirson (1958) developed a "coregraph" method using three independent cut-offs for k, ϕ and S_w. Another method from core and log analysis takes account of a different set of three net-pay cut-offs, shale factor V_{sh}, ϕ and S_w (Keener et al, 1967). McKenzie (1975) also defined "producible and non-producible rock types" by establishing an effective pore throat size correlated with the ratio $\frac{k}{\phi}$. A porosity cut-off ϕ_c below which there is no commercial permeability can be also considered. ϕ, S_w and a bulk-volume water ($\phi.S_w$) cut-off values are used for evaluating NP and NGR of oil-bearing carbonates of the Willinston Basin (Teti and Krug, 1987). The main advantage of these methods is that log-derived measurements are used instead of core data allowing to directly delineating NP on well log data without requiring further laboratory measurement. In the early stage of the discovery and appraisal of a field, these methods may give a reasonable evaluation of NP and NGR of a potential reservoir. However, the establishment of the cut-off values depends greatly on the way the surrogate are tied back to the permeability cut-off value, which might create additional errors.

USE OF CAPILLARY PRESSURE

Numerous models have been developed to predict permeability and delineate NP on the basis of the capillary pressure curves. Those curves are considered as direct indicator of permeability since capillary pressure curves are functions of the pore throat geometry and radius, grain sorting and to a smaller extent fluid properties (Vavra et al., 1992).

The Winland method (Kolodzie, 1980) intends to correlate porosity and permeability to pore throat radius (r) corresponding to different mercury saturations (Spearing et al., 2001). Pore throat sizes are derived from the Washburn equation, expressed as follows,

$$P_c = \frac{2\gamma \cos(\sigma)}{r}, \tag{2-1}$$

where P_c is the mercury/air capillary pressure, γ is the mercury/air interfacial tension, σ is the mercury/pore wall contact angle and r is the pore throat radius.

The percentage of non-wetting fluid saturation, i.e. mercury, giving the best correlation between ϕ, k and r, is assumed to correspond to the modal class of pore throat radius when the pore network becomes interconnected. Winland found that the 35th percentile, corresponding to 35 percent of the pore volume (“R35”, at which he observed an inflexion on the mercury injection capillary curve vs. mercury saturation), gave the best correlation for the Spindle Field data: it corresponds to a 0.5 μm pore throat threshold value. In order to delineate NP, a R35 throat radius vs permeability crossplot analysis is realized and the permeability cut-off is read off from a best fit line using the 0.5 μm threshold.

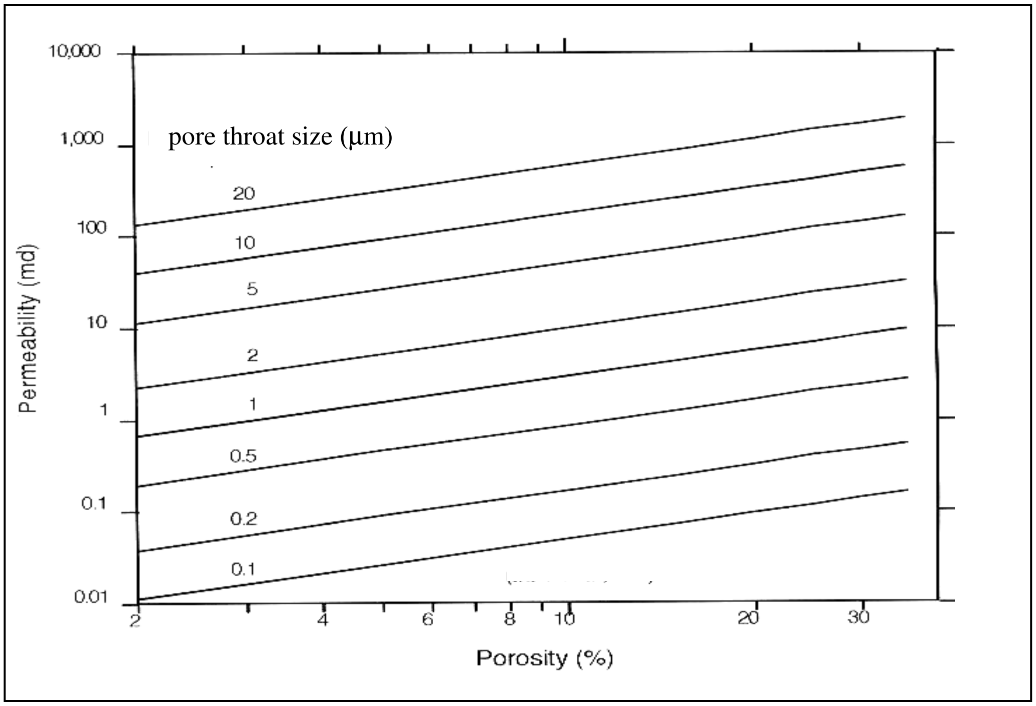


Fig. 2-1 Winland model based on pore geometry (Lucia, 1999).

The other way of defining porosity and permeability cut-offs is to display permeability vs. porosity cross plot with the isopore-throat radius lines (Figure 2-1). The permeability and porosity cut-off values may be calculated based on a pore throat radius value.

INFLUENCE OF THE RESERVOIR CHARACTERISTICS IN THE NP AND NGR EVALUATION PROCESS

In order to evaluate NGR, Egbele and Ezuka (2005) suggested associating each petrofacies (i.e. a rock type with defined petrophysical characteristics), with one unique pair of Log (k)-ϕ cut-off values instead of tying the traditional k_c to porosity values. This approach is based on the argument that pore-throat geometries are very dissimilar from one petrofacies to another one. This complies with the recommendations from Morton-Thompson and Woods (1993) who insist upon “a systematic, sedimentologically based reservoir zonation” as an essential component of effective pay determination.

Lucia (1999) demonstrated that by plotting interparticle porosity against permeability in carbonate reservoirs, one could derive the type of rock fabric and detect pore-size classes. Additional pore types (vuggy, dissolution-enhanced) might modify these relationships. The permeability and porosity cut-off values should be defined based on these considerations. A unique permeability cut-off value based on engineering considerations (i.e. mainly depending on the fluid mobility) will lead to several porosity cut-off values depending on the rock fabric, i.e. the particle size (Figure 2-2).

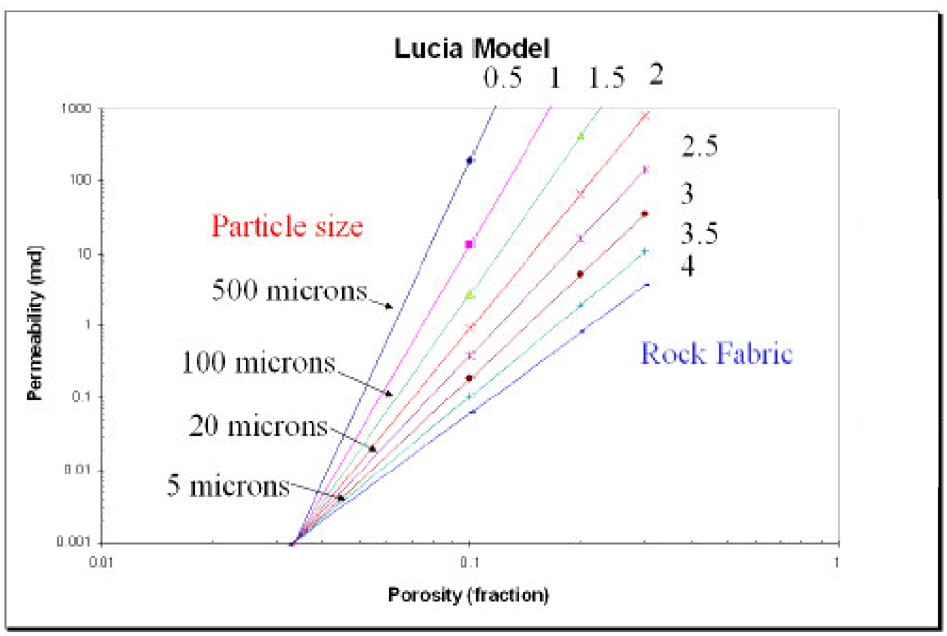


Fig. 2-2 Lucia model for porosity-permeability relationships based on rock fabric (Haro, 2004; modified from Lucia, 1999).

STATISTICAL ANALYSIS OF THE NP AND NGR DETERMINATION USING SURROGATE VARIABLES

In the case where either determination of reservoir NGR and/or NP is obtained by cross-plotting surrogate quantities as S_w, V_{sh}, and/or ϕ, investigating the errors inherent to the regression methods giving log (k) vs. ϕ best fit lines is crucial since the misuse of regression methods may lead to additional errors. Such statistical issues related to the selection of porosity cut-offs based on regression lines were investigated by Jensen and Menke (2006). Their study investigated the use of semilog porosity vs. permeability plot and the Y-on-X regression line to derive porosity cut-off values.

Jensen and Menke (2006) used a probabilistic approach to analyze the accuracy and errors in prediction of various porosity cut-off values. Their approach is based on defining four regions A, B, C, and D in the log(k)-ϕ (Figure 2-3), where the region boundaries are defined by the threshold values k_c and ϕ_c. Region A ($\phi < \phi_c$ and $k < k_c$) corresponds to the non-pay fraction of the data correctly identified using the porosity cut-off value and region D represents the pay intervals ($\phi > \phi_c$ and $k > k_c$) also correctly identified using the porosity cut-off value. Regions B $\left(\phi < \phi_c \text{ and } k \geq k_c\right)$ and C $\left(\phi \geq \phi_c \text{ and } k < k_c\right)$ represent the respectively erroneous misidentification of non-pay for pay and of pay for non-pay. The probability that an event, for instance A, occurs is defined as prob(A) and may be calculated as the ratio of the number of points, i.e. pairs of (k-ϕ), that are included in the area A, to the total of points displayed in the k-ϕ crossplot. The probabilities of events B, C, or D, are thus respectively defined as prob(B), prob(C), and prob(D).

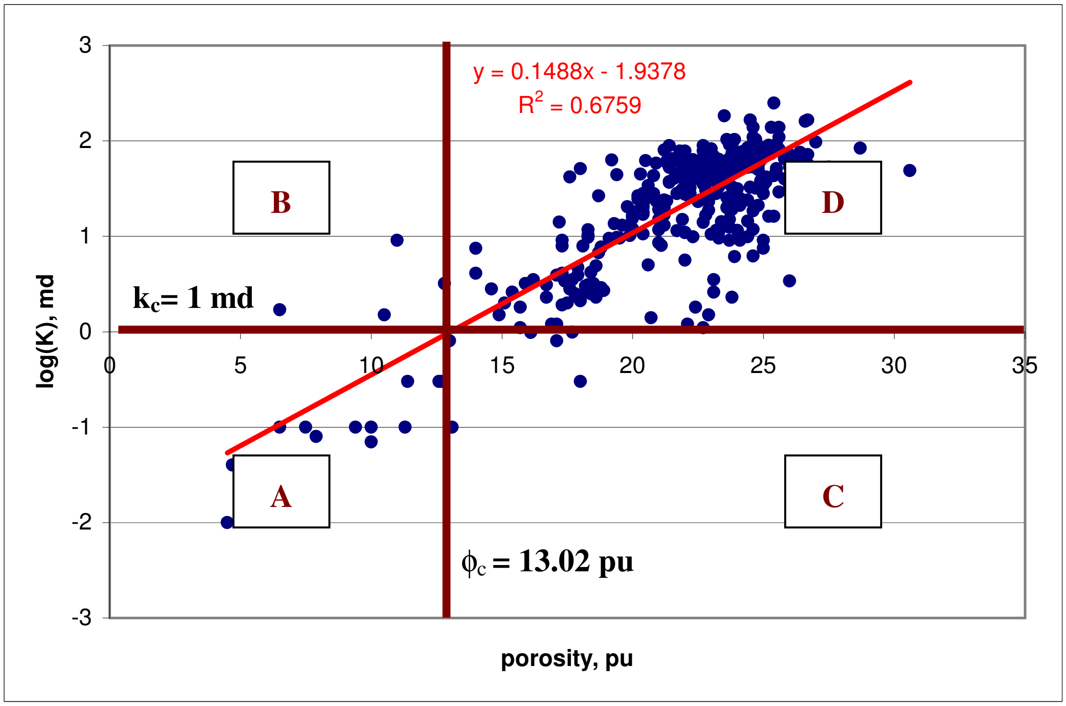


Fig. 2-3 Permeability vs porosity cross plot divided into four distinct regions, A, B, C, and D based on cut-offs values k_c and ϕ_c (data from Dutton et al., 2003; Nelson and Kibler, 2003).

Depending on whether NP or NGR is to be estimated, two separate criteria emerge for the best value of ϕ_c. One criterion is to minimize the sum of the probabilities prob(B) and prob(C) in order to minimize the errors of mistaking pay for non pay and non-pay for pay and thus delineate net pay intervals. The alternative criterion is to equalize the probabilities prob(B) and prob(C) in order to cancel out the misidentified parts of the reservoir for NGR evaluation.

The systematic use of the Y-on-X regression line to predict porosity cut-off values might induce errors and happen to significantly differ from the optimal cut-off values for delineating net pay and evaluating NGR. In the case that log-permeabilities and porosities are assumed to be JND, Jensen and Menke (2006) observed that the regression line provides the optimal results for estimating net pay whereas the RMA line gives optimal results for NGR.

USE OF DISCRIMINANT ANALYSIS

An alternative to using regression lines is to separate reservoir rocks from non-reservoir rocks based on their statistical properties and probability distribution functions (Kraznowski, 1988; Li and Dria, 1997; Jensen and Menke, 2006). In the case where the distributions of reservoir and non-reservoir rocks do not have distinctly separate ranges (Figure 2-4), establishing the boundary segregating the two distinct distributions, i.e., a discriminant function, could be more efficient and less erroneous than using a cut-off value.

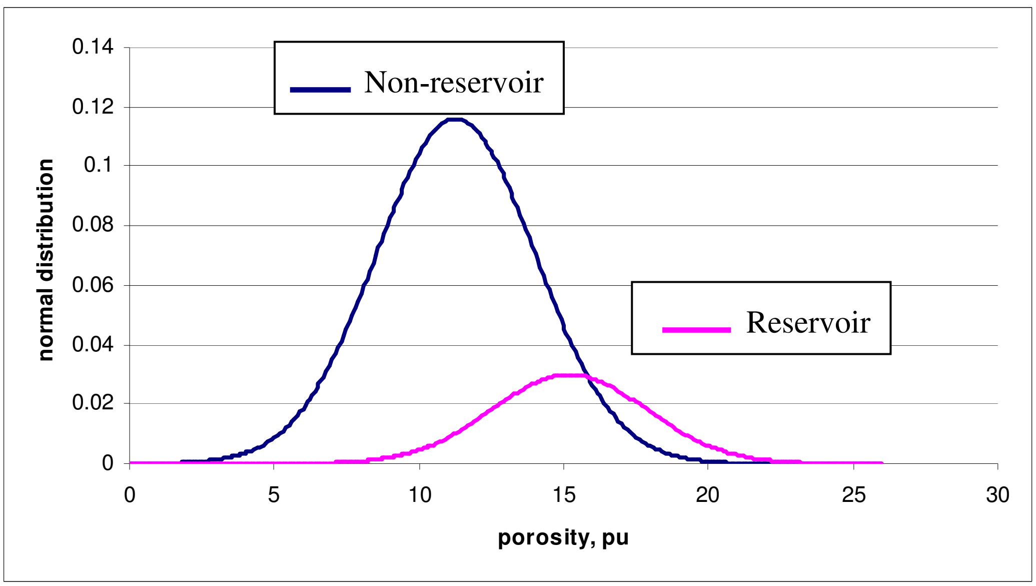


Fig. 2-4 Histogram of distributions of non-reservoir and reservoir rocks which do not have distinct ranges.

CHAPTER III

DESCRIPTION OF STATISTICAL METHODS

Y-ON-X REGRESSION AND REDUCED MAJOR AXIS LINES

General considerations concerning linear regression

The general problem of linear regression is to develop a predictor of a quantity Y (e.g. log permeability) from knowledge of the value of a variable X (e.g. porosity). The variable being investigated is the dependent or regressed variable, designated Y; individual observations of the dependent variable are indicated as y_i. The other variable is the predictor or regressor variable and is denoted X, with individual observations, x_i. The fitted line will cross the Y-axis at a point b_0 (the intercept), and will have a slope b_1. The expected relationship between Y and X is linear.
The regression line equation is as follows:

$$\hat{y}_i = b_0 + b_1.x_i\,, \tag{3-1}$$

where $\hat{y}_i$ is the estimated value of y_i for any value x_i.

Considering that only the variable Y is assumed to be measured with error gives specific coefficients b's referring to the Y-on-X line. In contrary, in the case that only the variable X is assumed to be with errors, it gives distinct coefficients b's that correspond to the X-on-Y line.

Y-on-X regression line

The b's (Equation 3-1) are usually determined by the least-squares regression and consists in minimizing the sum of the squared differences between the observed variable, y_i , and the predicted responses as expressed by equation (3-2).

$$\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2 = \min\,, \tag{3-2}$$

where n is the number of points.

The justification of the technique is given by Jensen et al. (2003, p184-186) using differential calculus. The coefficients b_0 and b_1 are defined as follows (Davis, 2002, p. 194-195):

$$b_1 = \frac{\sum_{i=1}^{n} x_i . y_i - \frac{\left(\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}, \tag{3-3}$$

$$b_0 = \frac{\sum_{i=1}^{n} y_i}{n} - b_1 \frac{\sum_{i=1}^{n} x_i}{n} = \overline{Y} - b_1 \overline{X} . \tag{3-4}$$

Reduced major axis line

Another line is one where both variables X and Y are assumed to have errors. Estimation of b_0 and b_1 minimizes the sum of the areas of the triangles formed by the observations and the fitted line, $(y_i - \hat{y}_i).(x_i - \hat{x}_i)$, i.e. the product of the deviations in both the X- and Y- directions is minimized. It results in what is called the reduced major axis, or more commonly referred as the "RMA line". The RMA line is in fact more appropriate than standard regression lines when the independent variable X is measured with significant error. In this case, estimates of slope will be biased.

The reduced major axis can also be expressed as an ordinary linear equation, such as equation (3-1).

The coefficients are estimated as follows (Davis, 2002, p. 216-217):

$$b_1 = \frac{\sigma_Y}{\sigma_X} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}}, \qquad (3\text{-}5)$$

$$b_0 = \overline{Y} - b_1\overline{X} \qquad (3\text{-}6)$$

The joint normal distribution and its related lines

The joint normal distribution (JND) is defined by the following probability density function (PDF) (Jensen et al., 2003, p. 172).

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left[-\frac{x^{*2} - 2\rho x^* y^* + y^{*2}}{2\left(1-\rho^2\right)}\right] \qquad (3\text{-}7)$$

where $x^* = \dfrac{x-\mu_X}{\sigma_X}$, $y^* = \dfrac{y-\mu_Y}{\sigma_Y}$, and $\rho = \dfrac{\sigma_{XY}}{\sigma_X\sigma_Y}$.

The marginal distributions of the variables X and Y, respectively ϕ and log (k), are normal. An example of a joint normal distribution is shown in Figure 3-1.

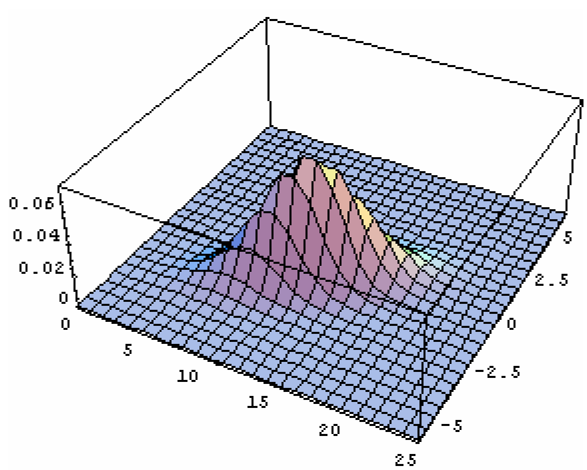


Fig. 3-1 A joint normal PDF with the following statistical quantities σ_X=3, σ_Y=1, μ_X=12, μ_Y=-1, and ρ=0.7.

JND ellipses may be formed by taking $f(x, y) = cte$ in the case where ρ is different from 0: the plane formed by setting the PDF to a constant intersects the PDF and forms an ellipse. If ρ is equal to 0, the plane formed by $f(x, y) = cte$ represents a circle. If a joint normal distributed dataset is generated from the PDF presented in figure 3-2, one may represent the RMA line and the associated regression lines in a standardized (X*, Y*) plane where $X^* = \frac{X - \mu_X}{\sigma_X}$ and $Y^* = \frac{Y - \mu_Y}{\sigma_Y}$.

The line which intersects the horizontal axis at 45°, corresponding to a slope of 1, thus represents the RMA line. On the basis of equations (3-5) and (3-6), the expression of the RMA axis may be expressed as follows:

$$Y = \mu_Y + \frac{\sigma_Y}{\sigma_X}(X - \mu_X). \qquad (3\text{-}8)$$

The associated regression lines, Y-on-X and X-on-Y lines have respectively slopes equal to ρ and $1/\rho$ (Figure 3-2). The three lines intersect at $(\overline{X}, \overline{Y})$ so that, near the center, the regression lines are not significantly different from the RMA line. This latter point, although obvious, will be discussed in more detail below because it affects whether choosing one or another of the possible lines makes any difference to the predicted porosity cut-off value.

In the (X*, Y*) plane, the RMA line, the Y-on-X, and the X-on-Y regression lines may therefore be respectively expressed as follows:

$$Y^* = 1.\, X^*, \qquad (3\text{-}9)$$

$$Y^* = \rho.\, X^*, \qquad (3\text{-}10)$$

$$Y^* = \frac{1}{\rho}.\, X^*. \qquad (3\text{-}11)$$

The closer ρ is to 0, the more different from 1 the slopes of Y-on-X and X-on-Y regression lines are: the regression lines are in this case more and more distinct with respect to the RMA line. The Y-on-X regression line is also expected to be more sensitive to noise and the degree to which the data are spread out than RMA line (Agterberg, 1974, Jensen et al., 2003).

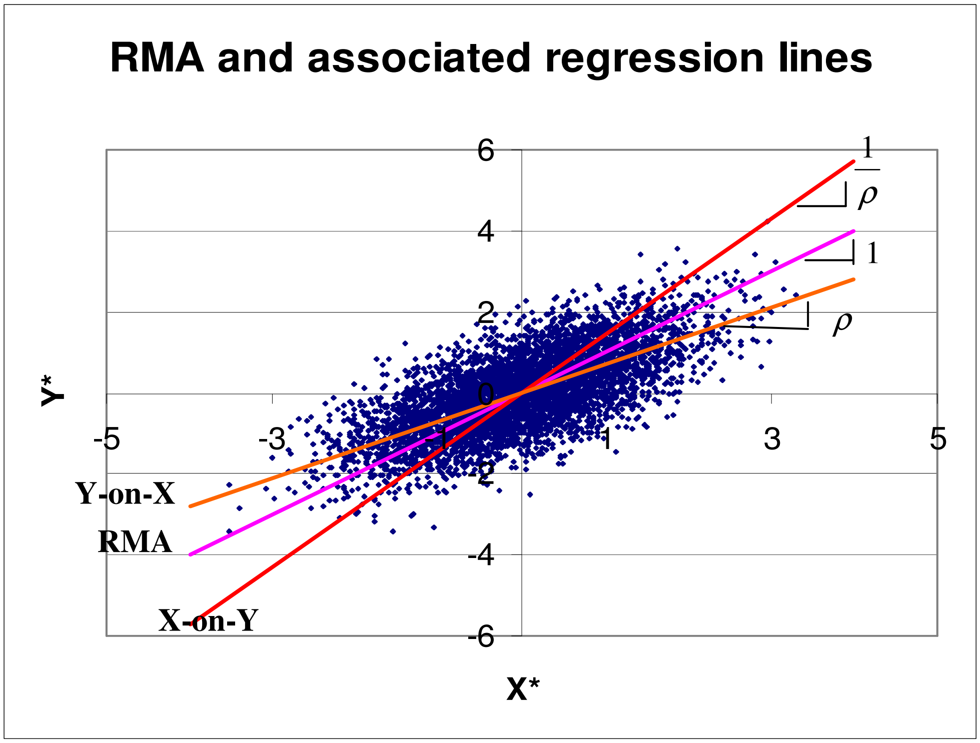


Fig. 3-2 1000 point bivariate dataset generated from the PDF shown in figure 3-1. The RMA line with a slope of 1 is displayed in pink. The associated regression lines, Y-on-X and X-on-Y, are respectively displayed in orange and red.

Considering the porosity and log permeability crossplot, the value of the permeability cut-off, k_c, and the coefficient of correlation, ρ, have a critical impact on the degree to which the lines are differentiated. In the case where k_c is not significantly different from the average of the permeability data and/or the coefficient of correlation is close to 0, the different lines are not differentiated, leading to a unique porosity cut-off value.

RELATIONSHIP OF REGRESSION AND RMA LINES TO ERROR BEHAVIOR

Jensen and Menke (2006) observed that the joint normality of Y = log (k) and X = ϕ makes the porosity cut-off value derived from the Y-on-X line optimal to delineate NP and the porosity cut-off value from the RMA line optimal to evaluate NGR. In order to evaluate NGR and equalize $prob(B)$ and $prob(C)$, i.e. to equalize the errors of misidentifying pay for non-pay and non-pay for pay, the RMA line is required. When delineating net pay, i.e. minimizing the sum of the errors in identifying pay and non-pay, the optimal porosity cut-off is given by the Y-on-X regression line. Those results are valid in case the log k-ϕ dataset is assumed to be joint normal distributed.

Detailed justifications of those observations are presented in Appendix A.

DISCRIMINANT ANALYSIS: AN ALTERNATIVE TO USING LINES TO DEFINE CUT-OFF VALUES

This technique consists of identifying a cut-off value which separates reservoir rocks from non-reservoir rocks based on their distribution functions. In the case of the present study, the permeability cut-off value has been determined so that the pay fraction of the dataset (i.e. $k \geq k_c$) may be segregated from the non-pay fraction of the data (i.e. $k \leq k_c$). The statistical quantities of porosities for these two subsets may be calculated and are defined as follows:

s_{NP}: Standard deviation of porosity for non-pay intervals

s_P: Standard deviation of porosity for pay intervals

$\overline{\phi_P}$: Average of porosity for pay intervals

$\overline{\phi_{NP}}$: Average of porosity for non-pay intervals

The procedure assumes that the porosity PDF's for both the pay and non-pay fractions are known. Here, we assume both PDF's are normal, whatever the value of k_c. In the case where the non-pay and pay porosity PDF's are not normally distributed, performance of the discriminant analysis might be altered. The normality of the two fractions of the porosities should be tested prior to using the normal-PDF approach (discussed in Appendix B). If the bivariate dataset is significantly corrupted by noise and the number of sample is sufficiently high, the data will be more spread out and it will make the non-pay and pay fractions tend to normality. Errors are expected to make the method more efficient to predict porosity cut-off values.

The PDF of the pay and non-pay porosities may be defined based on the expression of their normal distributions:

$$\Phi_{NP}(x) = \frac{1}{\sqrt{2\pi\, s_{NP}^2}}.\exp\left[-\frac{(x-\mu_{NP})^2}{2.s_{NP}^2}\right] \tag{3-12}$$

$$\Phi_{P}(x) = \frac{1}{\sqrt{2\pi\, s_{P}^2}}.\exp\left[-\frac{(x-\mu_{P})^2}{2.s_{P}^2}\right] \tag{3-13}$$

Considering that the misidentifications of pay for non-pay are equally undesirable and assuming that pay and non-pay are equally likely (Krzanowski, 1988, p. 332-348), it gives:

$$\int_{\phi_C}^{+\infty} \frac{1}{\sqrt{2\pi\, s_{NP}^2}}.\exp\left[-\frac{(x-\mu_{NP})^2}{2.s_{NP}^2}\right].dx = \int_{-\infty}^{\phi_C} \frac{1}{\sqrt{2\pi\, s_{P}^2}}.\exp\left[-\frac{(x-\mu_{P})^2}{2.s_{P}^2}\right].dx \tag{3-14}$$

where ϕ_c is the porosity cut-off value.

Introducing the estimated NGR into equation (3-14) so as to relax the assumption concerning the likelihood of pay and non-pay (Figure 3-3), it gives

$$(1-NGR).\int_{\phi_C}^{+\infty} \frac{1}{\sqrt{2\pi\, s_{NP}^2}}.\exp\left[-\frac{(x-\mu_{NP})^2}{2.s_{NP}^2}\right]dx = NGR.\int_{-\infty}^{\phi_C} \frac{1}{\sqrt{2\pi\, s_{P}^2}}.\exp\left[-\frac{(x-\mu_{P})^2}{2.s_{P}^2}\right]dx \tag{3-15}$$

Rearranging, it gives

$$(1-NGR).\int_{\phi_C}^{+\infty} \frac{1}{\sqrt{2\pi\, s_{NP}^2}}.\exp\left[-\frac{(x-\mu_{NP})^2}{2.s_{NP}^2}\right]dx - NGR.\int_{-\infty}^{\phi_C} \frac{1}{\sqrt{2\pi\, s_{P}^2}}.\exp\left[-\frac{(x-\mu_{P})^2}{2.s_{P}^2}\right].dx = 0 \tag{3-16}$$

The equation can be solved to obtain a porosity cut-off value ϕ_C which equalizes area A and area B (Figure 3-3), which represent respectively the probability of mistaking pay for non-pay and non-pay for pay.

This method is expected to be used for evaluating NGR since the errors, related to the prediction of pay and non-pay, are aimed at being cancelled out. The assumption about the normality of the non-pay and pay porosities is less restrictive than the assumption about the joint normality of the original dataset. This method should be compared to the performance of the RMA line for providing porosity cut-off value when evaluating NGR. As mentioned before, the discriminant analysis is expected to be less sensitive to the noise or the data scatters (inducing low correlation of the data) compared to the two previous methods, RMA and Y-on-X regression lines since it is a classic classification problem based on normal distributions, which are less affected by noise.

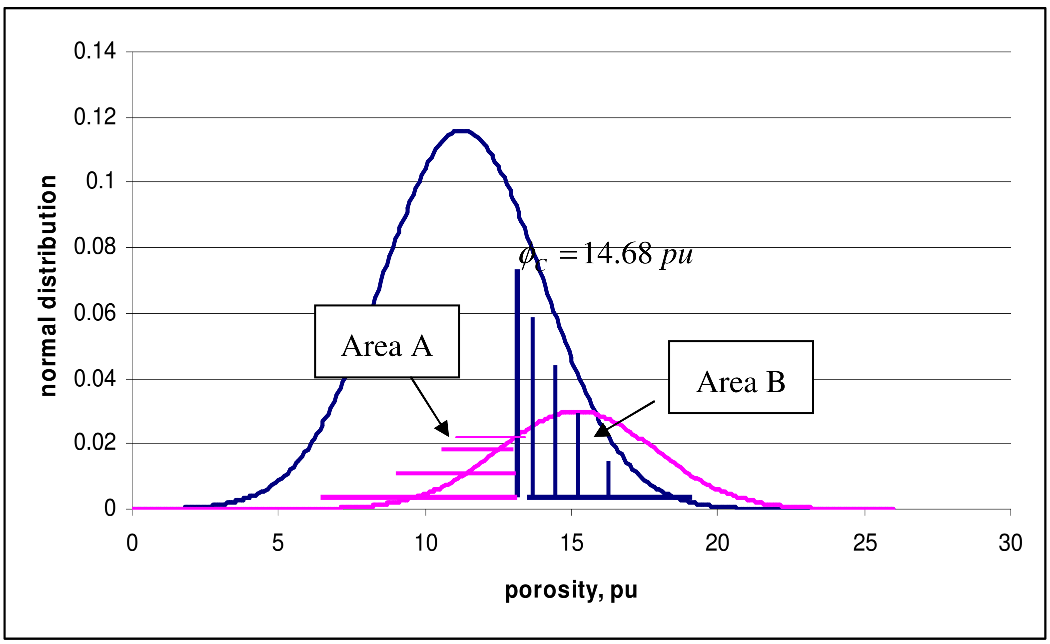


Fig. 3-3 Example of normal distributions of non-pay (blue) and pay porosity fraction (fuchsia), with the following characteristics: μ_{NP}=11.3, s_{NP}=2.75, μ_P=15.2, s_P=2.67 and NGR=0.2.

DATA-ORIENTED QUADRANT METHODS

Another approach to the problem of estimating optimal porosity cut-off value so as to evaluate NP and NGR is to consider a purely data-oriented method. On the basis of the probabilistic framework from Jensen and Menke (2006), the data points are classified using the threshold values k_c and ϕ_c. The probabilities A, B, C, and D may be estimated by counting the number of points belonging to one quadrant and making the ratio of this value over the total number of points.

For instance, $prob(A) = \dfrac{n_A}{n_{total}}$ (3-17)

Where n_A represents the number of points in the quadrant A and n_{total} represents the total number of points.

An algorithm may easily count points in each quadrant and calculate the probabilities for any given variable porosity cut-off value. The optimal porosity cut-off values minimizing the sum $prob(B) + prob(C)$ and equalizing the probabilities $prob(B)$ and $prob(C)$ may therefore be determined for any log (k)-ϕ dataset.

These methods are certainly interesting since they require no assumption concerning the distribution of the dataset. Two distinct porosity cut-off values may be obtained from those two methods to evaluate NGR and NP. There are, however, two limitations.

The first limitation of this approach is that the technique is sensitive to errors and the number of available samples. The sampling of a reservoir may be considered as the discretization of the PDF of variables k and ϕ. In the case where the number of samples is low, the sampled points may not be sufficient to represent and quantify the numerical diversity of the actual values for k and ϕ, leading to erroneous values for the porosity cut-off. It is especially the case for data set with less than 100 points: significant errors in the prediction of NP and NGR may be expected.

The second limitation concerns numerical issues. It is much easier to equalize $prob(B)$ and $prob(C)$, i.e. to find the porosity cut-off value so that the difference $prob(B) - prob(C)$ is nil, than try to find the cut-off value which minimizes the sum $prob(B) + prob(C)$. In the case where the dataset is relatively small, several minima may occur so that it is difficult to assess the actual optimal cut-off value. Despite these limitations, this method provides estimates of optimal porosity cut-off values, either to delineate NP or evaluate NGR without any assumptions regarding the PDF's of the variables.

SUMMARY

Four distinct methods have been described so as to provide porosity cut-off values to predict NP and evaluate NGR. Two linear regression methods are presented, the Y-on-X regression line and the RMA line.

The Y-on-X line method, usually given by the least-squares regression method, intends to minimize the errors in the predicted variable, i.e. log (k), assuming that no error is made on the regressor variable, i.e. ϕ. It has been confirmed from derivative analysis that this method provides the optimal porosity cut-off value to predict NP when log(k) and ϕ are assumed as JND. The latter assumption has led to propose that the optimal porosity cut-off value to evaluate NGR is given by the RMA line. The RMA line is designed to minimize the product of the variation in both directions of the cross plot so the line is expected to be less dependent of the degree of dispersion of the data, i.e. scale-dependence, than the Y-on-X line.

Discriminant analysis is also presented to give prediction of NGR using porosity cut-off values from the statistical properties of the pay and non-pay fraction of porosities. The PDF's of the two sub populations are assumed to be normal in this study. This method is less restrictive than the JND of log(k) and ϕ: PDF's for k and ϕ, which might be different from the normal distribution, may be determined for each fraction of porosities.

The last method is the quarter method which intends to determine the optimal porosity cut-off values from the statistical analysis of the k-ϕ cross plot (Jensen and Menke, 2006). Despite the dependency on the degree of sampling, this method does not require any hypothesis regarding the distribution of the dataset.

CHAPTER IV

EVALUATION METHODS FOR NP AND NGR ESTIMATORS

DEFINITIONS

Estimator assessment is a classic statistical problem covered by numerous authors. In practice, the true bivariate population of log(k) and ϕ value is sampled using core or log-derived measurements. A restricted number of points is measured, leading to an imperfect assessment of the characteristics of the true population, such as averages, standard deviations and coefficient of correlation. Limited sampling may lead to erroneous estimates of porosity cut-off values, $\hat{\phi}_c$ derived from the methods, i.e. the estimators, such as Y-on-X regression line, RMA line, discriminant analysis, and the data-oriented "quadrant" methods. A good estimator is expected to have specific characteristics as follows: small bias, good efficiency, robustness and consistency (Jensen et al., p. 96).

The bias may be defined by equation (4-1) as the difference between the expectation of the estimate, $E\left[\hat{\phi}_C\right]$, and the true value of the porosity cut-off, ϕ_C.

$$Bias = E\left[\hat{\phi}_C\right] - \phi_C = E\left[\hat{\phi}_C - \phi_C\right] \qquad (4\text{-}1)$$

In the case where the estimator bias is different from 0, i.e. meaning that the estimator tends to under-estimate or overestimate the true population optimal porosity cut-off ϕ_C, i.e. the estimator is referred to as biased. The confidence interval or standard error of the estimator allows assessing the accuracy and efficiency of the estimate

$$Std\ \ error = \left[Var\left(\hat{\phi}_C\right)\right]^{0.5} = \left[Var\left(\hat{\phi}_C - \phi_C\right)\right]^{0.5} \qquad (4\text{-}2)$$

The estimator robustness depends on the degree to which the estimates are influenced by errors occurring in the dataset. In the case where an estimate is unbiased, i.e. the bias is close to 0, and the standard error is minimized, the estimator is called a minimum variance unbiased estimator (MVUE). It can be also defined as a qualitative measure concerning how violation of the assumptions on which the estimator is based affect the results (e.g. is the estimator MVUE only if the samples are from a normal PDF, or is it still MVUE if the samples are from a log-normal or uniform PDF). The perturbation analysis allows evaluating the influence of variability on estimates by introducing noise, i.e. to assess the sensitivity of an estimator. Usually bias is ignored when it is small eg 10 % or less of the standard error. In the contrary, the root-mean squared error is considered to combine both bias and standard error.

For this purpose, the Monte Carlo method is really useful to assess the variability and performance of the estimates of porosity cut-off values given by the methods under investigation. The methods consist of generating stochastically values for k and ϕ from a population wit a known PDF. It allows simulating the behaviour of reservoir characteristics, characterized by a bivariate PDF, when sampled so as to assess the variability of the sub data sets and its influence on the estimates of porosity cut-off values.

SELECTION OF A BIVARIATE POPULATION DISTRIBUTION

To compare the various methods for NP and NGR estimation in terms of their bias, standard errors and their robustness, a true population has to be selected in order to use Monte Carlo methods to test the performance of the estimators on generated sub datasets. Data sets corresponding to tight reservoirs are taken into consideration. The newly discovered reservoirs are tighter and tighter, i.e. with lower NGR and permeabilities, since the conventional resources have been extensively exploited. Although it has always been an issue to define cut-off values, the problem is much more important and crucial when establishing cut-off value on tight reservoirs since a small error in the porosity cut-off may lead to a significant variation in the estimated NGR and have a strong impact on the economic feasibility of a project.

There are thousands of possibilities for defining the $k-\phi$ distribution. A good model to start with is the joint normality of $\log(k)-\phi$. It should reasonable to assume that the marginal distribution of porosities and log-permeabilities are normally distributed and that a correlation exists between those two variables. The assumption of JND is used in several studies for instance by Coker and Lindquist (1994). The use of this bivariate distribution also allows using the properties obtained from the derivative analysis: the optimal porosity cut-off values to evaluate NP and NGR may be determined from the quantities of the true population.

The statistical quantities of the population used for this study are defined as follows: μ_ϕ=12 pu, σ_ϕ=3, $\mu_{log(k)}$=-1, $\sigma_{log(k)}$=1, and ρ=0.7. It gives reservoir properties similar to a reservoir with an average permeability of $2.303*\exp(\mu_{log(k)} + 0.5*\sigma_{log(k)}\ ^\wedge 2) = 1.4$md (Figure 4-1).

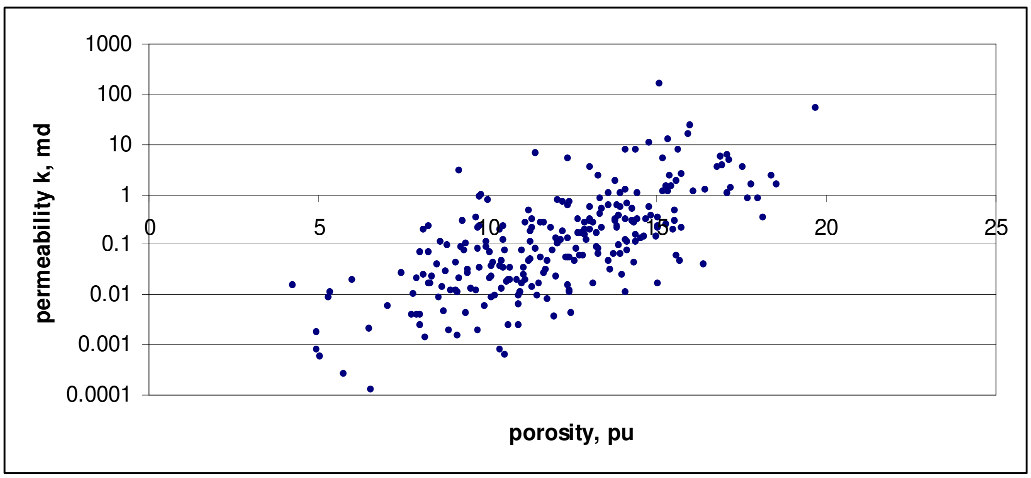


Fig. 4-1 Example of 250 sample joint normal log permeability-porosity dataset generated from the joint normal population with the following statistical quantities μ_ϕ=12 pu, σ_ϕ=3, $\mu_{log(k)}$=-1, $\sigma_{log(k)}$=1, and ρ=0.7.

EVALUATION OF ESTIMATOR PERFORMANCE

The study population may be sampled for any desired number of samples (a simple algorithm is presented in Appendix C). In order to assess bias and standard deviations of the estimators, numerous sub sets of bivariate data will be generated for several degrees of sampling, i.e. the number of samples N extracted from the population.

Numerous realizations for datasets with N= 25, 50, 75, 100, 250, 500, 1000, 5000 and 10,000 samples will be created and for each dataset the methods will be applied so as to obtain estimates of the porosity cut-off values. For cases where N < 1000, 1000 realizations are conducted to obtain a reliable estimation of the bias and standard error since the simulated results are expected to exhibit a significant variability. On the other hand, N ≥ 1000, only 100 realizations are done regarding that the simulated results are expected to be close to the irreducible values of bias and standard error, when N is large. On one hand, since the population is JND, the optimal porosity cut-off to delineate NP is that derived from the Y-on-X regression line using the statistical quantities of the population.

From equation (3-10),

$$\frac{\log(k_c) - \overline{\log(k)}}{\sigma_{\log(k)}} = \rho . \frac{\phi_{cNP} - \bar{\phi}}{\sigma_{\phi}} \tag{4-3}$$

and rearranging it gives

$$\phi_{cNP} = \frac{\log(k_c) - \overline{\log(k)}}{\sigma_{\log(k)} . \rho} . \sigma_{\phi} + \bar{\phi} . \tag{4-4}$$

On the other hand, the optimal porosity cut-off to evaluate NGR is that derived from the RMA line using the statistical quantities of the population.

From equation (3-9),

$$\frac{\log(k_c) - \overline{\log(k)}}{\sigma_{\log(k)}} = 1 . \frac{\phi_{cNGR} - \bar{\phi}}{\sigma_{\phi}} \tag{4-5}$$

Rearranging,

$$\phi_{cNGR} = \frac{\log(k_c) - \overline{\log(k)}}{\sigma_{\log(k)}} . \sigma_{\phi} + \bar{\phi} . \tag{4-6}$$

Using the statistical quantities defined as follows $\mu_{\phi} = 12$ pu, $\sigma_{\phi} = 3$, $\mu_{\log(k)} = -1$, $\sigma_{\log(k)} = 1$, and $\rho = 0.7$, the optimal porosity cut-off values may de determined for the population for any defined permeability cut-off value (Table 4-1).

Table 4-1 Porosity cut-off values derived from the RMA and Y-on-X regression line for several permeability cut-off values, k_c, using the statistical quantities from the true joint normal population.

k_c	0.01 md	0.1 md	0.5 md	1 md
ϕc(NGR), pu	9	12	14.09	15
ϕ_c(NP), pu	7.71	12	14.996	16.2857

For every sub dataset, an estimate of the porosity cut-off value will be calculated along with the calculated quantities for different permeability cut-off values by using equations (4-3) and (4-5). The porosity cut-off values will be generated using the methods described in Chap. 3: $\phi_{cY\text{-on-}X}$ derived from the Y-on-X regression line, ϕ_{cRMA} obtained from the RMA line, a porosity cut-off, ϕ_c, derived from the discriminant analysis by solving equation (3-16) and two distinct values from the data-oriented "quadrant" method (one porosity cut-of value minimizing the sum of the errors of mistaking pay for non-pay and pay for non-pay and the second one canceling out the errors). For each sub set, the difference between the estimated cut-off value and the true cut-off value (determined from the quantities of the population) will be computed. The mean, i.e. bias, and standard error of those differences $\hat{\phi}_C - \phi_C$ will be computed from those realizations: the four estimators are therefore assessed with respect to the degree of sampling by plotting the values of bias and standard errors versus the number of samples.

The biases of the estimates, expressed by equation (4-1), are plotted vs. the inverse of the number of sample since the bias is expected to be proportional to $1/N$ as the expectancy of a discrete variable X is defined as follows:

$$E(X) = \frac{1}{N}\sum_{i=1}^{N} X_i \qquad (4\text{-}7)$$

The standard errors of the estimates are plotted versus $1/\sqrt{N}$. In fact, the standard deviation of a discrete variable may be expressed as follows:

$$s^2 = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - E(X)\right)^2 \qquad (4\text{-}8)$$

where s represents the standard deviation and s^2 the variance.

The standard errors of the estimates are therefore expected to be proportional to $1/\sqrt{N}$. The analysis of the performance of the methods is realized by studying two plots, bias vs. $1/N$ and standard error vs. $1/\sqrt{N}$.

Various characteristics are investigated so as to evaluate the performance of the estimators. When the number of samples is sufficiently high, we can consider $1/N \to 0$ and $1/\sqrt{N} \to 0$. The values of bias and standard error may be therefore extrapolated to obtain the intercept to the Y-axis. These values of bias and standard error are irreducible since the degree of sampling is infinite, meaning they represent the minimum bias and standard error of estimates that can be obtained from a specific method.

First, bias at the intercept should be as low as possible, in order to obtain estimates that are not significantly different from the true value, i.e. “unbiased” (Figure 4-2). Second, the value of standard error at the intercept should be compared to 0, by using hypothesis test for instance. In the case where the intercept is different from 0, it means that an irreducible variability exists which is independent of the problem of estimation (Appendix D presents test on the intercept of the standard error lines for the different methods). The use of an estimator with a high irreducible variability is not recommended and should be avoided.

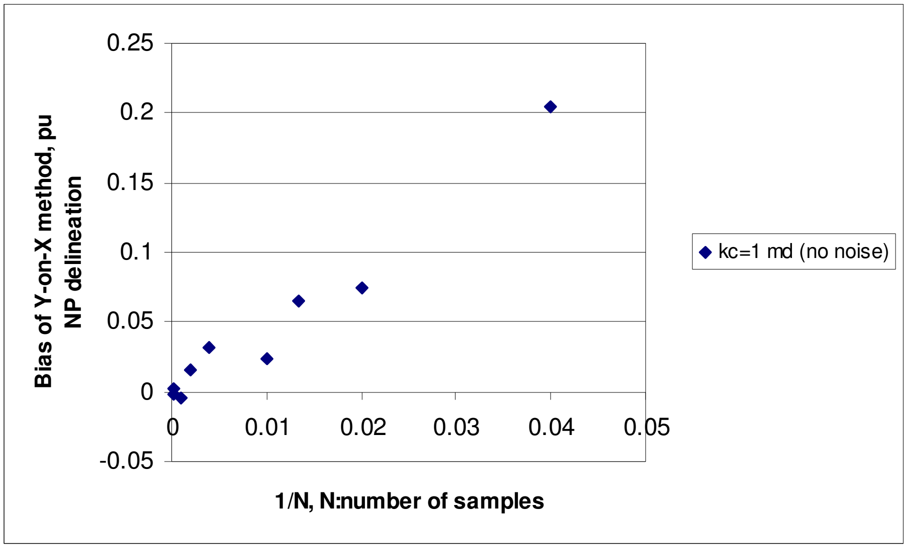


Fig. 4-2 Example of bias of estimates derived from the Y-on-X line with respect to the number of samples.

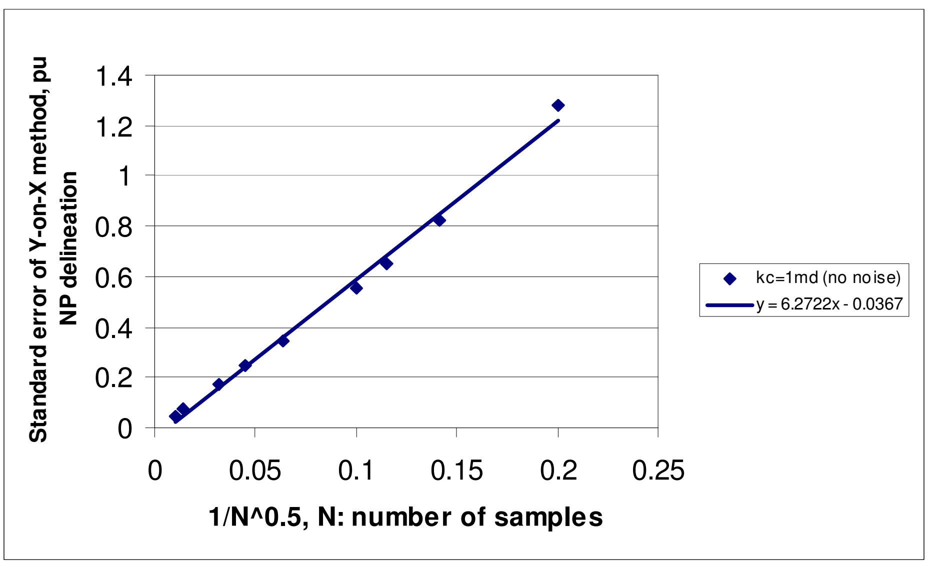


Fig. 4-3 Example of standard error of estimates derived from the Y-on-X line with respect to the number of samples.

The other component of the plot analysis is to investigate the variations and trends of bias and standard errors of the estimates for the different methods. The bias might be relatively constant or decrease with the number of samples. The standard error is expected to decrease linearly with the number of samples. The slope of the standard error line should be as low as possible to expect estimates with minimum variability. The standard error of an estimator also gives the confidence error of the estimates given by a method. For instance, for 50 sample datasets, the estimates of porosity cut-off values have a confidence interval of +/- 0.8 p.u. (Figure 4-3).

The four methods will be thus compared and tested for their ability to predict accurate and reliable porosity cut-off values, on the basis of their bias and errors in the case where either NP or NGR is required to be evaluated. The estimates from the discriminant analysis and derived from the RMA line are nonetheless expected to be significantly biased for delineating NP. The estimates obtained from the Y-on-X regression line are also expected to be biased and higher than the true porosity cut-off value for evaluating NGR owing to the slope ρ, less than 1 (Figure 4-4).

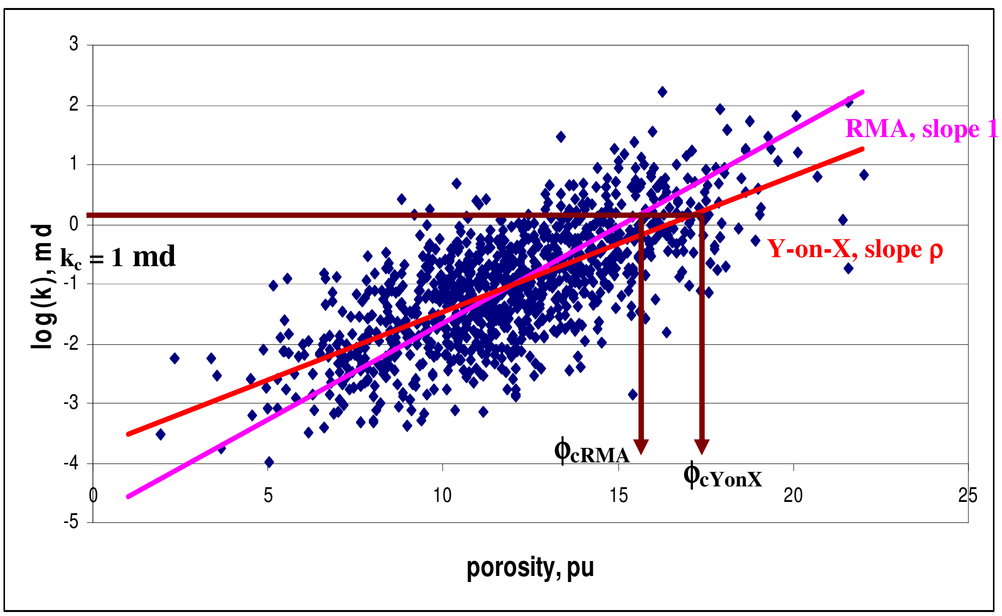


Fig. 4-4 The RMA and Y-on-X line and their respective slope, leading to distinct porosity cut-off values.

The bias and standard errors will be estimated for all methods to both delineate NP and evaluate NGR based on the number of samples. The robustness of those estimators should also be taken into consideration since the joint normality of log(k) and ϕ may be corrupted by the measurement errors or reservoir characteristics inherent to the depositional setting (e.g. heterogeneities, multi pore type especially in carbonate rocks), which leads to the mixing of different permeability laws, Log (k) = a . ϕ + b. The introduction of noise also significantly deviate the datasets from the joint normality.
Introducing noise is therefore necessary to assess the degree to which the estimates are influenced by errors occurring in the datasets and evaluate the robustness of the methods. The distribution of the noise is also an important issue and a uniform distribution was used in this study (Figure 4-5).

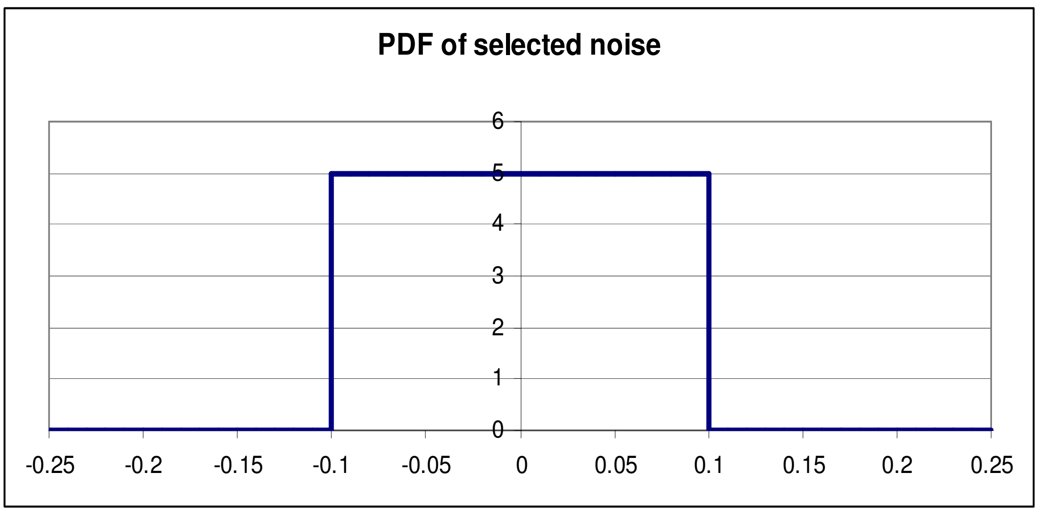


Fig. 4-5 Uniform distribution of noise, ε, whose range is comprised between -0.1 and 0.1.

Two different corrupted populations are generated by modifying the porosity and log (k) values of the original values as follows:

$(k)_{noise1} = (k)_{original} * (1 + 3*\varepsilon)$

$\phi_{noise1} = \phi_{original} * (1 + \varepsilon)$

$(k)_{noise2} = (k)_{original} *(1+ 5*\varepsilon)$

$\phi_{noise2} = \phi_{original} *(1+ 2*\varepsilon)$

where ε is a uniformly distributed random variable (Figure 4-5).

The relationships between porosity and log permeability are highly variable and mainly depend on the texture of the reservoir rock. In this study, we only consider the interparticle (or intercrystalline when the rock has been deposited by chemical process) case: the log (k) and ϕ relationship might be considered as linear with respect to the grain size (Lucia, 1999). A small variation of porosity expressed in porosity units, p.u. (driven by the pore diameter), thus induces a large variation in the permeability k expressed in md (driven by the pore throat size). Two magnitude of noise on porosity are selected, 10 % and 20 %, considering that measurement uncertainty of porosity is relatively low regarding the issues related to permeability measurements. However, these noises on porosity will lead to a greater variation in permeability which is fixed to respectively 30% and 50 %. These noises are intended to corrupt the bivariate population from the joint normality. The first noise, noted noise 1, is of mild magnitude whereas the second one, noted noise 2, is of higher magnitude.

Porosity cut-off values will be estimated using the different methods on the corrupted datasets. Bias and standard errors of estimates will be then estimated. The robustness of the estimators will be determined from a comparison of the abilities of the methods to predict correct porosity cut-off values with and without noise. The ability of the estimators to predict original cut-of value will be investigated since the noise will spread out the data and alter the quantities of the generated datasets. The coefficient of correlation is indeed very sensitive to the degree to which the data are spread out. Thus, the Y-on-X regression line estimates may be significantly affected by the introduction of noise.

CHAPTER V

RESULTS AND INTERPRETATIONS

Each method for estimating porosity cut-off values is successively presented in this chapter: bias and standard error of their estimates with regard to the degree of sampling are analyzed and evaluated in both qualitative and quantitative ways. All methods are evaluated for their ability to both predict NP and NGR, even though the assumption of JND would suggest a preference for the RMA line for NGR and the Y-on-X line for NP. In the same way, the discriminant analysis is also expected to provide better results for NGR than for NP since the method intends to cancel out the misidentification of pay and non-pay fraction of the porosities, which is a similar approach to equalize the probabilities B and C.

The influence of the permeability cut-off value is investigated for both RMA and the Y-on-X line methods. A permeability cut-off of 1 md, i.e. one decade higher than the median permeability of the JND population is selected to realize the sensitivity assessment and evaluate the robustness of the estimators. The data sets are also respectively corrupted with noise 1 and noise 2: bias and standard errors of the obtained estimates are computed.

Finally, the methods are compared to one another regarding their bias, standard error and robustness for the selected permeability cut-off value of 1 md and the cases where there is no noise, data are corrupted with noise 1, and when data are corrupted with noise 2.

BRIEF SUMMARY OF THE RESULTS

When evaluating NGR, the RMA line and the discriminant analysis provides estimates of porosity with good confidence interval and low bias. Those two methods are robust since the introduction of noise does not induce significant additional bias and standard error. Those methods, when used to delineate NP, give highly biased estimates since the Y-on-X line gives the analytical optimal porosity cut-off value. The quadrant method is significantly less efficient than the RMA line and the discriminant analysis to predict NGR but it is an interesting method since no assumptions concerning the distribution of the data are required.

Regarding the delineation of NP, the Y-on-X line remains the method to be used to obtain porosity cut-off value. The overall performance of its estimates is lower than the RMA line since standard errors are nearly twice as high and the method is clearly less robust since the introduction of noise introduces an irreducible bias. The quadrant method is for this case not recommended since its estimates are highly biased and have a poor confidence interval: it should be used in the case where no assumption may be made on the distribution of the data and when N is large.

PERFORMANCE OF THE Y-ON-X REGRESSION METHOD

The performance of the Y-on-X regression method for delineating NP and evaluating NGR is investigated by calculating the bias and the standard error of the estimator. Numerous realizations are made for each sampling case N: 1000 realizations for case where $N < 1000$ and 100 realizations for case where N is equal to 1000, 5000 and 10,000 samples. The Y-on-X regression method is applied and porosity cut-of values are derived for different permeability cut-of fvalues. The bias and the standard error of the estimates of the optimal cut-off value for delineating NP may be therefore determined by using equations (4-1) and (4-2).

The bias of the estimators when delineating net pay exhibits symmetry centred on the permeability cut-off value of 0.1 md, which corresponds to the median permeability of the JND population (Figure 5-1). In this case, the variability of the estimator is minimized so that it can be defined as an MVUE. Whatever the permeability cut-off value is, the biases are considered as insignificant: for instance, bias for the case where k_c= 1 md does not exceed 0.2 p.u., which represents less than 2 percent of error (Figure 5-1).

The bias of estimates is therefore proportional to the difference between the permeability cut-off value and the median permeability of the population, i.e. in our case 0.1 md. The higher or/and the lower the permeability cut-off value is than the permeability average, the higher the bias. We thus observe that bias of estimates is minimized for k_c=0.1 md: the Y-on-X and RMA lines are not significantly different from each other (Figure 5-1) at this location. It corresponds to the center of gravity of the ellipse representing the PDF for a JND population. The influence of the outliers is minimized in this area and the density of points is maximized, leading to an optimal estimation of the actual coefficient of correlation. The estimates are therefore unbiased.

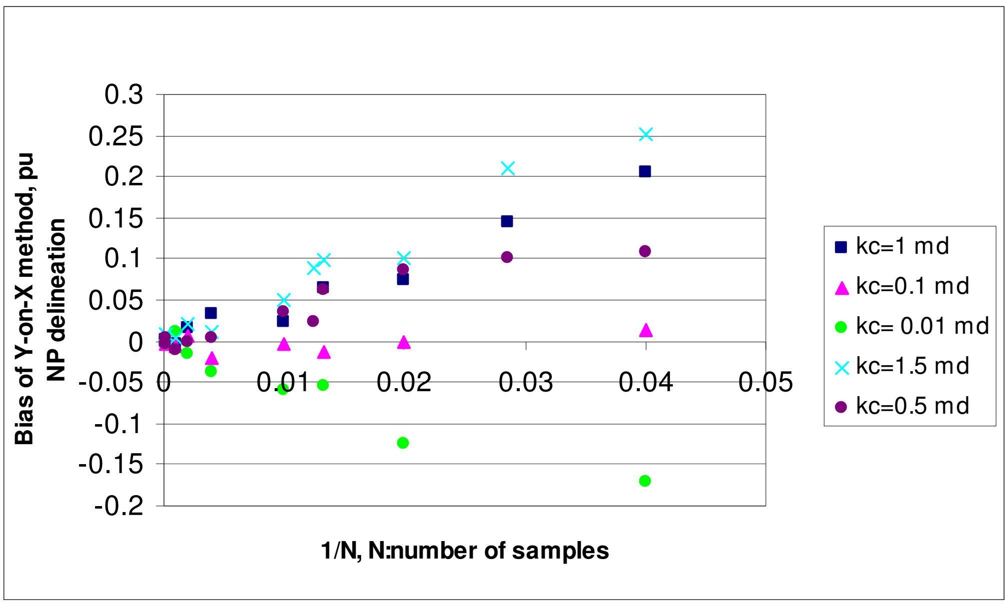


Fig. 5-1 Biases of the estimator for delineating NP for various permeability cut-off value, k_C.

As the permeability cut-off values are more and more decentred from the permeability average, the variability of the Y-on-X method will increase (Figure 5-2). It explains why the standard error of cases where kc=1 md and kc=0.01 md are the same. The variability of the Y-on-X method is significant: for instance, in the case N = 50 samples are measured from the population assuming that k_c = 1 md, the standard error is 1 p.u., i.e. the predicted porosity cut-off value will have a 68% confidence interval of +/-

1 p.u. In order to decrease this variability in the measurement by 20 percent, i.e. to ensure a relatively lower error of +/- 0.8 p.u., 25 additional samples should be taken (Figure 5-2). The root-mean square error (root-MSE) is therefore 0.175 p.u..

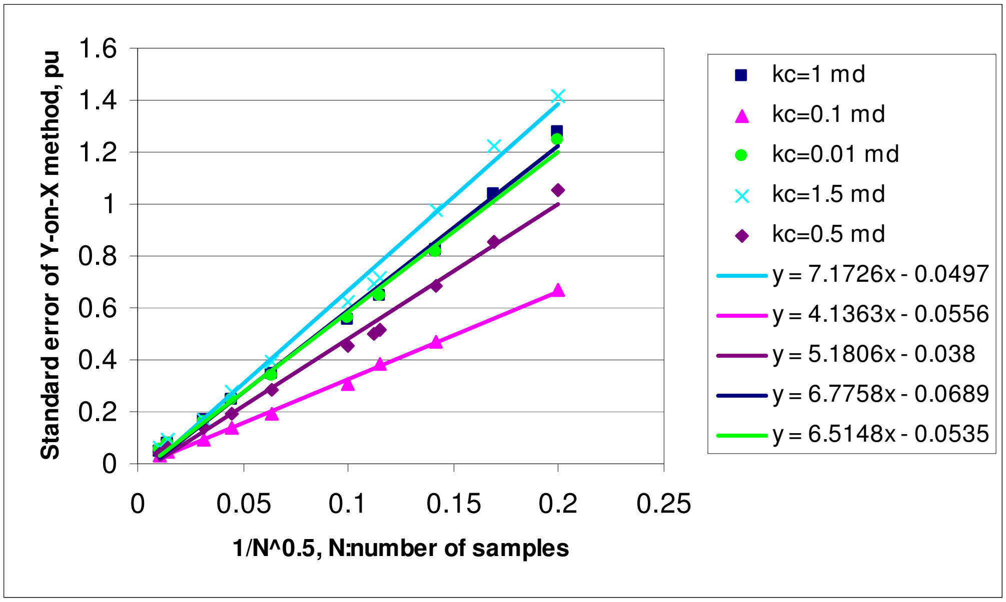


Fig. 5-2 Standard error of the estimator for delineating NP given by the Y-on-X regression line.

Best-fit lines may be fitted to the standard error points for the different permeability cut-off. The standard errors of estimates from the Y-on-X line exhibit a straight-line behaviour. The variability of the estimates is minimized for k_c= 0.1 md, as the slope of the line is the lowest (Figure 5-2). The higher the absolute difference between the median permeability value, 0.1 md, and the permeability cut-off, the higher the variability of the estimates. Using the equations of the best-fit lines, the intercept may be calculated in order to obtain the value of standard error where n tends to infinity. Those values are tested so as to determine whether they are statistically different from 0 (Appendix D). In the case where the intercept is 0, it means that no inherent variability exists when the degree of sampling is infinite. Obviously the intercepts are smaller than 0.1 p.u. and assumed to be nil.

On the basis of the analytical justification of the preferential use of the Y-on-X regression line to delineate NP, the use of the Y-on-X regression line so as to evaluate NGR will lead to an inherent bias (Figure 5-3). This bias derives from the fact that the Y-on-X line is not the analytical line for the NGR porosity cut-off. The standard error is not influenced by the value of the true optimal porosity value ϕ_c, either ϕ_{cNP} or ϕ_{cNGR}, as shown by equation (4-2) and Figure 5-4.

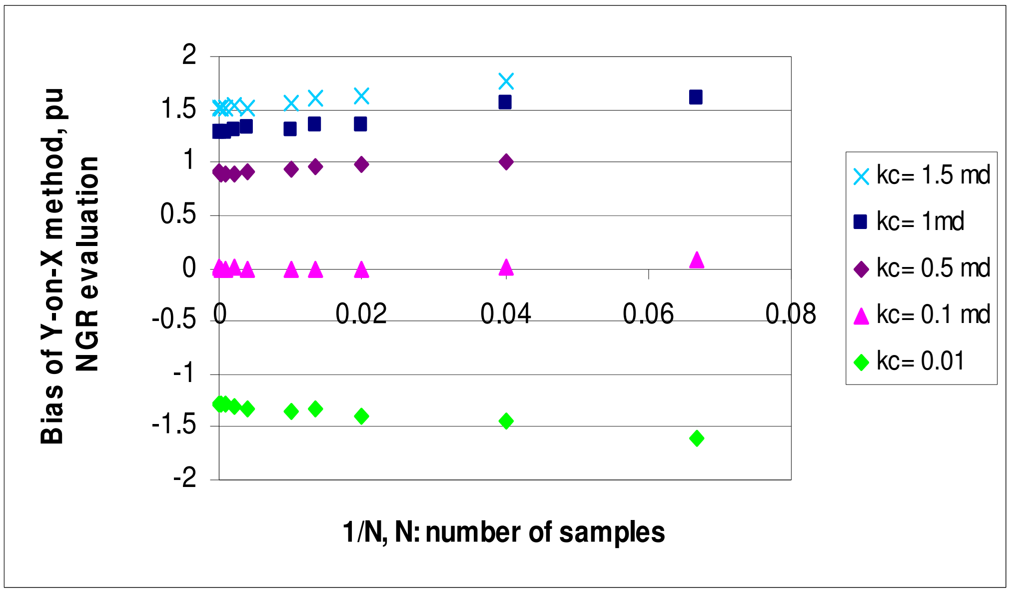


Fig. 5-3 **Biases of the estimator for evaluating NGR for various permeability cut-off value, k_C.**

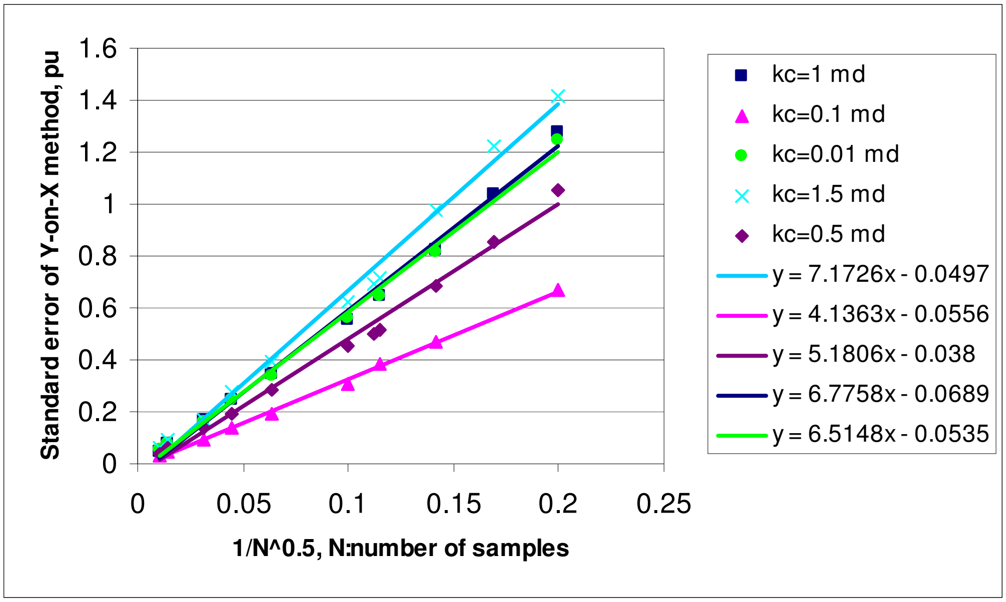
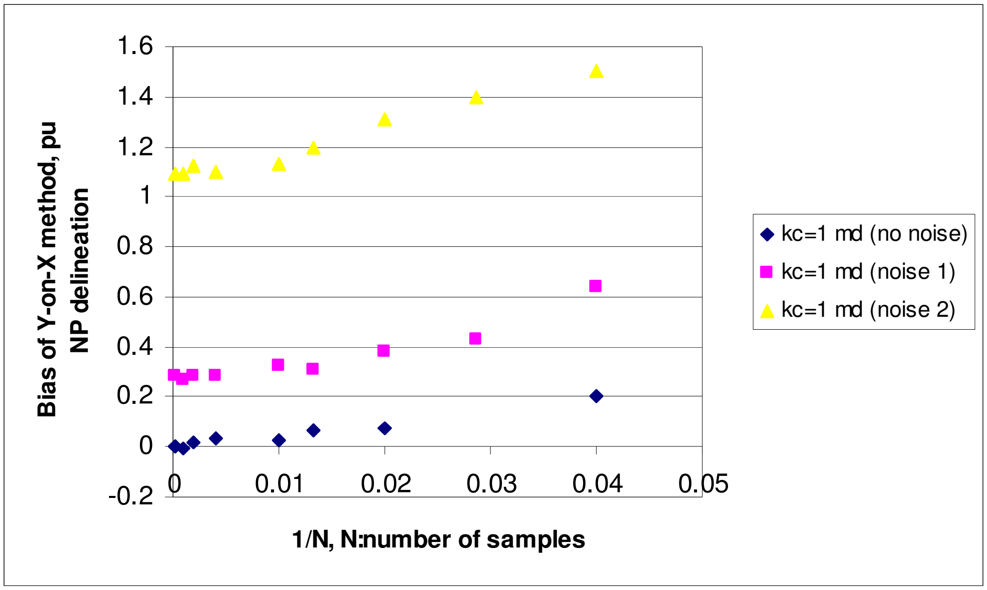


Fig. 5-4 Standard error of the estimator for evaluating NGR given by the Y-on-X regression line.

The robustness of the Y-on-X method is also investigated by using the two types of noise described in chapter IV. The corruption of the bivariate data drives an increase in the variability of the estimates in addition to higher biases (Figure 5-5). The behaviour of the standard error with respect to degree of sampling is no longer linear in the case where the datasets are corrupted with noise. In the case where the data sets are corrupted with mild noise (“noise 1”) the linearity of the standard error disappears for less than 75 samples. When the data sets are corrupted with noise of higher magnitude, the standard error no longer evolves linearly with the number of samples for data sets with less than 100 points (Figure 5-5).



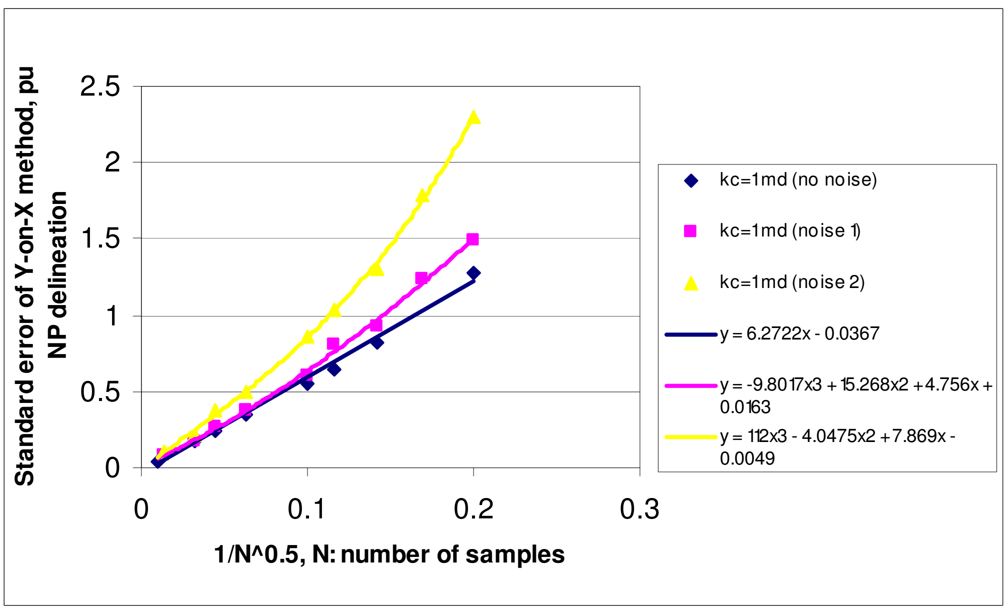


Fig. 5-5 Bias and standard error of estimates given by the Y-on-X method for non-corrupted data, data corrupted with noise of magnitude 1, and data corrupted with noise of magnitude 2. A permeability cut-off value of 1 md is used to derive the porosity cut-off value from the Y-on-X regression line.

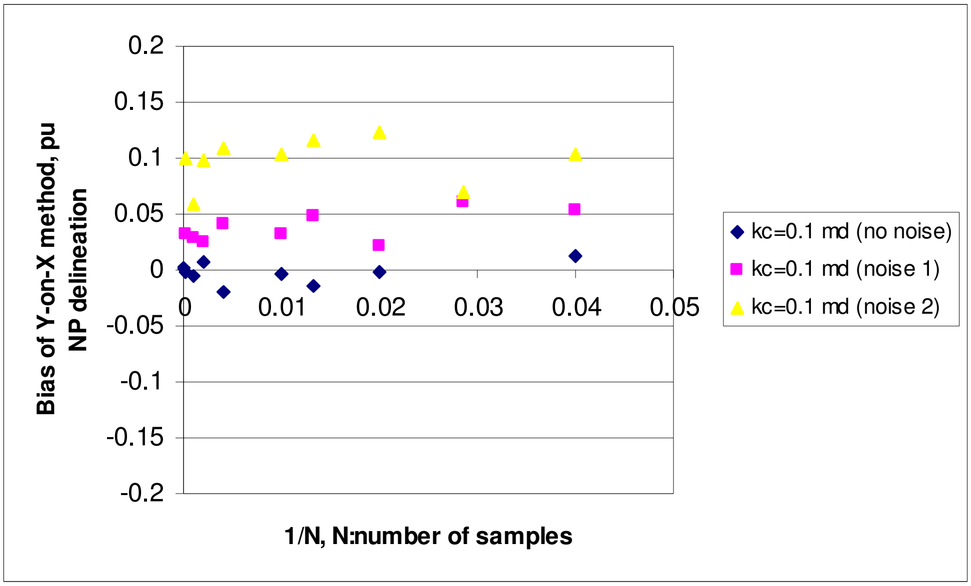
The corruption of the joint normality induces an increase in bias of estimates. It is no longer insignificant and represents a significant systematic error comprised between 1.1 and 1.5 p.u. when noise 2 is used.

The bias is not reduced to 0 when the number of samples tends to infinity, as shown on Figure 5-5. The estimates are still significantly biased even when N=5,000 samples for datasets corrupted with noise 1 and 2 (Figure 5-5). It may be explained by the influence of the noise which will alter the correlation of the data: it will lead to systematic biases even though the number of sample is really large.

All those observations highlight the great dependency of the Y-on-X regression on the degree to which data are spread out, i.e. the dependency on the value of ρ, the coefficient of correlation. This quantity is more affected by the corruption of the data than the others statistical quantities, such as averages and standard deviations. The bias of the estimates from the Y-on-X method increases as the number of samples decreases. In the case where 25 samples are available, for example, the initial bias of 0.2 pu with a joint normal dataset will reach 0.6 and 1.5 pu with datasets respectively corrupted with noises of magnitude 1 and noise of magnitude 2.

The non-linearity (for datasets with less than 75 samples) and the high variability of the estimator is still really significant for the case where k_c= 0.1 md (Figure 5-6). In this case, the standard error of the estimator is approximately twice as small as that for the k_c= 1 md. Bias does not differ significantly from 0 (when data sets are corrupted with noise 2 the bias does not exceed 0.1 p.u.), even though biases are increasing with fewer sampled data.

If 25 samples are measured, the porosity cut-off value estimates will have a bias of 0.1 pu and a 68 % confidence interval of +/- 1.1 p.u. for data sets corrupted with noise 2. When the permeability cut-off value equals the median permeability of the population, the estimator's bias and standard error are minimized.



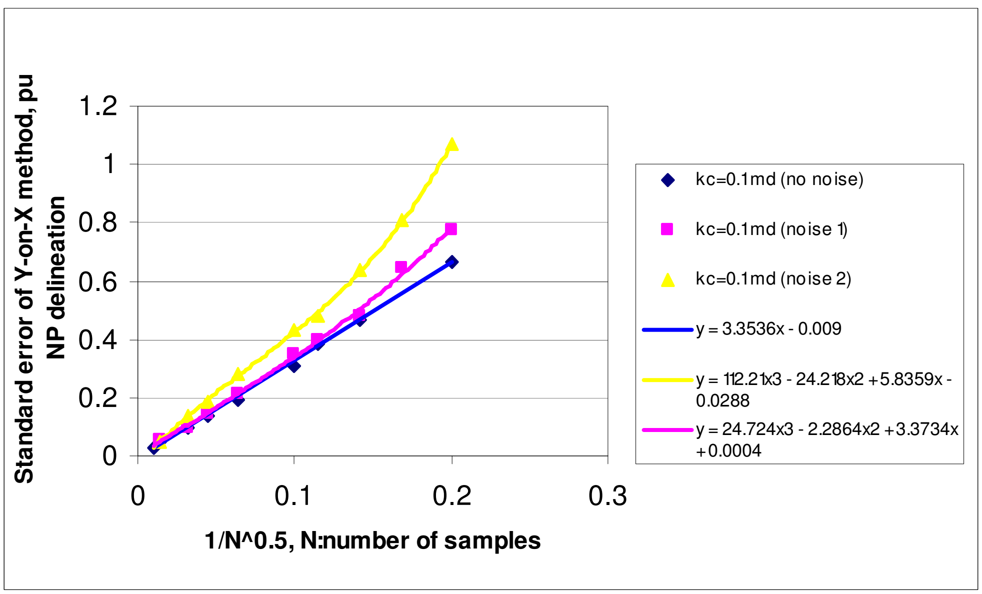


Fig. 5-6 Bias and standard error of estimates given by the Y-on-X method for non-corrupted data, data corrupted with noise of magnitude 1, and data corrupted with noise of magnitude 2. A permeability cut-off value of 0.1 md is used to derive the porosity cut-off value from the Y-on-X regression line.

If the Y-on-X regression line is used to predict the optimal porosity cut-off value for evaluating NGR, an inherent bias is introduced leading to systematic improper porosity cut-off values, whatever the noise is (Figure 5-7).

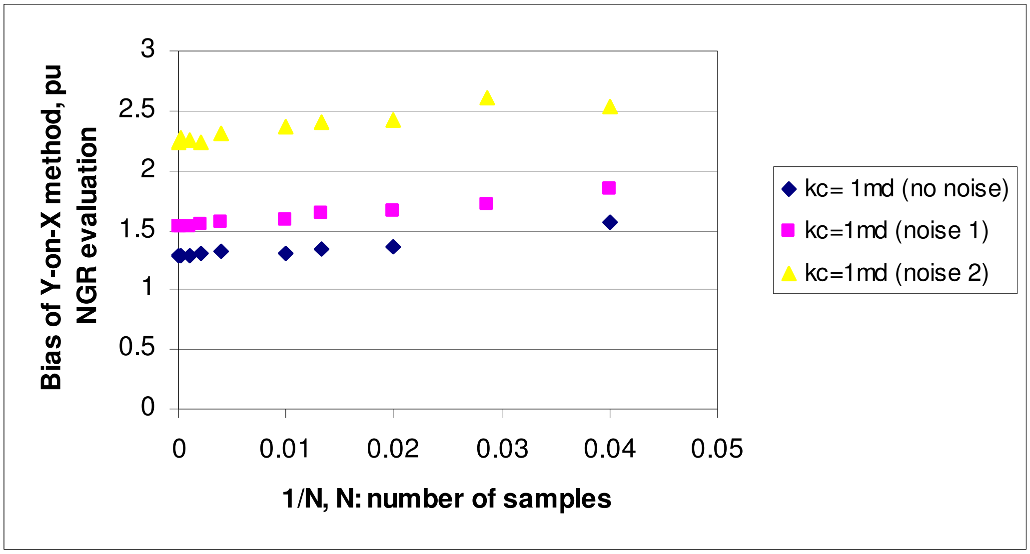


Fig. 5-7 Bias of estimates given by the Y-on-X method for non-corrupted data, data corrupted with noise of magnitude 1, and data corrupted with noise of magnitude 2. A permeability cut-off value of 1 md is used to derive the porosity cut-off value from the Y-on-X regression line.

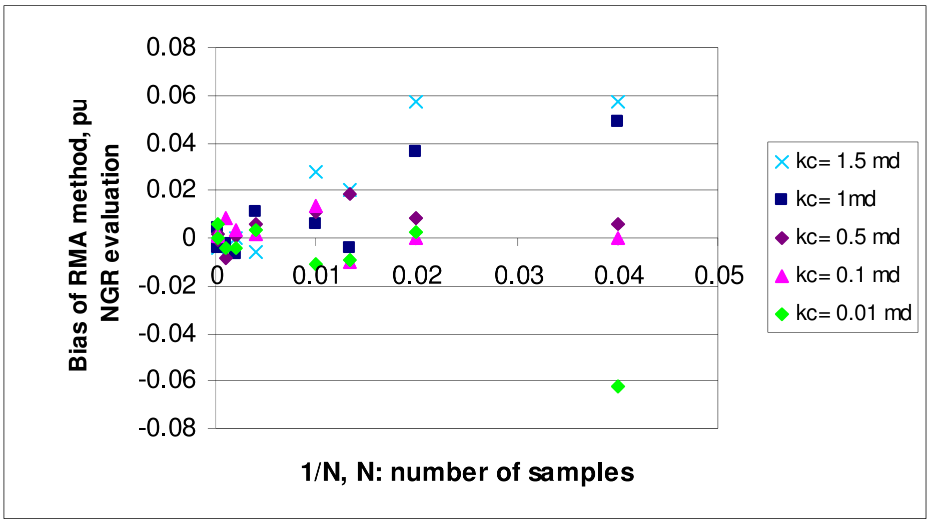
The Y-on-X regression line has been proven to provide optimal porosity cut-off values when log(k) and ϕ are joint normal distributed. However, the estimator presents a high dependency over the estimate of the coefficient of correlation, which induces an unavoidable bias, a high variability in the predicted porosity cut-off values and the non-linearity of the standard error of the estimates. The cause of this variability may be due to the limited number of samples or the corruption of the joint normality.

The prediction of NP obtained from those porosity cut-off values might be significantly erroneous. When the number of sample is les than 100 points (when the behaviour of the standard error is likely to be non-linear), the measurement of additional samples is in this case greatly recommended since increasing the degree to which the reservoir is sampled may significantly decrease the variability of the estimates of cut-off values from the Y-on-X line. For instance, considering a permeability cut-off value of 1 md, measuring 50 extra samples from a 25 point dataset decreases the error of the estimates by 40 percent.

PERFORMANCE OF THE RMA LINE METHOD

The same methodology is used to assess the performance of the RMA line to predict porosity cut-off values for delineating NP and evaluating NGR. The performance of the RMA method is therefore investigated by calculating the bias and the standard error of the estimates for delineating NP and evaluating NGR. For this purpose, numerous realizations are made for each sampling case: 1000 realizations for case where N < 1000 and 100 realizations for case where N is equal to 1000, 5000 and 10,000 samples. Different permeability cut-off values are used to obtain the estimates of the porosity cut-off values through the RMA line equation, expressed by equation (4-5).
The bias and the standard error of the estimates of cut-off value for evaluating NGR may be thereafter determined by using equations (4-1) and (4-2). The variability and the bias of the estimates will be dependent, in the same way as for the Y-on-X regression line, of the value of the permeability cut-off values. When the permeability cut-off value is close to the permeability average of the bivariate population, the RMA line and the Y-on-X line are not significantly different from one another. The biases are therefore relatively low, less than 0.1 pu, and may be considered as insignificant (Figure 5-8). The variability of the estimator is also minimized for the case where the permeability cut-off value is equal to 0.1 md: the estimator may de defined as a MVUE. Similarly to the Y-on-X method, symmetry with respect to the permeability average value is therefore observed on the bias and standard error.

The standard error of the estimates from the RMA line is significantly lower than those from the Y-on-X line (Figures 5-4 and 5-8). The performance of the RMA line is much higher since the variability, i.e. the ability of the method to provide estimates with good confidence intervals, is lower. It may be interpreted by the analytical expressions of the respective equations for both lines, given by equations (3-9) and (3-10). In contrast to the Y-on-X line, the RMA line is independent of the coefficient of correlation, meaning that this line is less sensitive to the outliers and the spatial distribution of the points on the log(k)-ϕ cross plot. It induces a lower variability of the estimates of porosity cut-off values. The Y-on-X line analytically remains the method which provides the optimal porosity cut-off for delineating NP whereas the RMA line gives the best estimates of porosity cut-off to evaluate NGR.



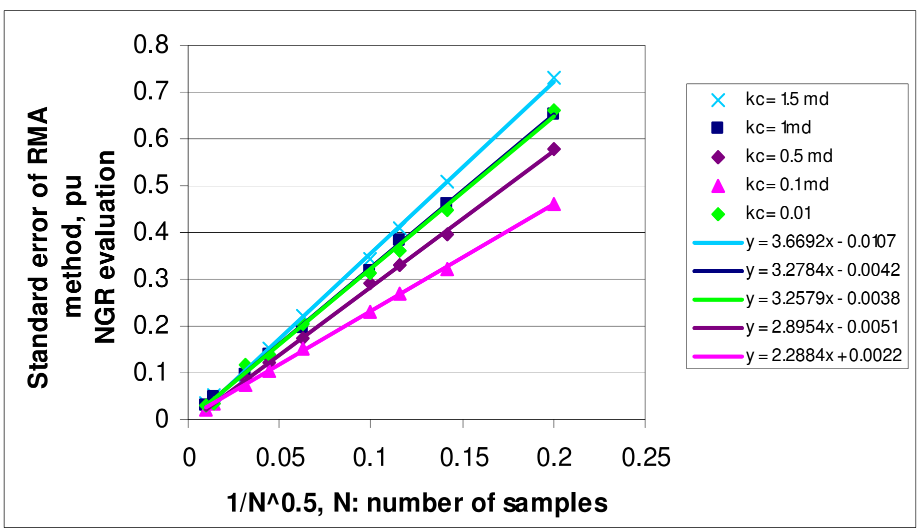


Fig. 5-8 Bias and standard error of porosity cut-off values given by the RMA line for different permeability cut-off values, k_C.

The behaviour of the standard error of the estimates of porosity cut-off is clearly linear and the intercepts of those lines is insignificantly different from 0.

For a permeability cut-off value of 1 md and for a 25 sample dataset, the estimate will be nearly unbiased, i.e. the expectancy of predictions made from the RMA line, $E\left|\hat{\phi}_C - \phi_C\right| = 15.05\ pu$, is nearly equal to the optimal porosity cut-off value, ϕ_{cNGR}=15 pu. The corresponding confidence interval will be +/- 0.65 pu (Figure 5-8).

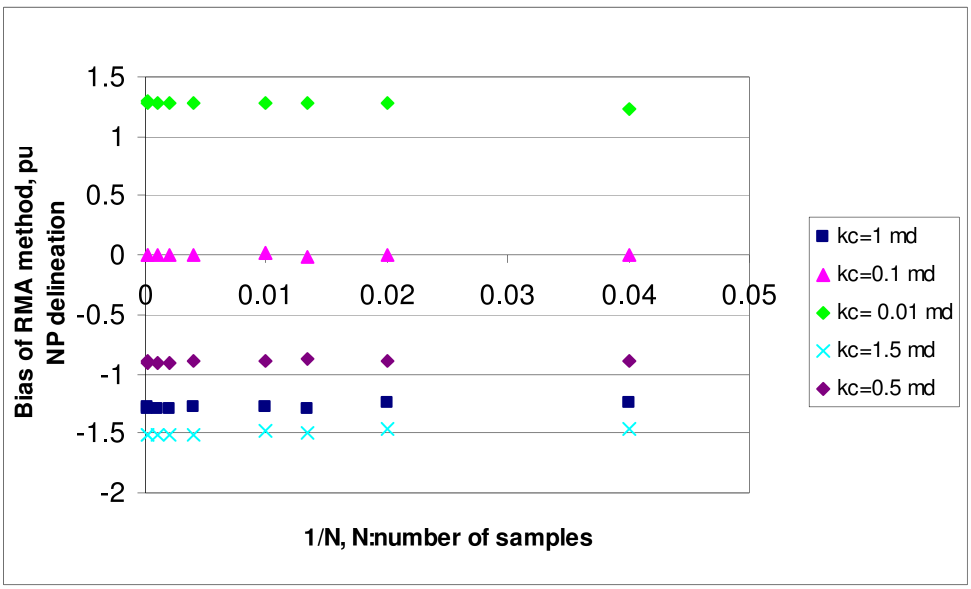


Fig. 5-9 Bias of estimates given by the RMA method. Various permeability cut-off values are used to derive porosity cut-off values so as to delineate NP.

For the case k_{c}=1 md and where 25 samples are available, sampling 25 additional points will decrease by 18 percent the variability of the estimates of the predicted porosity cut-off values. The use of RMA line is recommended since it provides reliable estimates of porosity cut-off values to evaluate NGR, i.e. with no bias and low variability.

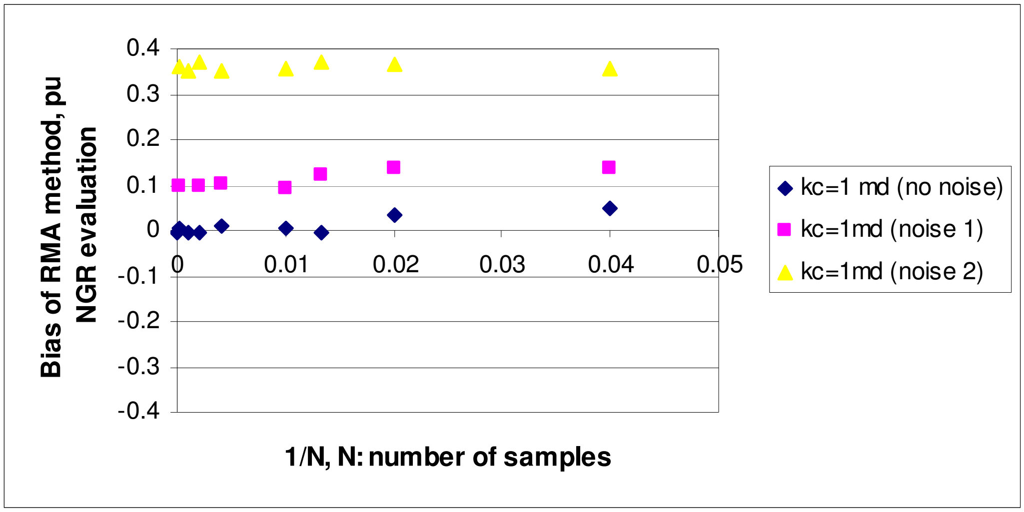
Using the RMA line to predict optimal porosity cut-off values for delineating NP will lead to erroneous NP delineation since the method provides estimates that are inherently biased as shown on Figure 5-9. For instance, for k_c = 1 md, the porosity cut-off values from the RMA line are systematically underestimated by 1.25 p.u., i.e. 7.7 %: the bias is therefore higher than the standard error and must no longer be ignored.

When the permeability cut-off value equals the average of permeability of the population, it seems reasonable to use the RMA line instead of the Y-on-X line since the variability of estimates is lower (the standard error of the estimates for both delineate NP and evaluate NGR are the same as explained previously).

The robustness of the RMA method is also investigated by corrupting the original dataset with noise 1 and 2 using the same procedure as for the study of the performance of the Y-on-X line. The ability of the estimator to predict the optimal porosity cut-off value for evaluating NGR is thus assessed when a significant deviation from the joint normality is observed. Figures 5-10 and 5-11 show the bias and standard error of estimates given by the RMA line for two permeability cut-off value and various magnitude of corruption of the original joint normality.

For instance, bias of estimates of porosity cut-off values is less sensitive to the corruption of data than the Y-on-X line: the introduction of the noise 2 lead to a bias of 0.35 p.u., i.e. that the expectancy of the estimates for predicting NGR is slightly higher than the optimal value, ϕ_{cNGR}= 15 pu. These biases are constant whatever the number of sample is available (Figure 5-10).

The higher the magnitude of the noise is, the higher the variability of the estimates is. The variability of the estimates increase by 48 percent if the datasets are corrupted with noise 1, and increase by 81 percent if the datasets are corrupted with noise 2. The influence of the noise on the performance of the estimator is thus really important but the evolution of the variability with respect to the degree of sampling remains linear, on contrary to the variability of the Y-on-X method. Second, the variability of estimates is largely lower than those from the Y-on-X line and may be reduced by increasing the degree of sampling. For instance, estimates of porosity cut-off for 25 sample data sets have a 68% confidence interval of +/- 0.8 p.u.. Decreasing this variability by 25 percent requires sampling 25 additional points (Figure 5-10).



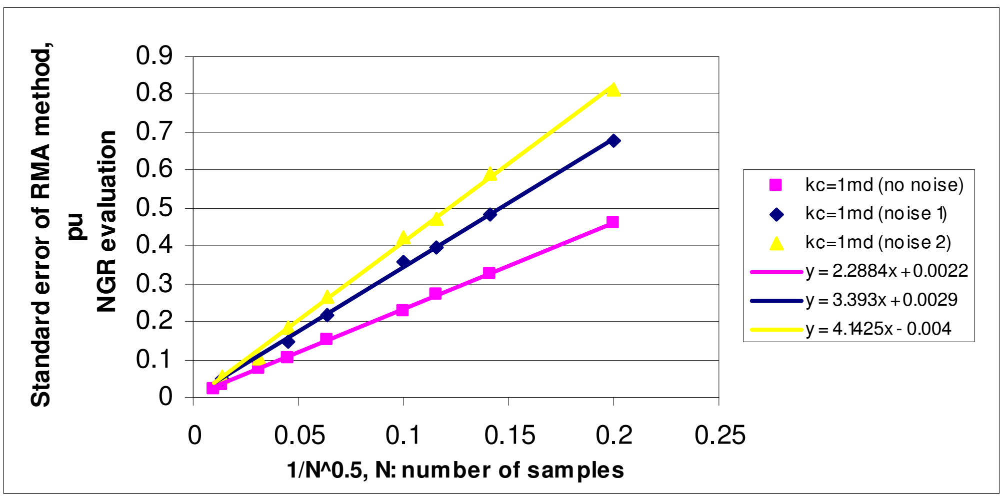


Fig. 5-10 Bias and standard error of estimates given by the RMA method for non-corrupted data, data corrupted with noise of magnitude 1, and data corrupted with noise of magnitude 2. A permeability cut-off value of 1 md is used to derive the porosity cut-off value from the RMA line.

As expected, the estimates given by the RMA line for a permeability cut-off value of 0.1 md is a MVUE, i.e. an estimator with minimum variance and unbiased. The variability of the estimates of porosity cut-off value given by the RMA line are unbiased even though datasets are corrupted with noise 1 and noise 2 as the bias remains lower than 0.1, as shown on Figure 5-11.

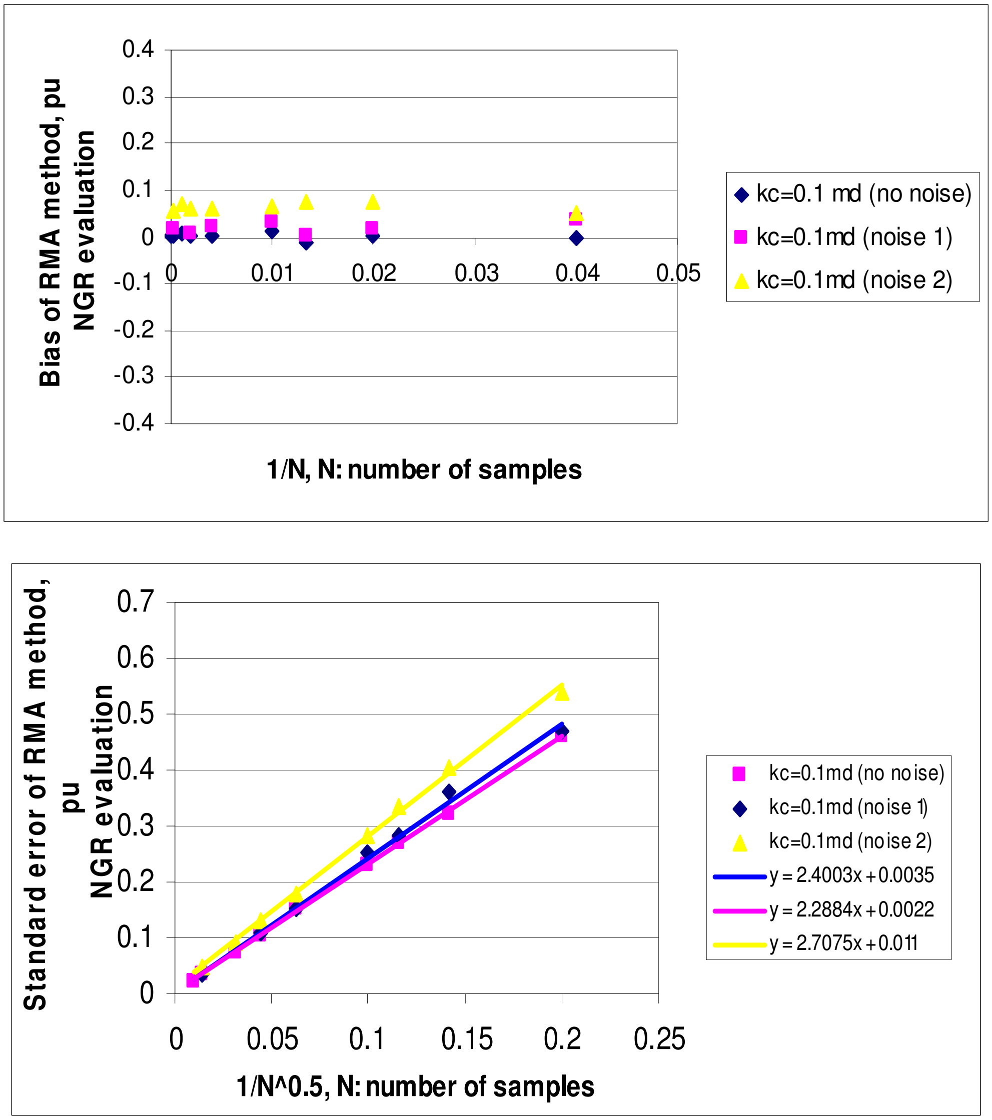
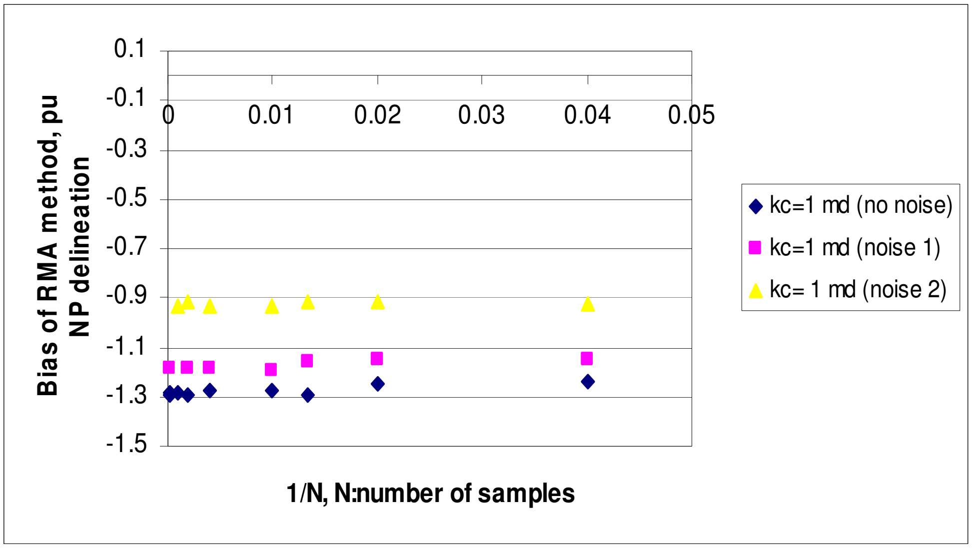


Fig. 5-11 Bias and standard error of estimates given by the RMA method for non-corrupted data, data corrupted with noise of magnitude 1, and data corrupted with noise of magnitude 2. A permeability cut-off value of 0.1 md is used to derive the porosity cut-off value from the RMA line.

The RMA line is, indeed, a scale-independent line meaning that it is uninfluenced by the degree to which the data are spread out. The equation of the RMA line depends on the accuracy of the estimation of the following statistical quantities, averages and standard deviations of the bivariates, and the line is independent of the coefficient of correlation. The latter is greatly influenced by the corruption of the data and the variability of estimates obtained from Y-on-X method is more affected by the corruption of data than those obtained from the RMA line. The linearity of the standard errors and the low bias of the RMA method, in response to corruption of data, illustrate these considerations.

The use of the RMA line to predict porosity cut-off values for delineating NP leads to significantly high biased estimates of porosity cut-off values for a permeability cut-off of 1 md, as shown on Figure 5-12. This systematic bias is due to the fact that the RMA line is not the line which gives analytically the best porosity cut-off value to delineate NP. In the case where the bias may be corrected, it is reasonable to use the RMA line instead of the Y-on-X line since the standard error of its estimates is expected to be twice as low as those from the Y-on-X line. In the case where the permeability cut-off value is a decade away from the permeability average, the bias of the estimator is relatively uninfluenced by the degree of sampling, i.e. the bias remains constant as shown on Figure 5-12.



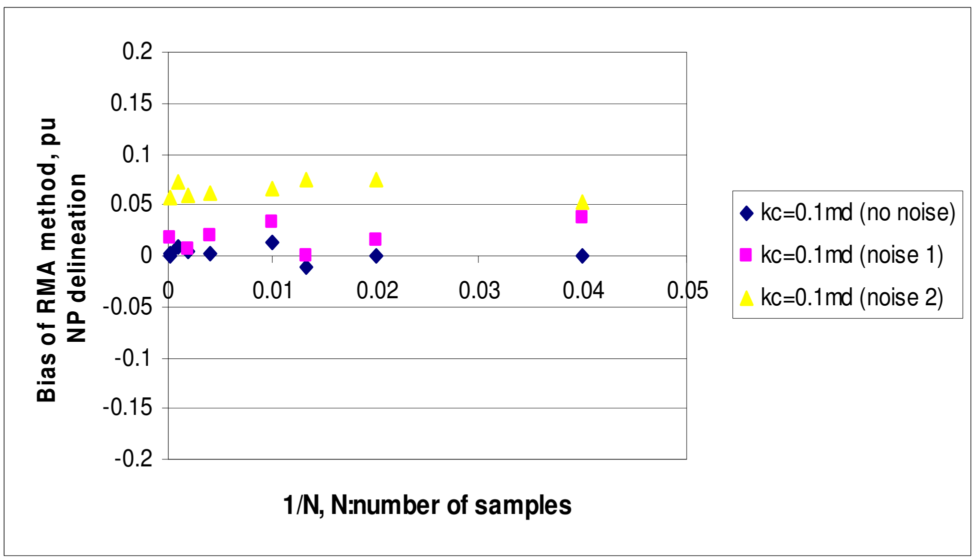


Fig. 5-12 Biases of estimates given by the RMA method for non-corrupted data, data corrupted with noise of magnitude 1, and data corrupted with noise of magnitude 2. Permeability cut-off values of 0.1 md and 1 md are used to derive porosity cut-off values so as to delineate NP.

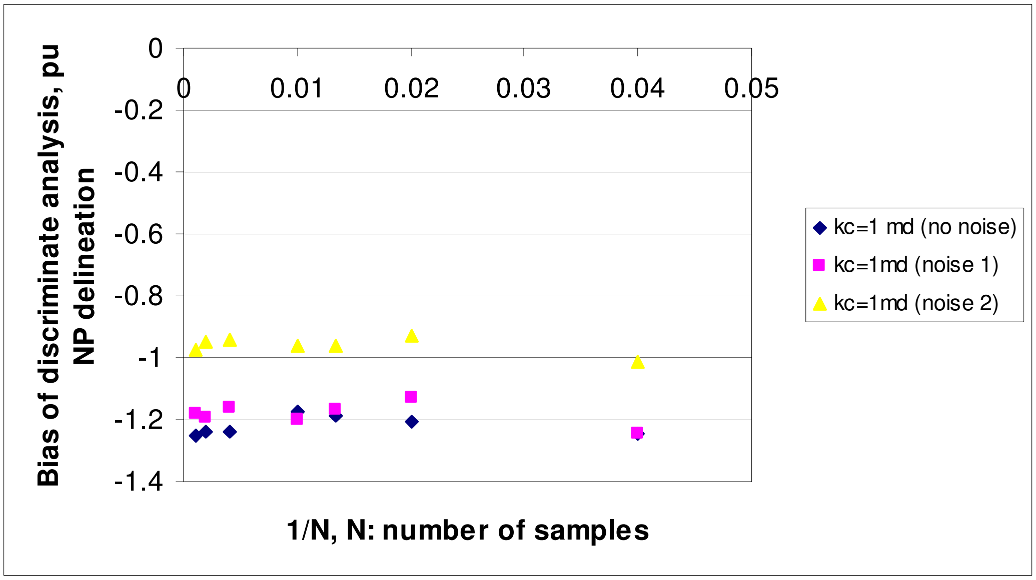
To conclude, the RMA line is therefore really efficient to provide unbiased and accurate estimates of optimal porosity cut-off values to evaluate NGR. Whatever the degree of noise, the number of sample, or the value of the permeability cut-off, the RMA method provides reasonably accurate estimates of porosity cut-off values to evaluate NGR. The analysis of the case where the RMA line is used to delineate NP raises the issue regarding the use of the systematic use of the RMA line to both delineate NP and evaluate NGR. For this purpose, the systematic bias which exists between the Y-on-X and the RMA line, proportional to the coefficient of correlation, should be corrected.

PERFORMANCE OF THE DISCRIMINANT ANALYSIS

The relevancy of using the discriminant analysis in our case and its derived sub datasets is considered by testing the normality of the pay and non-pay fraction of porosities, those having been determined on the basis of the permeability cut-off value. Appendix B presents an example of probability plots for a 100 sample dataset extracted from the joint normal population under consideration for the study with a permeability cut-off of 1 md. The normality of the two fractions of the porosity is not rejected for our case. The classic methodology is used to assess the performance of the discriminant analysis to predict porosity cut-off values for delineating NP and evaluating NGR.

For this purpose, numerous realizations are made for each sampling case: 400 realizations for case where $N < 1000$ and 100 realizations for the case where N is equal to 1000 samples. Fewer realizations are done than for the Y-on-X and RMA lines due to CPU time issues. The case where the permeability cut-off is defined to be 1 md is taken so that bias and standard error of this method will be compared to the other methods.This method is designed to provide estimates of the optimal porosity cut-off value when evaluating NGR as shown on Figure 3-3.

In the case there is no noise, the discriminant analysis gives unbiased (less than 0.1 p.u.) estimates of porosity cut-off values. When the dataset is corrupted with noise, a low bias is created, which does not exceed 0.35 pu in case where the noise is of magnitude 2. This method therefore provides accurate estimates of porosity cut-off values to evaluate NGR. When used to delineate NP, the method provides estimates that are highly biased and tends to underestimate the optimal value of the porosity cut-off value (Figure 5-13). The behavior of the standard error is linear and the intercept of the lines are not significantly different from 0. The performance of the method is relatively good to evaluate NGR in comparison to the RMA line.



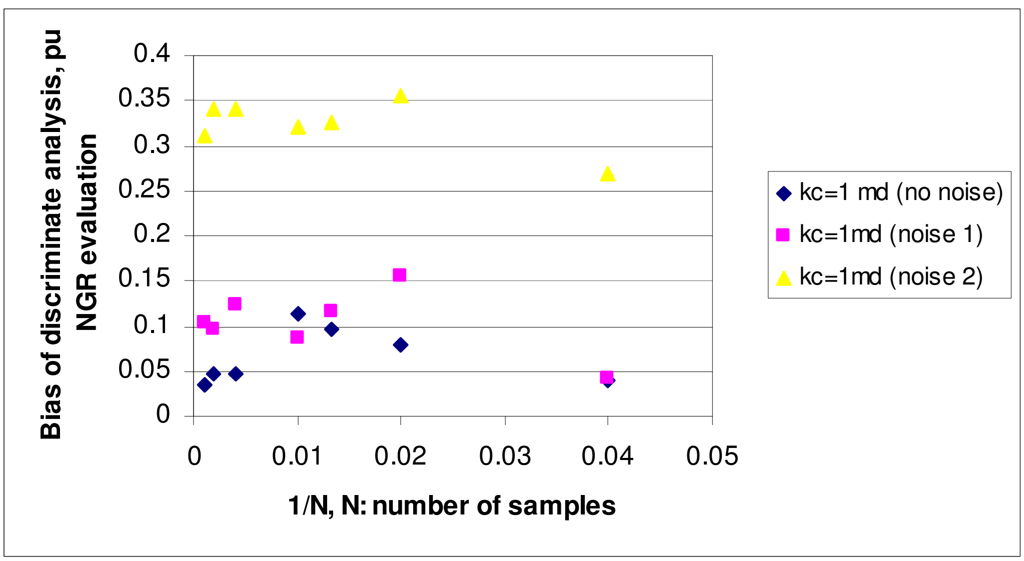


Fig. 5-13 Biases of estimates given by the discriminant analysis for non-corrupted data, data corrupted with noise of magnitude 1, and data corrupted with noise of magnitude 2. A permeability cut-off value of 1 md is used to derive porosity cut-off values so as to delineate NP and evaluate NGR.

The discriminant analysis actually provides a porosity cut-off value which cancels out the likelihood to mistake pay for non-pay and non-pay for pay. The performance of this statistical method is improved when the data are more corrupted. The robustness of the discriminant analysis is good since the introduction of noise 1 induces an increase of 1.5 percent in the standard error and the introduction of noise 2 increases by 19.6 percent the standard error of the estimates with respect to the standard errors of the original case (Figure 5-14).

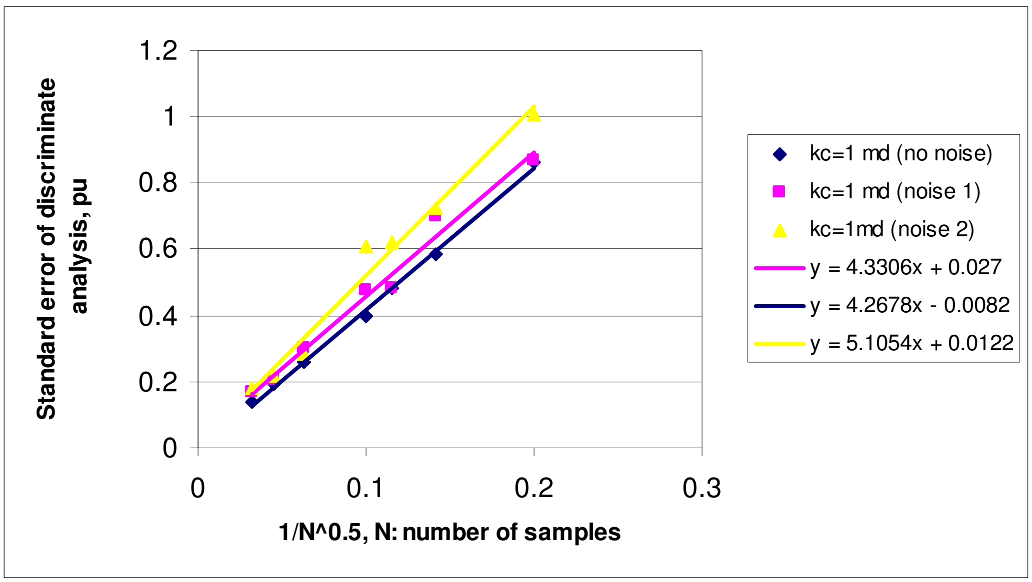


Fig. 5-14 Standard error of estimates given by the discriminant analysis for non-corrupted data, data corrupted with noise of magnitude 1, and data corrupted with noise of magnitude 2. A permeability cut-off value of 1 md is used to derive porosity cut-off values.

The discriminant analysis is therefore clearly robust since the introduction of noise induces a small increase in the variability of estimates in comparison to the previous method, i.e. the Y-on-X and RMA lines.

Discriminant analysis is not expected to be suitable to predict porosity cut-off values for delineating net pay. In fact, the method is rather designed to cancel out the errors of mistaking pay porosities for non-pay porosities and non-pay porosities for pay porosities. This method is a good alternative to evaluate NGR since the overall performance of the estimator is moderately poorer than that of RMA line; assumptions about the distribution of the pay and non-pay fraction are however less restrictive than the joint normality of the log (k)-ϕ.

PERFORMANCE OF THE QUADRANT METHOD

The same procedure is used to assess the performance of the quadrant method to predict porosity cut-off values for delineating NP and evaluating NGR. For this purpose, numerous realizations are made for each sampling case: 400 realizations for case where N < 1000 and 100 realizations for the case where N is equal to 1000 samples (limited number of realizations has been done due to CPU time issues). The case where the permeability cut-off is defined to be 1 md on the basis of geological and engineering considerations is taken.

This methodology is greatly dependent on the data and may be inaccurate, especially for cases where the samples are limited. The efficiency of the quadrant method to predict optimal porosity-cut-off value to delineate NP is poor in the case where noise is introduced. The estimates are moderately biased for non-corrupted data but the variability of the estimator is really significant owing to data-sampling issues (Figure 5-15): the cut-off value is designed to minimize the sum of the probabilities B and C, which is relatively difficult on datasets with few samples (issues related to multiple occurrences of minima).

Considering the non-corrupted case and the case where the quarter method is used to delineate NP, the estimates of the porosity cut-off values are moderately biased but the related standard error is high (for 1000 samples the 68% confidence interval is still +/- 0.5 p.u.): the method is not reliable to provide unbiased, non erroneous estimates for a joint normal log(k)-ϕ dataset. The robustness of the method to predict NP is therefore poor since the introduction of noise leads to an increase in the variability of the estimates (for 100 samples corrupted with noise 2, the confidence interval is increased from +/- 1.25 pu to +/- 1.75 pu and the bias is nearly tripled from slightly less than 0.5

pu to 1.5 pu. This methodology gives the optimal porosity cut-off value for any sub dataset, by minimizing the sum of probabilities B and C. It is obviously leading to the presence of a bias even for a significant number of samples.

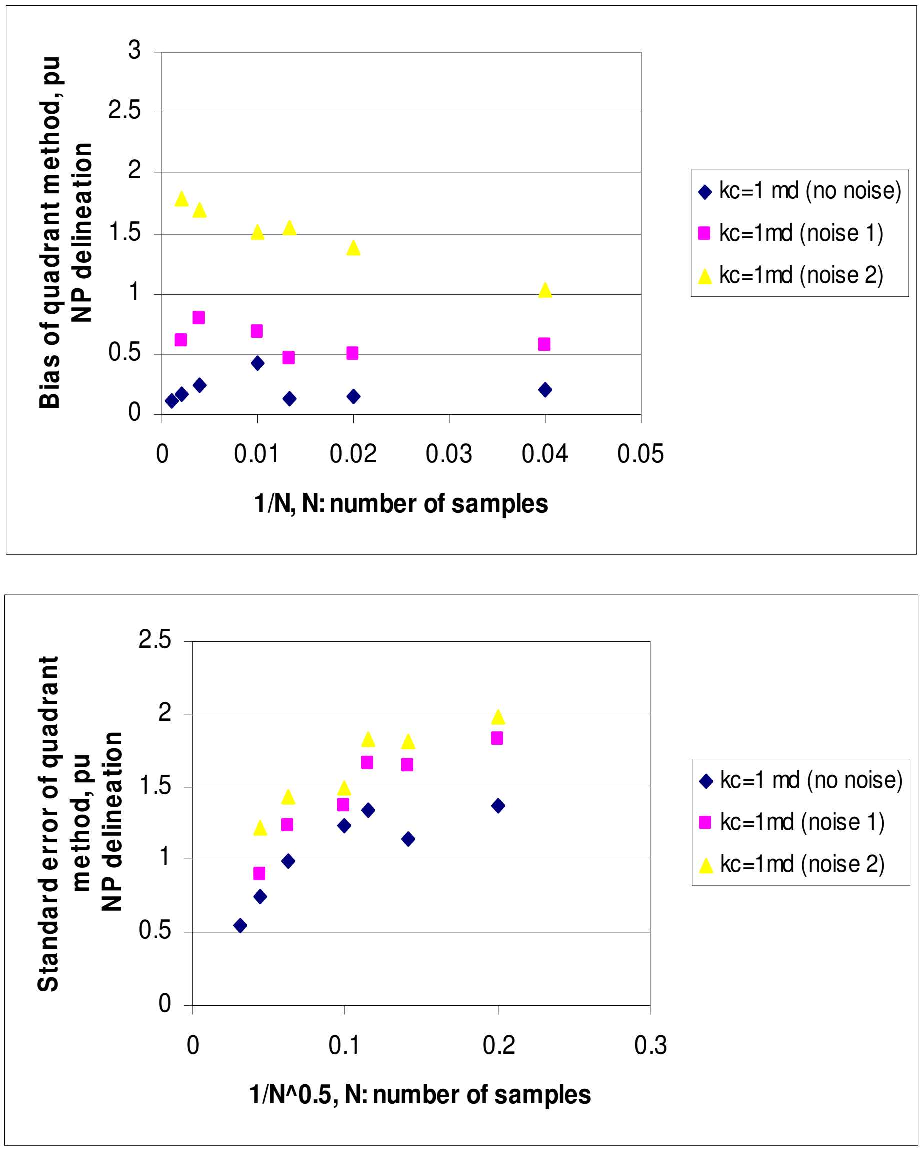


Fig. 5-15 Biases and standard errors of estimates given by the quadrant method for non-corrupted data, data corrupted with noise of magnitude 1, and data corrupted with noise of magnitude 2. A permeability cut-off value of 1 md is used to derive the porosity cut-off value to delineate NP.

The results for the case where the quadrant method is aimed to equalize the probabilities of B and C, i.e. evaluating NGR, are surprisingly different. The estimates of porosity cut-off values so as to evaluate NGR are fairly accurate and robust (Figure 5-16). First, bias of estimates is significantly lower than those when NP is delineated. Second, the standard errors' behavior of the estimates is linear and the intercept are not significantly different from 0.

The estimates obtained from the quadrant method on samples from the non-corrupted population are unbiased and have a moderate variability. In order to decrease the variability by 40 percent for a 25 sample dataset, it requires doubling the number of samples, i.e. to sample 25 additional k-ϕ measurements. The estimator is robust since the introduction of noise induces moderate bias and variability (Figure 5-16).

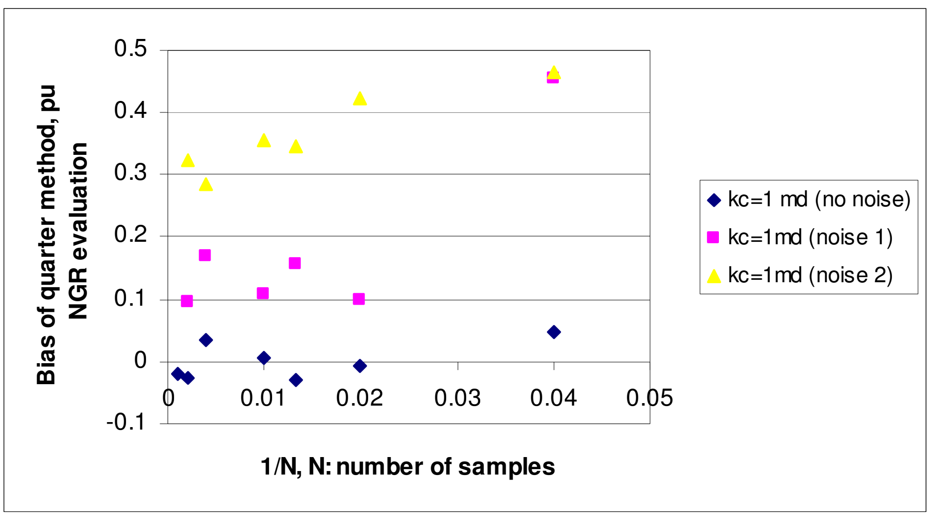


Fig. 5-16 Biases and standard errors of estimates given by the quadrant method for non-corrupted data, data corrupted with noise of magnitude 1, and data corrupted with noise of magnitude 2. A permeability cut-off value of 1 md is used to derive the porosity cut-off value to evaluate NGR.

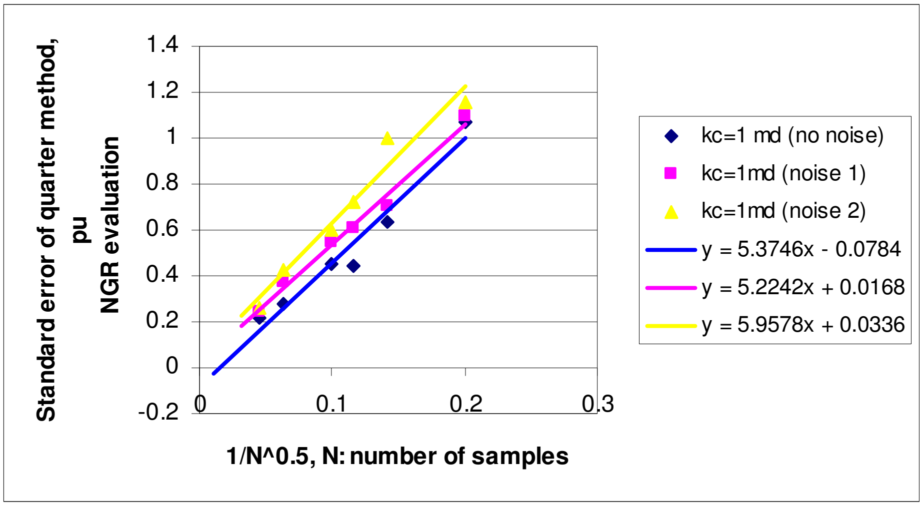


Fig. 5-16 Continued.

The method is therefore significantly more efficient to evaluate NGR than to predict NP. This may be explained by the numerical issues related to minimize the number of points in the two quadrants B and C, as it may occur that a dataset presents several minima.

Their estimates are indeed significantly biased. Even though these estimates have a systematic variability and are non robust, they are from a hypothesis-free method and gives the optimal value for a specific dataset, independently from the statistical characteristics of the original population from which data have been sampled.

COMPARISON OF THE METHODS TO PREDICT NP AND EVALUATE NGR

We consider the case where k_c = 1 md is determined from engineering and geological considerations a decade away from the median permeability, i.e. 0.1 md, and investigate standard error and bias of estimates. The different methods are first compared to one another on the basis of their ability to assess the optimal method to either predict NP or evaluate NGR on the basis of bias, standard error, and robustness. The reference optimal cut-off values are determined from the equations of the Y-on-X and RMA lines, as shown in chapter III for a permeability cut-off value of 1 md.

NP delineation

The RMA method is obviously the method with the lower standard errors whatever the magnitude of noise is. However, this method provides optimal porosity cut-off estimates for evaluating NGR causing an inherent bias if the method is applied to predict porosity cut-off values to delineate NP. In the case where the data are not corrupted, the Y-on-X method provides nearly unbiased estimates of porosity cut-off, ranging between 0 and 0.2 pu (Figure 5-17). The standard error is however higher than those of the discriminant analysis and RMA line. Those two methods are nonetheless designed to provide cut-off values so as to predict NGR: the use of one of them to delineate NP will lead to a systematic, inherent bias, which tends to overestimate NP. The results from the quadrant method show that the estimator is not a good predictor of optimal porosity cut-off values since its bias and standard error are really high in the case where the data are joint normal.

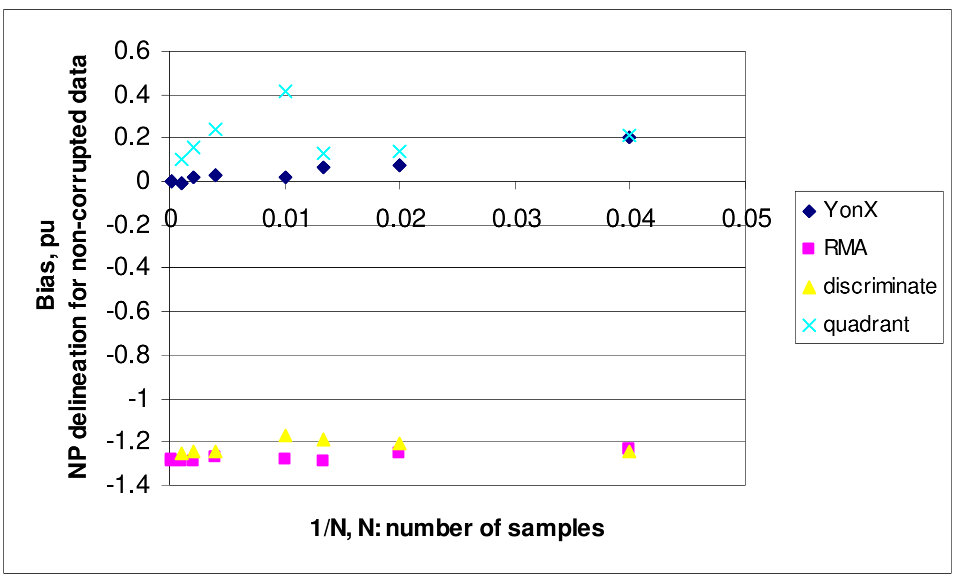


Fig. 5-17 Bias and standard error of estimates given by the methods under investigation for non-corrupted data. A permeability cut-off value of 1 md is used to derive the porosity cut-off value to delineate NP.

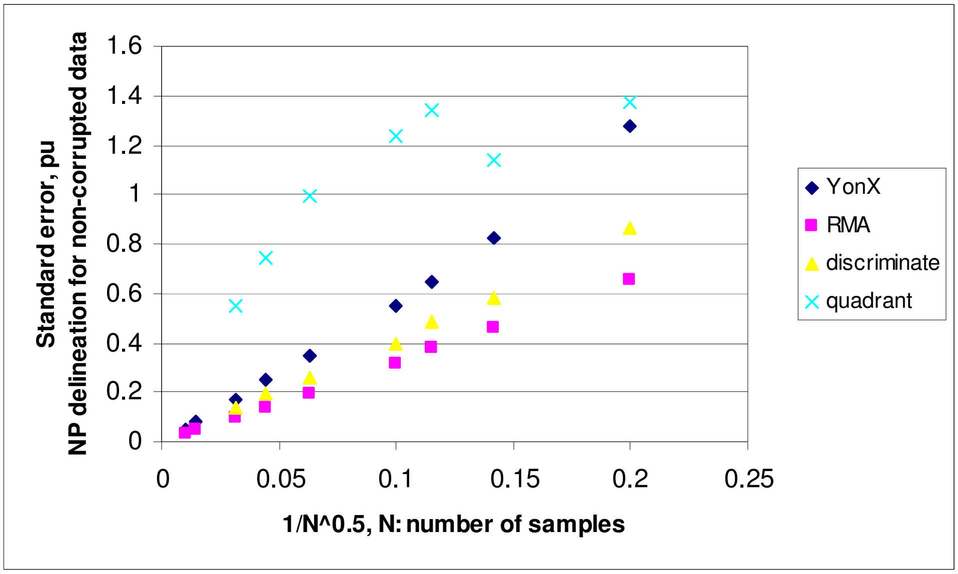
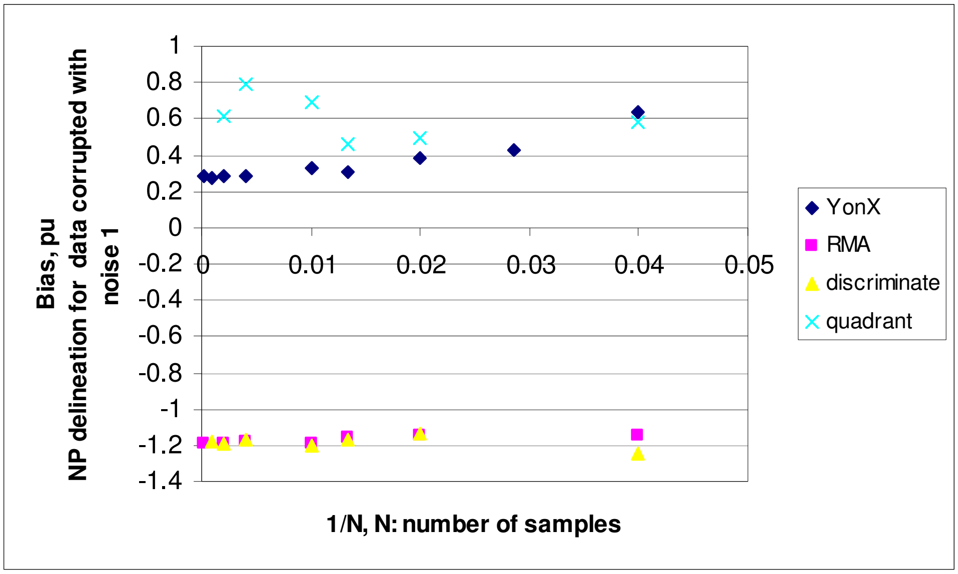


Fig. 5-17 Continued.

The corruption of data with a noise of moderate magnitude does not modify the hierarchy between the different methods. The Y-on-X line still gives porosity cut-off values with a low standard error, i.e. the confidence interval is relatively good, with the lowest bias. The standard errors of the estimates obtained from the RMA line and the discriminant analysis are lower than that of the Y-on-X method but with a significant inherent bias (Figure 5-18).



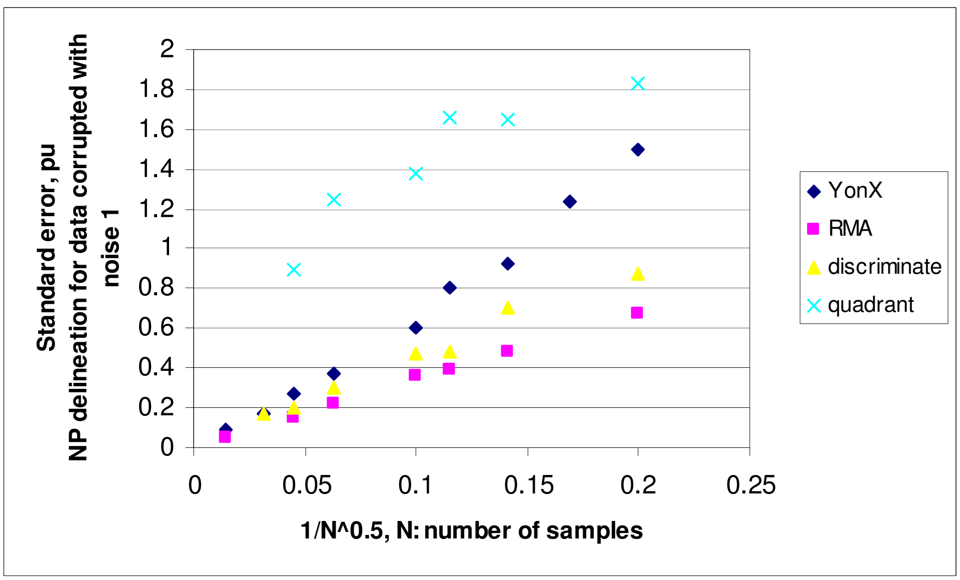
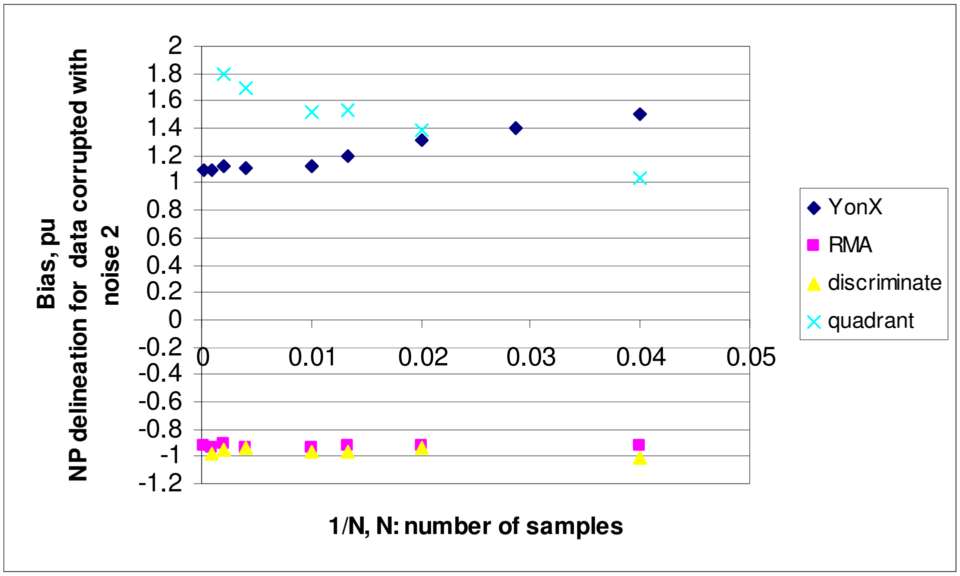


Fig. 5-18 Bias and standard error of estimates given by the methods under investigation for data corrupted with noise of moderate magnitude. A permeability cut-off value of 1 md is used to derive the porosity cut-off value to delineate NP.

When datasets are corrupted with a noise of high magnitude, the Y-on-X method gives estimates of porosity cut-off values whose variability is much higher than the standard error of the RMA line in particular. Since the Y-on-X axis is much more sensitive to the degree to which data are spread out, the bias of the estimates derived from this method will be increased so that the bias of estimates from RMA line are lower than those from the Y-on-X method for dataset with less than 50 points, as shown Figure 5-19.



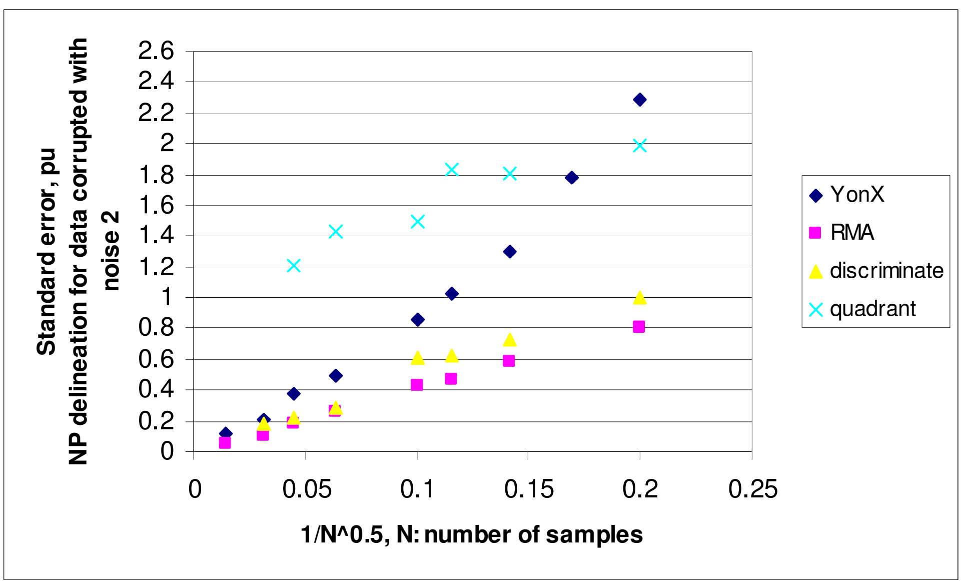


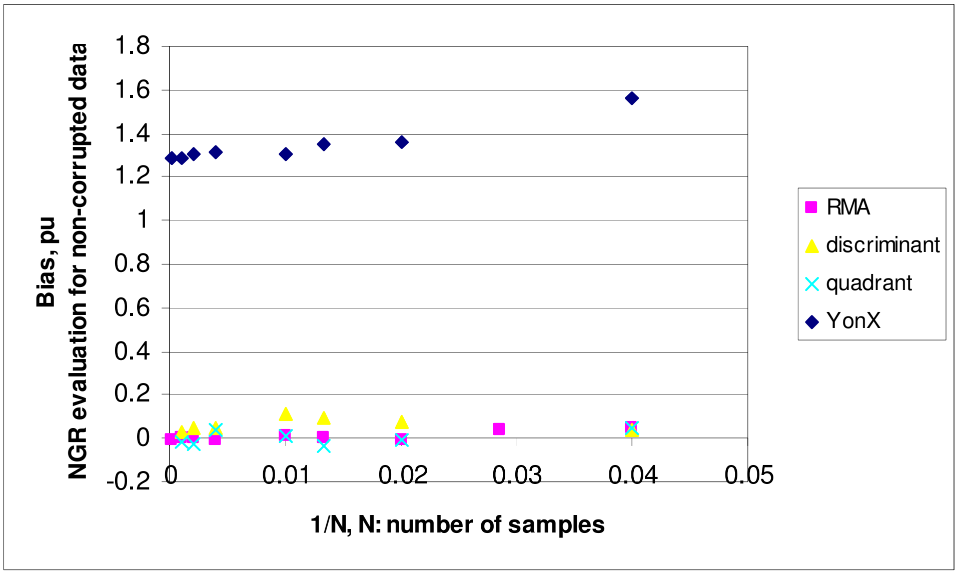
Fig. 5-19 Bias and standard error of estimates given by the methods under investigation for data corrupted with noise of high magnitude. A permeability cut-off value of 1 md is used to derive the porosity cut-off value to delineate NP.

It raises the problem regarding the possibility to correct the inherent bias related to the use of the RMA line to predict NP instead of using the Y-on-X. In the case where it is possible to correct the inherent bias of the RMA line to predict NP, it would be clearly interesting to use the RMA line instead of the Y-on-X line to predict NP. The performance of the RMA line is indeed higher than that of the Y-on-X line since the RMA axis is independent of the coefficient of correlation which is strongly affected by any type of noise or perturbation. The bias of the estimates from the RMA line appear to be constant whatever the number of sample is so that it can be envisioned to try to quantify this bias. When the joint normality is tested and validated, the Y-on-X method remains nonetheless the optimal estimator for predicting porosity cut-off value to delineate NP, despite a significant variability.

NGR delineation

In any case, the Y-on-X method provides biased estimates of porosity cut-off values with poor confidence intervals. The discriminant analysis, the quadrant method and the RMA method provide fairly good estimates of porosity cut-off values so as to evaluate NGR. Bias of their estimates is closed to zero, i.e. that the expectancy of estimates is centred on the optimal porosity cut-off value to evaluate NGR, as shown on Figure 5-20. The RMA line remains the best estimator by far since it shows the lowest standard error. The quadrant method and the discriminant analysis are more sensitive to the corruption of the joint normality than those from the RMA line; bias of these methods approximately equals 0.2 pu for datasets corrupted with noise of moderate magnitude and 0.4 pu for datasets corrupted with noise of high magnitude (Figures 5-21 and 5-22).The estimates from RMA line are nearly unaffected by the corruption of the data.

As the magnitude of noise increases the variability of the discriminant analysis tends to be closer to the variability of the RMA line. It illustrates that the performance of the discriminant analysis is not significantly affected by the introduction of noise and the corruption of the joint normality. When evaluating NGR, three methods are available to predict in a reliable way porosity cut-off value. The RMA line is the method which provides unbiased estimates with small confidence intervals.



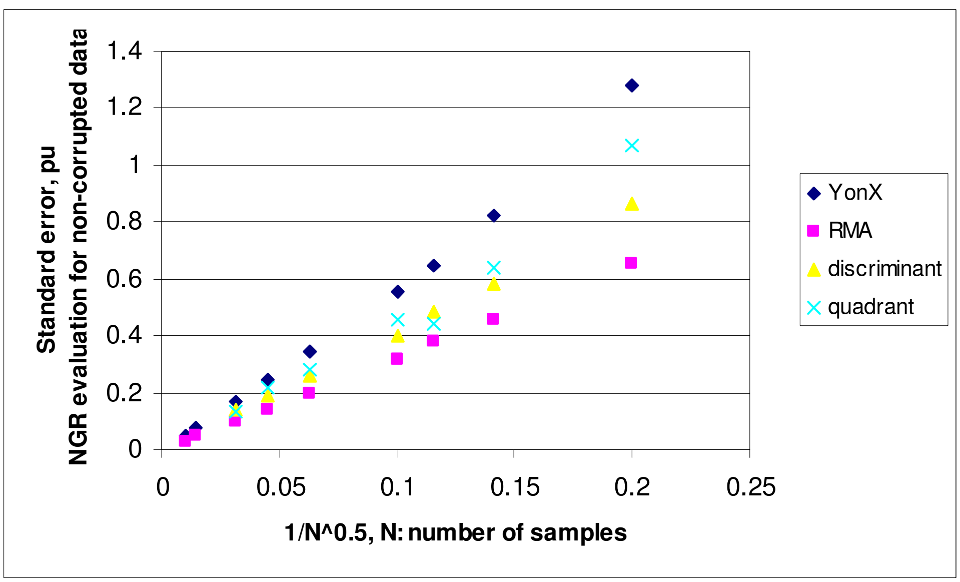
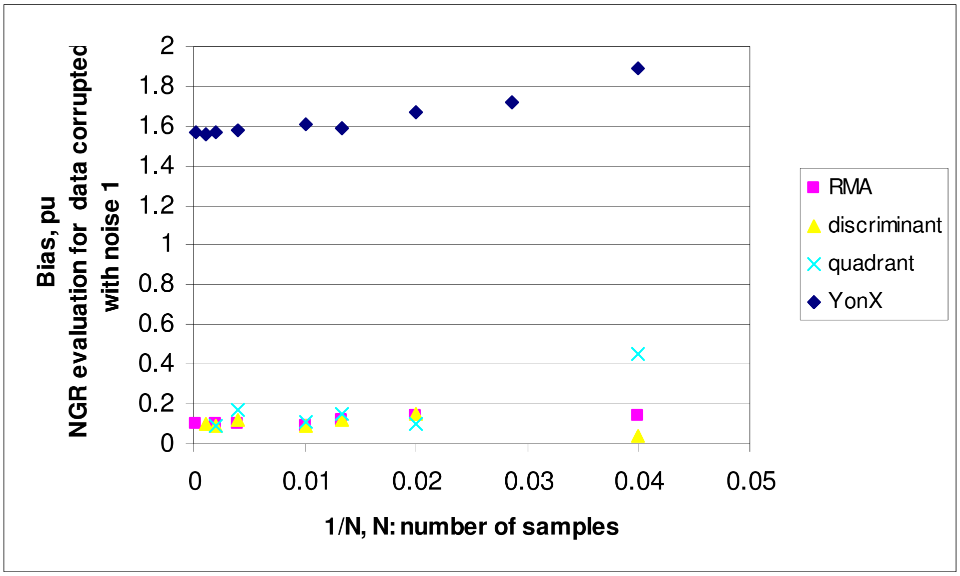


Fig. 5-20 Bias and standard error of estimates given by the methods under investigation for non-corrupted data. A permeability cut-off value of 1 md is used to derive the porosity cut-off value to evaluate NGR.

The issue is that the use of RMA line is dependent of the assumption of joint normality. In the case where the joint normality is too restrictive, the discriminant analysis is an interesting alternative in order to evaluate NGR.



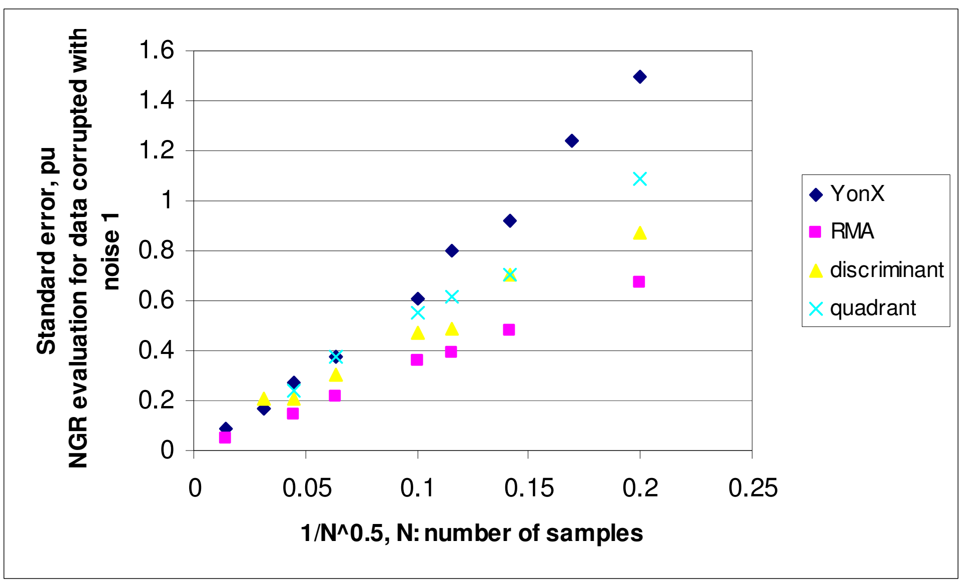
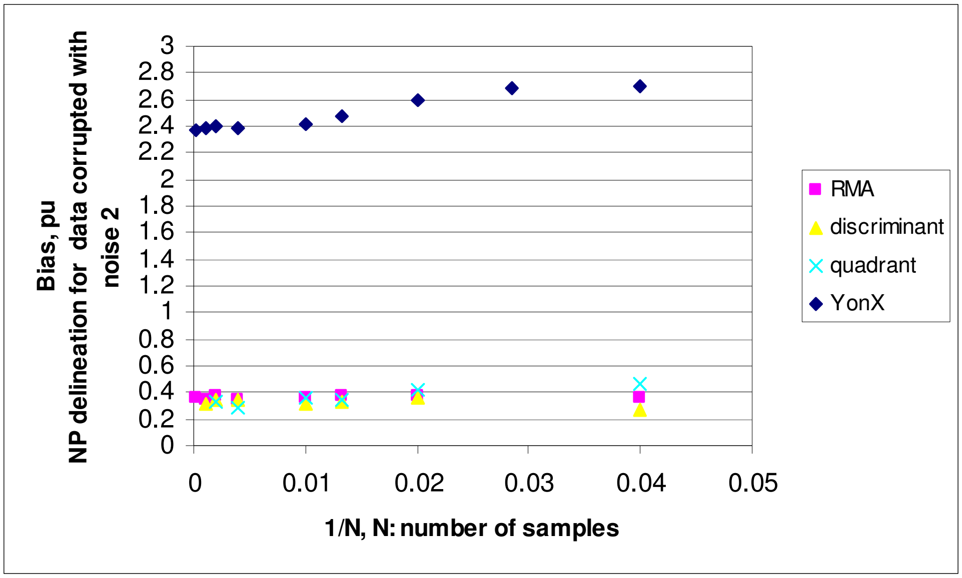


Fig. 5-21 Bias and standard error of estimates given by the methods under investigation for data corrupted with noise of moderate magnitude. A permeability cut-off value of 1 md is used to derive the porosity cut-off value to evaluate NGR.

This comparison has shown that the RMA line is providing the optimal estimates of porosity cut-off values to evaluate NGR, in terms of their bias, standard error, and robustness. These considerations are only valid in the case where log(k)-ϕ are joint normal.



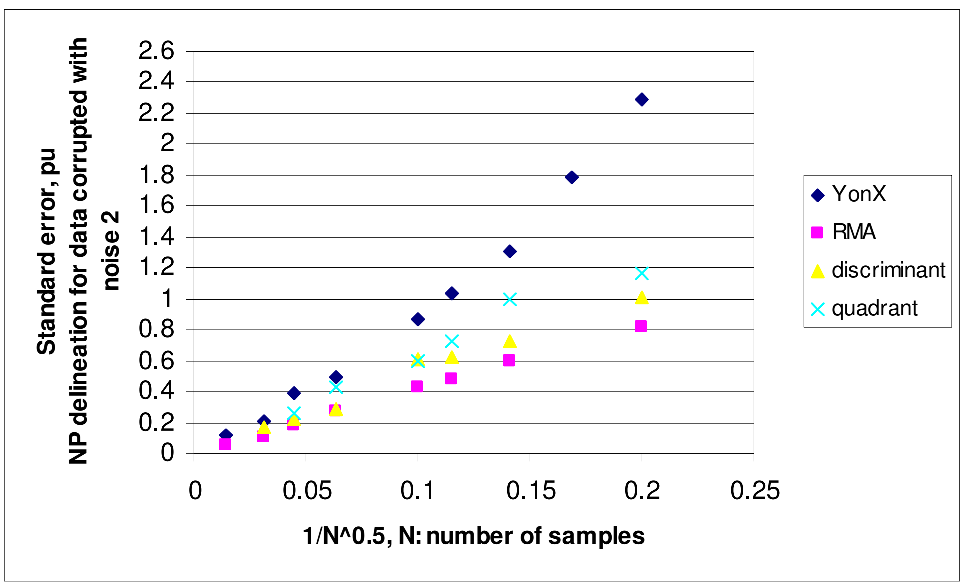


Fig. 5-22 Bias and standard error of estimates given by the methods under investigation for data corrupted with noise of high magnitude. A permeability cut-off value of 1 md is used to derive the porosity cut-off value to evaluate NGR.

The quadrant and the discriminant analysis methods offer two interesting alternatives since the variability of their estimates are slightly higher than that of the RMA but require less restrictive assumptions.

Delineating NP point by point by using a porosity cut-off value on the log-derived reservoir porosities is expected to be more problematic, especially for datasets in which fewer than 75 samples are present. In fact, the Y-on-X line might be in this case expected to be biased and highly variable, leading to erroneous estimation of NP.

Test case for the application of the methods

A core permeability and porosity dataset measured at reservoir pressure is presented from the Travis Peak formation in East Texas (from Luffel et al., 1991, Nelson and Kibler, 2003). The deposits consist of Lower Cretaceous complexes of delta lobes reworked by fluvial and marine processes. One specific facies of the formation, corresponding to fine-grained, moderately to well sorted sandstones (quartzarenites and subarkoses) is used for testing the methods.

The sandstone permeabilities are relatively tight, with an average of 0.03 md (Table 5-1). The porosity average is 7 pu. The variabilities, as shown by the standard deviations, are of the same magnitude as the quantities used in the Monte Carlo analysis of the different methods. The coefficient of correlation of the field data is high, i.e. 0.879, but does not characterize the heterogeneities of density of points in the dataset (Figure 5-23). In fact, data in the lower permeability range, e.g. $k < 0.1$ md, are well correlated and the corresponding density of points is greater than that in the high permeability range. On the other hand, the density of points for $k > 0.1$ md is low and the correlation between porosity and permeability is poor. As a consequence, one may expect to observe significant variations in the performance of the estimators depending on the value of permeability cut-off value.

Table 5-1 Statistical quantities of the Travis Peak core measurements (N=320) and the equations of the Y-on-X regression and RMA lines.

	porosity, pu	log (k), md
mean	7.028	-1.495
std deviation	3.033	1.5157
rho	0.8786	
Y-on-X	$\log(k) = 0.439\ \phi - 4.5809$	
RMA	$\log(k) = 0.4997\ \phi - 5.0073$	

For this purpose one will investigate cases where three different permeability cut-off values are defined, 1, 0.1, and 0.01 md, so as to illustrate the impact of this value on the NP and NGR evaluation process. The case where the permeability cut-off value equals the log-permeability average is not investigated in this analysis since the RMA and the Y-on-X lines intersect at this point. It is recommended in this case to use the RMA line both to delineate NP and evaluate NGR by using one unique porosity cut-off value.

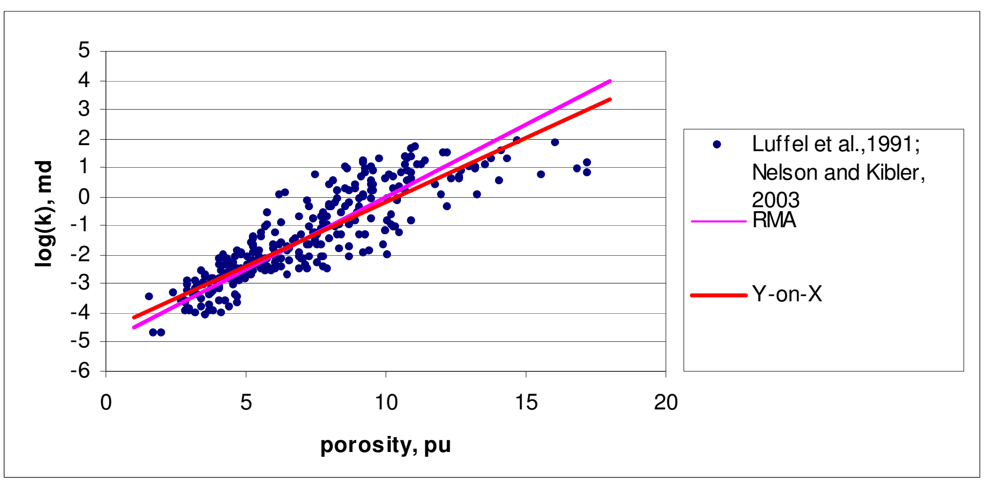


Fig. 5-23 Log permeability vs. porosity cross plot from Travis Peak formation core measurements (data from Luffel et al., 1991, Nelson and Kibler, 2003).

The second objective of this analysis is to compare the results of analyzing this dataset with those from the Monte Carlo analysis. From the comparison, we aim to make recommendations so that anyone faced with producing optimal cut-offs from a dataset will have a guide.

First, a test of the joint normality of the Travis Peak dataset (the details are presented in Appendix E) shows that the hypothesis about the joint normality of the dataset is not rejected. The field data are thus assumed to be approximately joint normally distributed. Second, a test for the normality of the pay and non-pay fractions of porosities (Appendix B) shows that whatever the permeability cut-off value is, the normality of the sub fraction of porosities is not rejected, with the exception of the case where k_c = 0.01 md. In this case, the pay fraction is not normal since the probability plot exhibits a significant deviation from linearity. These results suggest that the Travis Peak dataset and the datasets used from the Monte Carlo assessments have similar distributions. Therefore, we expect that the RMA line will be the most efficient method to evaluate the cut-off for NGR, or at least as efficient as the discriminant analysis and the quadrant method. The Y-on-X and the quadrant methods should provide unbiased estimates of porosity cut-off values for delineating NP.

The first component of the analysis of the dataset is to evaluate the performance of estimates of porosity cut-off value given by the Y-on-X regression line and the quadrant method in order to delineate NP. Unambiguous absolute minima are obtained for this dataset for the quadrant method with all three permeability cut-off values (Figure 5-24). As the dataset is assumed to be joint normally distributed, the porosity cut-off values given by the discriminant analysis and the RMA line are expected to be significantly biased so those methods are not considered.

One obtains two different porosity cut-off values, respectively derived from the quadrant method and the Y-on-X line for the different permeability cut-off values (Table 5-2). As expected, the cut-off values estimated by the quadrant method and the Y-on-X line differ by only a few percent and are not significantly different. The observed variations are caused by the imperfect discretization of the PDF of the formation by the sampling as observed in the model (Figure 5-17).

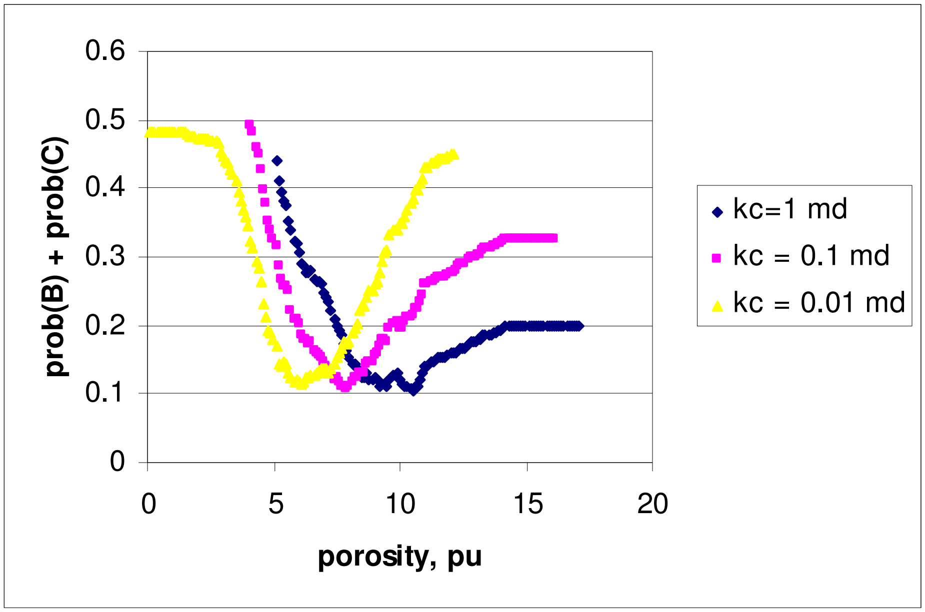


Fig. 5-24 Example results for the quadrant method for the cases where $k_c = 1$ md, $k_c = 0.1$ md, and $k_c = 0.01$ md. The curves exhibit mimima for $\phi_c = 6.1$ p.u., $\phi_c = 7.9$ p.u., and $\phi_c = 10.5$ p.u..

Table 5-2 Porosity cut-off values given by the Y-on-X line and the quadrant method. The total probabilities of making misidentification of non-pay for pay and pay for non-pay are calculated, i.e. estimating the values of p(B)+p(C).

k_c= 0.01 md	cut-off, pu	% error
Y-on-X	5.89	12.5
quadrant	6.1	11.25

k_c= 0.1 md	cut-off, pu	% error
Y-on-X	8.16	12.18
quadrant	7.9	10.31

Table 5-2 Continued.

k_c= 1 md	cut-off, pu	% error
Y-on-X	10.43	10.6
quadrant	10.5	10.31

For this dataset, we do not know the true population porosity cut-off value for NP identification. Nonetheless, one can estimate confidence intervals for the values listed in Table 5-2 using the Monte Carlo analyses described earlier. For instance, for the case where k_c = 1 md, Figure 5-17 gives the standard error of estimates from the Y-on-X line and the quadrant method for N = 320 samples to be

ϕ_c (quadrant) = 10.5 ± 0.9 p.u.

ϕ_c (Y-on-X) = 10.43 ± 0.3 p.u.

Obviously, the errors in both mistaking pay for non-pay and non-pay for pay are the same for both methods but the confidence interval is smaller for estimates given by the Y-on-X regression line. As a consequence, it is obviously recommended to use the Y-on-X line to delineate NP in this case (Table 5-2).

A similar procedure can be applied to obtain porosity cut-off values for NGR estimation. The Y-on-X line is discarded since it has been demonstrated that it provides biased estimates with large intervals of confidence when used to evaluate NGR. The discriminant analysis, the RMA line and the quadrant methods are thus considered for predicting NGR. An actual NGR is computed from the permeability cut-off values for comparison with NGR values estimated from the porosity cut-off values given by the RMA line, the discriminant analysis, and the quadrant methods (Table 5-3).

Table 5-3 Predicted NGR and errors of prediction determined from porosity cut-off values given by the methods under consideration.

	kc= 0.01 md	kc= 0.1 md	kc= 1 md
Actual NGR, percent	51.8	34.68	22.19

kc= 0.01 md	cut-off, pu	predicted NGR	% error
RMA	6.02	52.8	1.93
Discriminant	6.20	50.62	-2.28
Quadrant	6.1	51.25	-1.06

kc= 0.1 md	cut-off, pu	predicted NGR	% error
RMA	8.02	34.06	-1.79
Discriminant	7.8	36.56	5.42
Quadrant	8.05	34.06	-1.79

kc= 1 md	cut-off, pu	predicted NGR	% error
RMA	10.02	17.18	-22.58
Discriminant	9.11	25	12.66
Quadrant	9.3	22.5	1.40

For the two first cases, where the permeability cut-off values respectively equal 0.01 and 0.1 md, the errors in predicting NGR are not significantly different from one another. For the case k_c = 0.01 md, errors in predicting NGR by using the discriminant analysis are slightly higher than those by using the RMA line and quadrant methods. The use of the discriminant analysis in spite of the non-normality of the pay fraction of porosities may have caused additional errors. These results confirm the results from the Monte

Carlo testing: for datasets with N = 300 samples, the three estimators are providing unbiased estimates of porosity cut-off with small confidence intervals. It is however suggested to use the RMA line since it has been demonstrated that the estimates from the RMA line presents the lowest variability.

For the case where k_c = 1 md, the RMA line does not, surprisingly, provide the best prediction for NGR (Table 5-3). It can be explained by issues related to the degree of sampling (higher density of points in the low permeability domain than in the high permeability domain) or that the joint normality is altered in the high permeability domain. The performance of the discriminant analysis is significantly better than that of the RMA line but remains lower than that of the quadrant method. The standard error of the estimators, i.e. the 68 % confidence interval, is used from Figure 5-20 so as to evaluate the uncertainty in the NGR evaluation for the case k_c = 1 md:

ϕ_c (quadrant) = 9.3 ± 0.275 p.u.

ϕ_c (RMA) = 10.02 ± 0.18 p.u.

ϕ_c (discriminant) = 9.11 ± 0.25 p.u.

The confidence intervals of estimates of porosity cut-off values seem to be relatively small and insignificant. However, the small variability of the estimates given by the three methods induces a high variability in predicted NGR values. In fact, by considering the 68 % confidence interval of porosity cut-off values from the RMA line, it gives NGR values ranging from 15.31 to 18.5 percent. In the same way, the range of predicted NGR values by using the discriminant analysis is between 21.87 and 27.5 percent. The 68 % confidence interval for the NGR values predicted from the quadrant method gives a range between 18.7 and 26.3 percent. All these values indicate that the 95% confidence interval (± 2 σ) would include an NGR of 22.19.

In the same way, the evaluation of the reliability of the NGR value obtained directly from the permeability cut-off value can also be investigated. A common procedure usually involves the use of the “leave-one-out” method, or jack-kniffing (Jensen et al., 2003, p. 111). It consists of successively removing one different sample from the original dataset with N samples so as to obtain N subdatasets with N-1 samples.

This procedure is not used in our analysis since the removal of one point from the Travis Peak dataset (N = 320 samples) will only induce an insignificant change in the NGR values: for instance, for the case k_c = 1 md, the removal of one sample from either pay or non-pay fraction of data will modify the NGR values by about 0.3 %. As an alternative approach, the confidence intervals for the NGR values determined from the permeability cut-off values, 1, 0.1, and 0.01 md, and shown in Table 5-3 are evaluated heuristically by setting the error of the permeability cut-offs to ± 10 %. For the case where k_c = 1 md ± 0.1, for example, the predicted NGR values vary from 21.56 to 22.5 percent, i.e. an absolute maximum variation of 3 percent with respect to the evaluated NGR value of 22.19. For the case where k_c = 0.1 md, it gives a range of predicted NGR values from 33.125 to 35.31 percent, i.e. an absolute maximum variation of 4.5 percent with respect to the evaluated NGR value of 34.68 percent. Finally, for the case where k_c = 0.01 md, it gives a range of predicted NGR values from 50.62 to 52.187 percent, i.e. a absolute maximum variation of 2.3 percent with respect to the evaluated NGR value of 51.8 percent. Therefore, whatever the permeability cut-off value is, the variations in NGR values due to uncertainties are similar. Also, one can observe that for the case where k_c = 1 md, the variation in the NGR values evaluated from the porosity cut-off values is larger than the variation in the NGR values evaluated from the permeability cut-off values. This may be explained by the small density of points having permeabilities higher than 0.1 md, leading to the significant perturbation of the joint normality and the presence of outliers. Obviously the quadrant method should be preferred in the case where k_c = 1 md since the estimates of NGR given by the RMA line and the discriminant analysis have larger errors. The normality of the pay fraction for these permeability cut-off values may be also altered because of the lack of data, which induces an increase in the variability of the estimates. In fact, the probability plot presenting the pay fraction of porosities for k_c = 1 md in Appendix B presents slight deviation from the normality for the bigger porosities.

In conclusion, Monte Carlo analysis of hypothetical datasets and analysis of the Travis Peak dataset suggests the following procedures for determining appropriate porosity cut-off value:

For NGR estimation, the RMA line and the discriminant analysis methods perform best. The values of predicted NGR given by those two methods should be compared to those given by the quadrant to verify the efficiency of the prediction and evaluate the possible limitations of the application of the methods. If no assumption can be made about the joint normality of the field dataset or the distribution of the pay and non-pay porosities, one has to resort to use the quadrant method. The Y-on-X method should be systematically discarded when the NGR is evaluated. The efficiency of the RMA line and the discriminant analysis methods may be significantly reduced in the case where the permeability cut-off is located in a crossplot area where the density of points is small and the influence of outliers is significant. In this case, the uncertainties in NGR values determined from the porosity cut-off (inherent to the methods used for obtaining the porosity cut-off values from the permeability cut-off values) may be significantly higher than uncertainties in NGR values determined from the permeability cut-off values. In this case, it is recommended to use the quadrant method in the case where it gives a better estimate for NGR.

When delineating NP, both Y-on-X line and the quadrant may be used. The former method, however, is preferred since it presents smaller variability of its estimates of porosity cut-off values than the latter method.

CHAPTER VI

DISCUSSION AND RECOMMENDATIONS

The assumption of joint normality of log (k) and ϕ remains a good starting point to assess the performance of the estimators which are to be evaluated. The choice of the joint normality is both a theoretical convenience and a plausible distribution for log(k) and porosity.

From the observations made by Jensen and Menke (2006), confirmed by analytical calculus in Chapter II, the joint normality of bivariate dataset is convenient so as to know the optimal porosity cut-off values to delineate NP and evaluate NGR. These properties are convenient for estimating the performance of the methods and to assess, in a quantitative and qualitative way, the bias, standard errors and robustness of the estimates. In the case where NGR is evaluated, the study has shown that the estimates of the porosity cut-off values from the RMA line are only slightly affected in terms of their bias and standard errors by noise and the corruption of the joint normality. The use of RMA line and/or the discriminant analysis is therefore recommended to predict porosity cut-off values even if the joint normality is not clearly demonstrated. The estimates of their porosity cut-off values are relatively accurate, slightly biased and with low variability. The data-oriented quadrant method should be used in last resort because of its significant bias and higher variability, e.g. 10 % errors.

In the case where the intended use of the porosity cut-off value is to delineate NP point by point, the Y-on-X line gives the optimal porosity cut-off value for joint normal log (k) $-$ ϕ. The main issue is that the estimator is expected to provide estimates with high variability, especially for datasets in which few points are present. It has also been demonstrated that the estimator is not robust since the introduction of significant noise, so that the joint normality is no longer present, leads to an increase in variability and the introduction of an irreducible and inherent significant bias of 1.1 p.u., when datasets are perturbed with noise 2. It leads to systematic errors in the prediction of the NP and it may be interesting to use the quadrant method rather than the Y-on-X line. The Y-on-X method nonetheless remains a convenient and efficient method to provide porosity cut-off values so as to delineate NP especially when the assumption of joint normality is not rejected. Bias is insignificant in the noise-free case and the standard error is smaller than +/- 1.3 p.u, i.e. +/- 8% of error, which is moderate. The RMA line and the discriminant analysis are designed to evaluate NGR: it is thus not adequate to use them directly to delineate NP because of the introduction of a significant bias.

Despite its biased estimators with a poor interval of confidence, it is particularly interesting to consider the quadrant method when the assumption of joint normality is rejected.

When delineating NP, the correction of inherent bias of estimates given by the RMA line should be taken into consideration since this axis is scale invariant and less sensitive to the errors in measurements or deviations from the normality. The idea would be to use the RMA line to delineate NP and correct the systematic bias since the standard errors are always smaller than those of estimates given by the Y-on-X line.

The value of the permeability cut-off value with respect to the permeability average of the data is also an important parameter to take into account, so as to evaluate whether the two lines, RMA and Y-on-X are significantly different from each other and the consequences on the derived porosity cut-off values, as shown with the application of the methods on field data. Field datasets may present heterogeneous densities of points. In the domain where the density of points is small, the performance of the discriminant analysis, RMA line, and the Y-on-X regression line may deteriorate because of the presence of outliers. It may result in the choice of the quadrant method to evaluate NGR in the case where the performance of the method is better than that of the RMA line and discriminant analysis methods.

CHAPTER VII

FURTHER WORK

The study and the comparison of the estimators for porosity cut-off values have been conducted by using joint normal datasets. This specific assumption is clearly restrictive but it remains a good start to investigate the performance of various methods under consideration for the study, in terms of their bias, standard error, and robustness. Other bivariate distributions should be investigated for comparison and ensure that the results and the qualitative classification of the methods are still valid and not significantly different. Joint normal bivariate (k) ^{p}- ϕ datasets may be used for this purpose, $|p| > 0$. The case where $p \rightarrow 0$ was investigated in this study. The optimal porosity cut-off values should be analytically determined by using derivative calculus on the PDF of the bivariate population under consideration.

The tentative correction of the inherent bias of estimates from the RMA line when delineating NP should be therefore investigated. The systematic bias seems to be in a certain extent invariant whatever the number of samples is. This bias might be assessed and quantified by using the expressions of the two equations, RMA and Y-on-X lines.

Regarding the discriminant analysis, a new equation may be considered by attempting to minimize the sum of the errors of misidentification of pay for non-pay and non-pay for pay, still assuming the normality of the two sub fraction of porosities. The sum of areas A and B (Figure 3-3) may be minimized with respect to the porosity cut-off, by conducting a derivative calculus.

The case where only one surrogate variable was investigated: the study and the considerations may be extended to multi-variate dataset so as to determine optimal V_{shc}, S_{wc}, or R_{tc}.

CHAPTER VIII

CONCLUSIONS

The study has shown that the systematic use of the ordinary least-squares regression for selecting porosity cut-off values from a permeability cut-off may lead to erroneous values and does not guarantee optimal predictions of NP and NGR. The use of a simple and interesting statistical bivariate distribution has led to the comparison of various methods in addition to the classic least-squares regression. As confirmed by Jensen and Menke (2006), this regression line does systematically predict the optimal porosity cut-off values: the use of another line, the Major Reduced Axis, presents interesting potential to derive porosity cut-off values because of the statistical properties of the axis.

The study provides a new vision the problem of selecting porosity cut-off values, or in a larger extent, of any surrogate variable, to delineate NP and evaluate NGR. It also suggests recommendation so as to predict accurately and efficiently porosity cut-off values so as to evaluate NP and NGR.

Here are several tables summarizing the results, i.e. bias and standard errors of the estimates of porosity cut-off values, given by the different methods (Table 8-1).

Table 8-1 Bias and standard error of estimates given by the different methods under consideration in the study for noise-free and noised datasets.

NP delineation	N k_c=1 md	25	50	100	500	1000
No noise Bias	Y-on-X	0.204	0.075	0.023	0.015	-0.0038
	RMA	-1.237	-1.249	-1.2798	-1.293	-1.288
	Discriminant	-1.246	-1.208	-1.17	-1.2388	-1.252
	Quadrant	0.209	0.14	0.419	0.16	0.1023
No noise Standard error	Y-on-X	1.278	0.82	0.55	0.247	0.17
	RMA	0.65	0.46	0.32	0.14	0.096
	Discriminant	0.86	0.58	0.399	0.19	0.139
	Quadrant	1.37	1.137	1.239	0.74	0.55

NP delineation	N k_c=1 md	25	50	100	500	1000
Noise 1 Bias	Y-on-X	0.64	0.38	0.32	0.28	0.27
	RMA	-1.15	-1.15	-1.19	-1.19	-1.19
	Discriminant	-1.24	-1.13	-1.2	-1.19	-1.18
	Quadrant	0.58	0.50	0.69	0.75	0.55
Noise 1 Standard error	Y-on-X	1.49	0.92	0.60	0.27	0.17
	RMA	0.68	0.48	0.36	0.15	-
	Discriminant	0.87	0.70	0.48	0.20	0.17
	Quadrant	1.83	1.64	1.38	0.89	-

Table 8-1 Continued.

NP delineation	N k_c=1 md	25	50	100	500	1000
Noise 2 Bias	Y-on-X	1.5	1.31	1.13	1.12	1.10
	RMA	-0.93	-0.92	-0.93	-0.91	-0.93
	Discriminant	-1.01	-0.93	-0.96	-0.94	-0.97
	Quadrant	1.03	1.38	1.51	1.69	1.79
Noise 2 Standard error	Y-on-X	2.29	1.30	0.86	0.38	0.21
	RMA	0.81	0.59	0.43	0.18	0.11
	Discriminant	1.00	0.726	0.61	0.22	0.18
	Quadrant	1.98	1.81	1.50	1.21	-

NGR evaluation	N k_c=1 md	25	50	100	500	1000
No noise Bias	Y-on-X	1.56	1.36	1.31	1.30	1.28
	RMA	0.05	0.036	0.006	-0.007	-0.0027
	Discriminant	0.04	0.08	0.11	0.047	0.034
	Quadrant	0.05	-0.0061	0.007	-0.25	-0.02
No noise Standard error	Y-on-X	1.278	0.82	0.55	0.247	0.17
	RMA	0.65	0.46	0.32	0.14	0.096
	Discriminant	0.86	0.58	0.399	0.19	0.139
	Quadrant	1.07	0.64	0.45	0.22	0.14

Table 8-1 Continued.

NGR evaluation	N k_c=1 md	25	50	100	500	1000
Noise 1 Bias	Y-on-X	1.89	1.67	1.61	1.57	1.56
	RMA	0.14	0.14	0.09	0.1	-
	Discriminant	0.04	0.15	0.086	0.095	0.139
	Quadrant	0.45	0.1	0.11	0.095	-
Noise 1 Standard error	Y-on-X	1.49	0.92	0.60	0.27	0.17
	RMA	0.68	0.48	0.36	0.15	-
	Discriminant	0.87	0.70	0.48	0.20	0.17
	Quadrant	1.09	0.70	0.55	0.24	-

NGR evaluation	N k_c=1 md	25	50	100	500	1000
Noise 2 Bias	Y-on-X	2.7	2.6	2.4	2.41	2.38
	RMA	0.36	0.37	0.36	0.15	-
	Discriminant	0.27	0.35	0.48	0.205	0.206
	Quadrant	0.46	0.42	0.35	0.32	-
Noise 2 Standard error	Y-on-X	2.29	1.30	0.86	0.38	0.21
	RMA	0.81	0.59	0.43	0.18	0.11
	Discriminant	1.00	0.726	0.61	0.22	0.18
	Quadrant	1.16	1.0	0.6	0.26	-

REFERENCES

Agterberg, F.P., 1974, *Geomathematics*, Elsevier Scientific Publishing Co., Amsterdam, p. 119-123.

Cobb, W.M. and Marek, F.J., 1998, Net pay determination and waterflood depletion mechanisms, SPE 48952, *in* the Annual Technical Conference and Exhibition, 14p.

Coker, D.A. and Lindquist, W.B., 1994, Structural reliability analysis for one-dimensional, two phase miscible flow, *SIAM Journal of Applied Mathematics*, vol. 54, no. 4, p. 935-953.

Davis, J.C., 2002, Statistics and data analysis in geology, John Wiley & Sons, New York, 638p.

Dutton, S.P., Flanders, W.A. and Barton, M.D., 2003, Reservoir characterization of a Permian deep-water sandstone, east Ford field, Delaware basin, Texas, *American Association of Petroleum Geologists Bulletin*, vol. 87, no. 4, p. 609-627.

Egbele, E. and Ezuka, I., 2005, Net-to-gross ratios: implications in integrated reservoir management studies, SPE 98808, *in* the Annual Technical Conference and Exhibition, 13p.

George, C.J., Stiles, L.H., 1978, Improved techniques for evaluating carbonate waterfloods in West Texas, *Journal of Petroleum Technology*, vol. 30, no. 6, p. 1547-1554.

Haro, C.F., 2004, The perfect permeability transform using logs and cores, SPE 89516, *in* the Annual Technical Conference and Exhibition, 17p.

Jensen, J.L., Lake, W.L., Corbett, P.W.M., Goggin, D.J., 2003, Statistics for petroleum engineers and geoscientists, Elsevier, Amsterdam, 338p.

Jensen, J. L. and Menke, J. Y., 2006, Some statistical issues in selecting porosity cut-offs for estimating net pay, *Petrophysics*, vol. 47, no. 4, p. 315-320.

Keener, M. H. L., 1967, Computer coordinated core and log interpolation, Trans., *in* SPWLA 6th Annual logging Symposium, Denver, A1-7.

Kolodzie, S. Jr., 1980, Analysis of pore throat size and use of the Waxman-Smits equation to determine OOIP in Spindle Field, Colorado, SPE 9382, *in* the Annual Fall Technical Conference, 10p.

Krzanowski, W. J., 1988, Principles of multivariate analysis, Oxford University Press, Oxford, 586p.

Li, D., and Dria, M.A., 1997, Cut-off or separation? A new approach to distinguish reservoir rock from non-reservoir for reservoir modeling and simulation, SPE 38380, *in* the Rocky Regional Meeting, 9p.

Lucia, F.J., 1999, Carbonate reservoir characterization, Heidelberg, Springer-Verlag, 226p.

Luffel, D.L., Howard, W.E., and Hunt, E.R., 1991, Travis Peak core permeability and porosity relationships at reservoir stress, *Society of Petroleum Engineers Formation Evaluation,* vol. 6, no. 3, p. 310-318.

McKenzie,W.T., 1975, Petrophysical study of the Cardium Sand in the Pembina field, SPE 5541, *in* the Annual Technical Conference and Exhibition, 12p.

Morton-Thompson, D. and Woods, A.M., 1993, AAPG methods in exploration series, No. 10, Development Geology Reference Manual, Tulsa, Oklahoma, 565p.

Nelson, Philip H. and Kibler, Joyce E., 2003, A catalog of porosity and permeability from core plugs in siliciclastic rocks, Open file Report 03-420.
http://pubs.usgs.gov/of/2003/ofr-03-420/ofr-03-420.html

Pirson, S. J., 1958, Oil reservoir engineering, McGraw-Hill, New York City, 735p.

Snyder, R.H., 1971, A review of the concepts and methodology of determining 'net pay', SPE 3609, *in* the Annual Technical Conference and Exhibition, 12p.

Spearing M., Allen, T., McAulay G., 2001, Review of the Winland R35 method for net pay definition and its application in low permeability sands, paper *in* the Special Core Analysis (SCA) Symposium, 7p.

Teti, M.J. and Krug, J.A., 1987, Log analysis methods and empirical data derivation of net pay parameters for carbonate reservoirs in the eastern Montana part of the Williston Basin, *The Log Analyst*, vol. 28, no. 3, p. 259.

Vavra, C.L., Kaldi, J.G., and Sneider, R.M., 1992, Geological applications of capillary pressure: a review, *American Association of Petroleum Geologists Bulletin*, vol. 76, no. 6, p. 840.

Worthington, P.F. and Consentino, L., 2005, The role of cut-offs in integrated reservoir studies: *SPE Reservoir Evaluation and Engineering*, vol. 8, no. 4, p. 276-290.

NOMENCLATURE

NP : net pay

NGR : net-to-gross ratio

ϕ : porosity expressed in porosity units, p.u.

k : permeability expressed in md.

ϕ_c : porosity cut-off value, expressed in p.u.

k_c : permeability cut-off value, expressed in md.

ϕ_{cRMA} : porosity cut-off value given by the Reduced Major Axis line.

$\phi_{cY\text{-}on\text{-}X}$: porosity cut-off value given by the Y-on-X line.

ϕ_{cNP} : optimal porosity cut-off value to delinate net pay.

ϕ_{cNGR} : optimal porosity cut-off value to evaluate net-to-gross ratio.

log(k) : Base 10 logarithm of permeability in md.

p.u. : porosity unit, expressed in percent.

s_{NP} : standard error of the non-pay fraction of porosities.

s_P : standard error of the pay fraction of porosities.

$\overline{\phi_{NP}}$ μ_{NP} : mean of the non-pay fraction of porosities.

$\overline{\phi_P}$: mean of the pay fraction of porosities.

APPENDIX A

RELATIONSHIP OF REGRESSION AND RMA LINES TO ERROR BEHAVIOR

Assuming that log (k) and ϕ are assumed to have a joint normal distribution, their corresponding probability density function (PDF) is defined by equation (A-1):

$$f(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left[-\frac{x^{*2}-2\rho x^* y^* + y^{*2}}{2\left(1-\rho^2\right)}\right] \tag{A-1}$$

where $x^* = \dfrac{x-\mu_X}{\sigma_X}$, $y^* = \dfrac{y-\mu_X}{\sigma_Y}$, and $\rho = \dfrac{\sigma_{XY}}{\sigma_X\sigma_Y}$.

On the basis of the probabilistic framework shown in Figure 2-3, Chapter II, the probabilities of respectively misidentifying pay for non-pay (represented by region B) and non-pay for pay (represented by region C), i.e. prob (B) and prob (C), may be expressed as follows:

$$prob\,(B) = prob\,(\log k \ge \log k_c \;\; and \;\; \phi \le \phi_c)$$

$$prob\,(C) = prob\,(\log k \le \log k_c \;\; and \;\; \phi \ge \phi_c)$$

From equations (A-1), (A-2), and (A-3), $prob\,(B)$ and $prob\,(C)$ might be expressed as follows:

$$prob(B) = \int_{\log kc}^{+\infty}\int_{-\infty}^{\phi_C}\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left[-\frac{x^{*2}-2\rho x^* y^* + y^{*2}}{2\left(1-\rho^2\right)}\right]dxdy \tag{A-4}$$

$$prob(C) = \int_{-\infty}^{\log kc}\int_{\phi_c}^{+\infty}\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left[-\frac{x^{*2}-2\rho x^* y^* + y^{*2}}{2\left(1-\rho^2\right)}\right]dxdy \tag{A-5}$$

where $x^* = \dfrac{x-\mu_\phi}{\sigma_\phi}$, $y^* = \dfrac{y-\mu_{\log(k)}}{\sigma_{\log(k)}}$, and $\rho = \dfrac{\sigma_{\phi.\log(k)}}{\sigma_\phi\sigma_{\log(k)}}$.

The following change of variables simplifies equations (A-4) and (A-5).

$$x^* = \frac{x-\mu_\phi}{\sigma_\phi} \text{ hence } dx^* = \frac{1}{\sigma_\phi}dx. \tag{A-6}$$

$$y^* = \frac{y-\mu_{\log(k)}}{\sigma_{\log(k)}} \text{ hence } dy^* = \frac{1}{\sigma_{\log(k)}}dy. \tag{A-7}$$

It gives:

$$prob(B) = \int_{\log kc^*}^{+\infty}\int_{-\infty}^{\phi_c^*} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{x^{*2}-2\rho x^* y^* + y^{*2}}{2(1-\rho^2)}\right] dx^* dy^* \tag{A-8}$$

$$prob(C) = \int_{-\infty}^{\log kc^*}\int_{\phi_c^*}^{+\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{x^{*2}-2\rho x^* y^* + y^{*2}}{2(1-\rho^2)}\right] dx^* dy^* \tag{A-9}$$

where $\phi_c{}^* = \dfrac{\phi_c - \mu_\phi}{\sigma_\phi}$ and $\log k_c{}^* = \dfrac{\log k_c - \mu_{\log(k)}}{\sigma_{\log(k)}}$.

The first part of derivative analysis is aimed to verify that the optimal porosity cut-off value for equalizing $prob(B) = prob(C)$ is given by the RMA line. The RMA line, expressed by equation (A-10), gives for $\phi^* = \phi_c^*$ in the standardized log(k*)-ϕ^* crossplot:

$$\log k_C^* = \phi_C^*. \tag{A-10}$$

Substituting by equation (A-10) into equations (A-8) and (A-9) it gives:

$$prob(C) = \int_{-\infty}^{\log kc^*}\int_{\log kc^*}^{+\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{x^{*2}-2\rho x^* y^* + y^{*2}}{2(1-\rho^2)}\right] dx^* dy^*, \tag{A-11}$$

$$prob(B) = \int_{-\infty}^{\log kc^*}\int_{\log kc^*}^{+\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{x^{*2}-2\rho x^* y^* + y^{*2}}{2(1-\rho^2)}\right] dx^* dy^*. \tag{A-12}$$

It gives $prob(B) = prob(C)$.

The porosity cut-off value derived from the RMA line is therefore the optimal value $\phi_c=\phi_{cR}$ for NGR estimation. ϕ_{cR} equalizes $prob(B)$ and $prob(C)$ and cancels out the errors of mistaking pay as non-pay and non-pay as pay. In the case where the assumption about the joint normality of the porosity-log permeability dataset applies, the use of the

RMA line is recommended to evaluate the NGR cut-off.

Next we verify that the porosity cut-off value derived from the Y-on-X regression line minimizes the sum of the errors of misidentification of pay and non-pay $prob(B) + prob(C)$, in the case where the porosity-log permeability dataset is assumed to be joint normal distributed.

On the basis of the properties related to the joint normality, the Y-on-X regression line gives for $\phi^* = \phi_c^*$ in the standardized log(k*)-ϕ^* crossplot

$$\log k_C^* = \rho . \phi_C^* . \qquad \text{(A-13)}$$

where ρ is the coefficient of correlation of the variables $\log(k)$ and ϕ.

In order to demonstrate that $prob(B) + prob(C)$ is minimized for the porosity cut-off value derived from the Y-on-X regression line, i.e. $\phi_c = \phi_{cYX}$, a derivative analysis confirms that $prob(B) + prob(C)$ admits a minimum for $\phi_c = \phi_{cYX}$. For this purpose it should be demonstrated that:

(i) The derivative of $prob(B) + prob(C)$ with respect to ϕ_c is equal to 0 for $\phi_c = \phi_{cYX}$

(ii) The second derivative of $prob(B) + prob(C)$ with respect to ϕ_c is positive for $\phi_c = \phi_{cYX}$

From equations (A-8) and (A-9), we obtain the derivatives of $prob(B)$ and $prob(C)$ with respect to ϕ_c are therefore expressed as follows:

$$\frac{\partial}{\partial \phi_c^*} prob(B) = \frac{\partial}{\partial \phi_c^*} \left(\int_{-\infty}^{\log kc^*} \int_{\phi_C^*}^{+\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[- \frac{x^{*2} - 2\rho x^* y^* + y^{*2}}{2(1-\rho^2)} \right] dx^* dy^* \right)$$

Considering $F(x,y) = \int_a^x \int_c^y f(u,v)\, dudv$, it gives:

$$\frac{\partial}{\partial x}\left(F(x,y)\right) = \frac{\partial}{\partial x}\left(\int_a^x \int_c^y f(u,v) dudv \right),$$

$$\frac{\partial}{\partial x}\left(F(x,y)\right)=\int_c^y\left(\frac{\partial}{\partial x}\int_a^x f(u,v)du\right)dv\ ,$$

$$\frac{\partial}{\partial x}F(x,y)=\int_c^y f(x,v)dv\ . \tag{A-16}$$

$$\frac{\partial}{\partial y}\left(F(x,y)\right)=\frac{\partial}{\partial y}\left(\int_a^x\int_c^y f(u,v)dudv\right),$$

$$\frac{\partial}{\partial y}\left(F(x,y)\right)=\int_c^x\left(\frac{\partial}{\partial y}\int_a^y f(u,v)dv\right)du\ ,$$

$$\frac{\partial}{\partial y}F(x,y)=\int_a^x f(u,y)dv\ . \tag{A-17}$$

$$\frac{\partial^2}{\partial x\partial y}F(x,y)=f(x,y)=\frac{\partial^2}{\partial y\partial x}F(x,y) \tag{A-18}$$

The probabilities $prob(B)$ and $prob(C)$ may be expressed as follows, where f represents the PDF of the Log (k)-ϕ joint normal distribution:

$$prob(B)=\int_{-\infty}^{\phi_C^*}\int_{\log kc^*}^{+\infty} f(u^*,v^*)du^*dv^*=\int_{\phi_{\min}^*}^{\phi_C^*}\int_{\log kc^*}^{\log k_{\max}^*} f(u^*,v^*)du^*dv^* \tag{A-19}$$

$$prob(C)=\int_{\phi_C^*}^{+\infty}\int_{-\infty}^{\log kc^*} f(u^*,v^*)du^*dv^*=\int_{\phi_C^*}^{+\infty}\int_{\log k_{\min}{}^*}^{\log kc^*} f(u^*,v^*)du^*dv^*\ . \tag{A-20}$$

Using equation (A-16), equation (A-20) becomes

$$\frac{\partial}{\partial\phi_c^*}prob(B)=-\int_{\log k_{\max}^*}^{\log kc^*} f(\phi_c^*,v^*)dv^*=\int_{\log kc^*}^{\log k_{\max}^*} f(\phi_c^*,v^*)dv^*\ . \tag{A-21}$$

Developing equation (A-21) gives

$$\frac{\partial}{\partial\phi_c^*}prob(B)=\int_{\log kc^*}^{\log k_{\max}^*}\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left[-\frac{\phi_c^{*2}-2\rho.\phi_c^*\ v^*+v^{*2}}{2\left(1-\rho^2\right)}\right]dv^* \tag{A-22}$$

$$\frac{\partial}{\partial\phi_c^*}prob(B)=\int_{\log kc^*}^{\log k_{\max}^*}\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left[-\frac{\phi_c^{*2}-\rho^2.\phi_c^{*2}+\rho^2.\phi_c^{*2}-2\rho.\phi_c^*\ v^*+v^{*2}}{2\left(1-\rho^2\right)}\right]dv^*$$

$$\frac{\partial}{\partial\phi_c^*}prob(B)=\int_{\log kc^*}^{\log k_{\max}^*}\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left[-\frac{\phi_c^{*2}-\rho^2.\phi_c^{*2}}{2(1-\rho^2)}\right]\exp\left[-\frac{\rho^2.\phi_c^{*2}-2\rho.\phi_c^*v^*+v^{*2}}{2(1-\rho^2)}\right]dv^*$$

$$\frac{\partial}{\partial\phi_c^*}prob(B)=\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left[\frac{-\phi_c^{*2}+\rho^2.\phi_c^{*2}}{2(1-\rho^2)}\right].\int_{\log kc^*}^{\log k_{\max}^*}\exp\left[-\frac{(v^{*2}-\rho\ .\phi_c^*)^2}{2(1-\rho^2)}\right]dv^*$$

$$\frac{\partial}{\partial\phi_c^*}prob(B)=\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left[\frac{-\phi_c^{*2}+\rho^2.\phi_c^{*2}}{2(1-\rho^2)}\right].\int_{\log kc^*}^{\log k_{\max}^*}\exp\left[-\frac{1}{2}\frac{(v^*-\rho\ .\phi_c^*)^2}{\left(\sqrt{(1-\rho^2)}\right)^2}\right]dv^* \quad \text{(A-23)}$$

A new change of variable is conducted on equation (A-23):

$$X=\frac{(v^*\ -\rho\ .\phi_c^*)}{\sqrt{(1-\rho^2)}}\ \text{hence}\ dX=\frac{dv^*}{\sqrt{(1-\rho^2)}}. \quad \text{(A-24)}$$

The limits of the integral become:

$$v^*=\log k_c^*\ \text{hence}\ X=\frac{(\log k_c^*-\rho\ .\phi_c^*)}{\sqrt{(1-\rho^2)}}, \quad \text{(A-25)}$$

$$v^*=\log k_{\max}^*\ \text{hence}\ X=\frac{\log k_{\max}^*-\rho\ .\phi_c^*}{\sqrt{(1-\rho^2)}}. \quad \text{(A-26)}$$

Using the equations (A-24), (A-25), and (A-26), equation (A-23) becomes

$$\frac{\partial}{\partial\phi_c^*}prob(B)=\frac{1}{\sqrt{2\pi}}\exp\left[\frac{-\phi_c^{*2}}{2}\right].\int_\alpha^\beta\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{X^2}{2}\right]dX \quad \text{(A-27)}$$

where $\alpha=\dfrac{(\log k_c^*-\rho\ .\phi_c^*)}{\sqrt{(1-\rho^2)}}$

$$\beta=\frac{\log k_{\max}^*-\rho\ .\phi_c^*}{\sqrt{(1-\rho^2)}}.$$

By setting $\log k_{c\max}\longrightarrow+\infty$, equation (A-27) becomes

$$\frac{\partial}{\partial \phi_c^*} prob(B) = \frac{1}{\sqrt{2\pi}} \exp\left[\frac{-\phi_c^{*2}}{2}\right] . \int_{\alpha}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{X^2}{2}\right] dX \tag{A-28}$$

Recalling the expression of $prob(C)$,

$$prob(C) = \int_{\log k_{\min}^*}^{\log kc^*} \int_{\phi_c^*}^{\phi_{\max}^*} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{u^{*2} - 2\rho u^* v^* + v^{*2}}{2\left(1-\rho^2\right)}\right] du^* dv^*$$

The expression of $prob(C)$ may be slightly modified, as follows:

$$prob(C) = -\int_{\log k_{\min}^*}^{\log kc^*} \int_{\phi_{\max}^*}^{\phi_c^*} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{u^{*2} - 2\rho u^* v^* + v^{*2}}{2\left(1-\rho^2\right)}\right] du^* dv^* \tag{A-29}$$

Rearranging equation (A-29),

$$\frac{\partial}{\partial \phi_c^*} prob(C) = -\frac{\partial}{\partial \phi_c^*} \int_{\phi_{\max}^*}^{\phi_c^*} \int_{\log k_{\min}^*}^{\log k_c^*} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{u^{*2} - 2\rho u^* v^* + v^{*2}}{2\left(1-\rho^2\right)}\right] du^* dv^* \tag{A-30}$$

On the basis on equation (A-17), equation (A-30) may be expressed as follows:

$$\frac{\partial}{\partial \phi_c^*} prob(C) = -\int_{\log k_{\min}^*}^{\log k_c^*} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{\phi_c^{*2} - 2\rho.\phi_c^* \; v^* + v^{*2}}{2\left(1-\rho^2\right)}\right] dv^* \tag{A-31}$$

$$\frac{\partial}{\partial \phi_c^*} prob(C) = -\int_{\log k_{\min}^*}^{\log k_c^*} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{\phi_c^{*2} - \rho^2.\phi_c^{*2} + \rho^2.\phi_c^{*2} - 2\rho.\phi_c^* \; v^* + v^{*2}}{2\left(1-\rho^2\right)}\right] dv^* \tag{A-32}$$

Developing equation (A-32), it gives

$$\frac{\partial}{\partial \phi_c^*} prob(C) = -\frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[\frac{-\phi_c^{*2} + \rho^2.\phi_c^{*2}}{2\left(1-\rho^2\right)}\right] . \int_{\log k_{\min}^*}^{\log k_c^*} \exp\left[-\frac{\left(v^{*2} - \rho \; .\phi_c^*\right)^2}{2\left(1-\rho^2\right)}\right] dv^* \tag{A-33}$$

$$\frac{\partial}{\partial \phi_c^*} prob(C) = -\frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[\frac{-\phi_c^{*2} + \rho^2.\phi_c^{*2}}{2\left(1-\rho^2\right)}\right] . \int_{\log k_{\min}^*}^{\log k_c^*} \exp\left[-\frac{1}{2} \frac{\left(v^* - \rho \; .\phi_c^*\right)^2}{\left(\sqrt{\left(1-\rho^2\right)}\right)^2}\right] dv^* \tag{A-34}$$

The same change of variable is applied to the integral of equation (A-34).

Let consider for this purpose a new variable X so as:

$$X = \frac{\left(v^* \;- \rho\;\; .\phi_c^*\right)}{\sqrt{\left(1-\rho^2\right)}} \text{ hence } dX = \frac{dv^*}{\sqrt{\left(1-\rho^2\right)}}. \tag{A-35}$$

The limits of the integral become

$$v^* = \log k_c^* \text{ hence } X = \frac{\left(\log k_c^* - \rho\;\; .\phi_c^*\right)}{\sqrt{\left(1-\rho^2\right)}}, \tag{A-36}$$

$$v^* = \log k_{\min}^* \text{ hence } X = \frac{\log k_{\min}^* - \rho\;\; .\phi_c^*}{\sqrt{\left(1-\rho^2\right)}}. \tag{A-37}$$

Using the equations (A-35), (A-36), and (A-37), equation (A-34) becomes

$$\frac{\partial}{\partial \phi_c^*} prob(C) = -\frac{1}{\sqrt{2\pi}} \exp\left[\frac{-\phi_{cM}^{*2}}{2}\right] . \int_\chi^\alpha \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{X^2}{2}\right] dX$$

where $\alpha = \dfrac{\left(\log k_c^* - \rho\;\; .\phi_c^*\right)}{\sqrt{\left(1-\rho^2\right)}}$

$$\chi = \frac{\log k_{\min}^* - \rho\;\; .\phi_c^*}{\sqrt{\left(1-\rho^2\right)}}.$$

By setting $\log k_{\min}^* \longrightarrow -\infty$, equation (A-38) becomes

$$\int_\chi^\alpha \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{X^2}{2}\right] dX = \int_{-\infty}^\alpha \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{X^2}{2}\right] dX$$

In summary, the derivatives of $prob(B)$ and $prob(C)$ with respect to ϕ_c can therefore be expressed as follows:

$$\frac{\partial}{\partial \phi_c^*} prob(C) = -\frac{1}{\sqrt{2\pi}} \exp\left[\frac{-\phi_c^{*2}}{2}\right] . \int_{-\infty}^\alpha \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{X^2}{2}\right] dX \tag{A-40a}$$

$$\frac{\partial}{\partial \phi_c^*} prob(B) = \frac{1}{\sqrt{2\pi}} \exp\left[\frac{-\phi_c^{*2}}{2}\right] . \int_\alpha^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{X^2}{2}\right] dX \tag{A-40b}$$

By adding the equations (A-40a) and (A-40b), it gives

$$\frac{\partial}{\partial\phi_c^*}\left(prob(B)+prob(C)\right)=\frac{\partial}{\partial\phi_c^*}prob(C)+\frac{\partial}{\partial\phi_c^*}prob(B)$$

$$\frac{\partial}{\partial\phi_c^*}\left(prob(B)+prob(C)\right)=\frac{\partial}{\partial\phi_c^*}\left(\frac{1}{\sqrt{2\pi}}\exp\left[\frac{-\phi_c^{*2}}{2}\right]..\left(\begin{array}{l}\int_{\alpha}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{X^2}{2}\right]dX\\ -\int_{-\infty}^{\alpha}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{X^2}{2}\right]dX\end{array}\right)\right)$$

(A-41)

Substituting equation (A-13) into equation (A-41) gives

$$\alpha=\frac{\left(\log k_c^*-\rho\ .\phi_c^*\right)}{\sqrt{\left(1-\rho^2\right)}}=\frac{\left(\rho\ .\phi_c^*-\rho\ .\phi_c^*\right)}{\sqrt{\left(1-\rho^2\right)}}=0$$

$$\frac{\partial}{\partial\phi_c^*}\left(prob(B)+prob(C)\right)=\frac{\partial}{\partial\phi_c^*}\left(\frac{1}{\sqrt{2\pi}}\exp\left[\frac{-\phi_c^{*2}}{2}\right]..\left(\begin{array}{l}\int_{0}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{X^2}{2}\right]dX\\ -\int_{-\infty}^{0}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{X^2}{2}\right]dX\end{array}\right)\right) \quad \text{(A-42)}$$

On the basis of the properties of the PDF of the normal distribution, the following integrals are defined (Jensen et al., 2003):

$$\int_{0}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{X^2}{2}\right]dX=\frac{1}{2}\ \text{ And }\ \int_{-\infty}^{0}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{X^2}{2}\right]dX=\frac{1}{2}$$

Equation (A-43) becomes

$$\frac{\partial}{\partial\phi_c^*}\left(prob(B)+prob(C)\right)=\frac{\partial}{\partial\phi_c^*}\left(\frac{1}{\sqrt{2\pi}}\exp\left[\frac{-\phi_c^{*2}}{2}\right]..\left(\frac{1}{2}-\frac{1}{2}\right)\right)=0 \quad \text{(A-44)}$$

In the case where $\log k_C^*=\rho.\phi_c^*$, i.e. the porosity cut-off value is derived from the Y-on-X regression line, the sum $prob(B)+prob(C)$ admits an extremum for $\phi_c=\phi_{cYX}$. The variations of $prob(B)+prob(C)$ with respect to ϕ_c should be investigated so as to determine whether this extremum is a minimum or a maximum.

$$\frac{\partial}{\partial \phi_c^*}\left(prob(B) + prob(C)\right) =$$

$$\frac{1}{\sqrt{2\pi}}\exp\left[\frac{-\phi_c^{*2}}{2}\right].\left(\int_{\alpha}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{X^2}{2}\right]dX - \int_{-\infty}^{\alpha}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{X^2}{2}\right]dX\right) \tag{A-45}$$

where $\alpha = \frac{\left(\log k_c^* - \rho\ .\phi_c^*\right)}{\sqrt{\left(1-\rho^2\right)}}$

The sign of $prob(B) + prob(C)$ depends on the sign of α.

In the case where $\alpha > 0$, i.e. $\frac{\left(\log k_c^* - \rho\ .\phi_c^*\right)}{\sqrt{\left(1-\rho^2\right)}} > 0$, $\log k_c^* - \rho\ .\ \phi_c^* > 0$ and $\phi_c^* < \frac{\log k_c^*}{\rho}$:

(i) $\int_{\alpha}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{X^2}{2}\right]dX < \frac{1}{2}$

(ii) $\int_{-\infty}^{\alpha}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{X^2}{2}\right]dX > \frac{1}{2}$

(iii) $\frac{\partial}{\partial \phi_c^*}\left(prob(B) + prob(C)\right) < 0$

In the case where $\alpha < 0$, i.e. $\frac{\left(\log k_c^* - \rho\ .\phi_c^*\right)}{\sqrt{\left(1-\rho^2\right)}} < 0$, $\log k_c^* - \rho\ .\ \phi_c^* < 0$ and $\phi_c^* > \frac{\log k_c^*}{\rho}$:

(i) $\int_{\alpha}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{X^2}{2}\right]dX > \frac{1}{2}$

(ii) $\int_{-\infty}^{\alpha}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{X^2}{2}\right]dX < \frac{1}{2}$

(iii) $\frac{\partial}{\partial \phi_c^*}\left(prob(B) + prob(C)\right) > 0$

The variations of the derivative of $prob(B) + prob(C)$ with respect to ϕ_c is summarized in the table below:

Table of variations of the derivative of prob(B)+prob(C) with respect to ϕ_c.

ϕ_c^*	$-\infty$	ϕ_{cYX}^*	$+\infty$
$\frac{a}{d\phi_c^*}\left(prob(B)\right)$	+		+
$\frac{d}{d\phi_c^*}\left(prob(C)\right)$	-		-
$\frac{d}{d\phi_c^*}\left(p(B)+p(C)\right)$	-	O	+
$prob(B)+prob(C)$	$\searrow$	O	$\nearrow$

The sum $prob(B)+prob(C)$ is therefore minimized for $\phi_c = \phi_{cYX}$. In order to evaluate NGR and equalize $prob(B)$ and $prob(C)$, i.e. to equalize the errors of misidentifying pay for non-pay and non-pay for pay, the RMA line is required. When delineating net pay, i.e. minimizing the sum of the errors in identifying pay and non-pay, the optimal porosity cut-off is given by the Y-on-X regression line. Those results are valid in case the log k-ϕ dataset is assumed to be joint normal distributed.

APPENDIX B

TEST FOR NORMALITY: PROBABILITY PLOTS

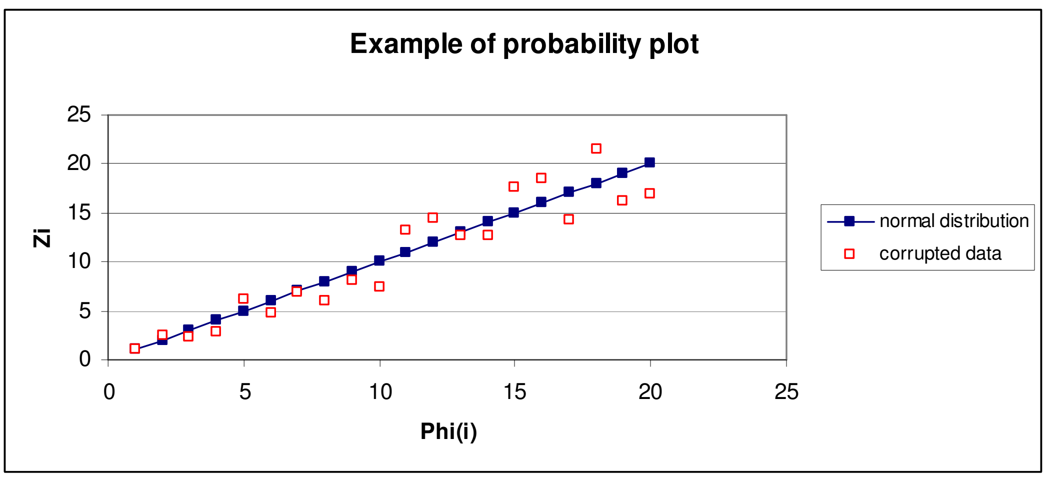
The use of probability plot is a convenient method to test the normality of a univariate population. Let consider a population of n realizations $\{\phi_1, \phi_2, ..., \phi_{n-1}, \phi_n\}$, where $\phi_1 \le \phi_2 \le \phi_3 \le ..... \le \phi_{n-1} \le \phi_n$. The likelihood for any realization, ϕ_i, to occur may be defined by equation (B-1).

$$p_i = \frac{i - \frac{1}{2}}{n}, \ \forall i \in [1, n]$$

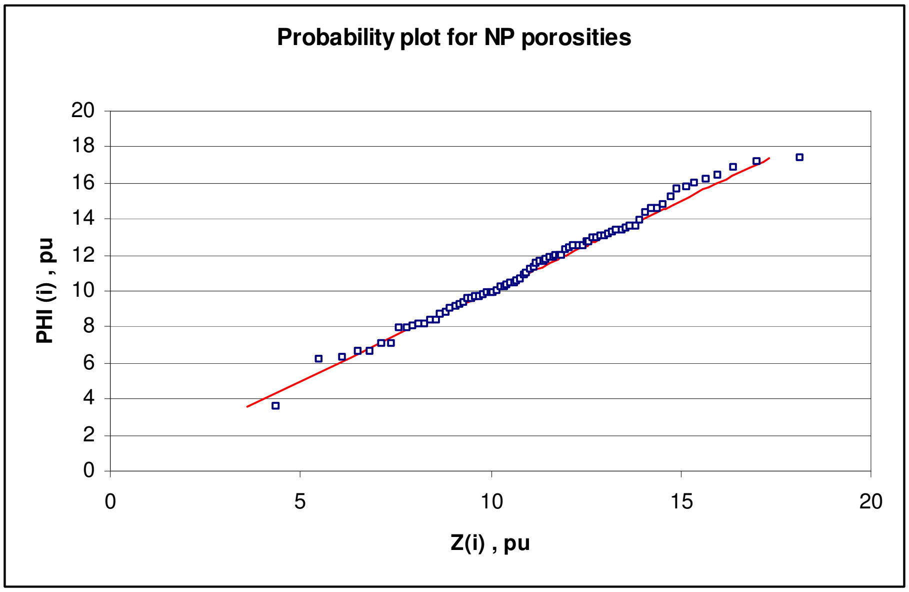
These probability values are inverted so as to obtain normal distributed values Z_i :

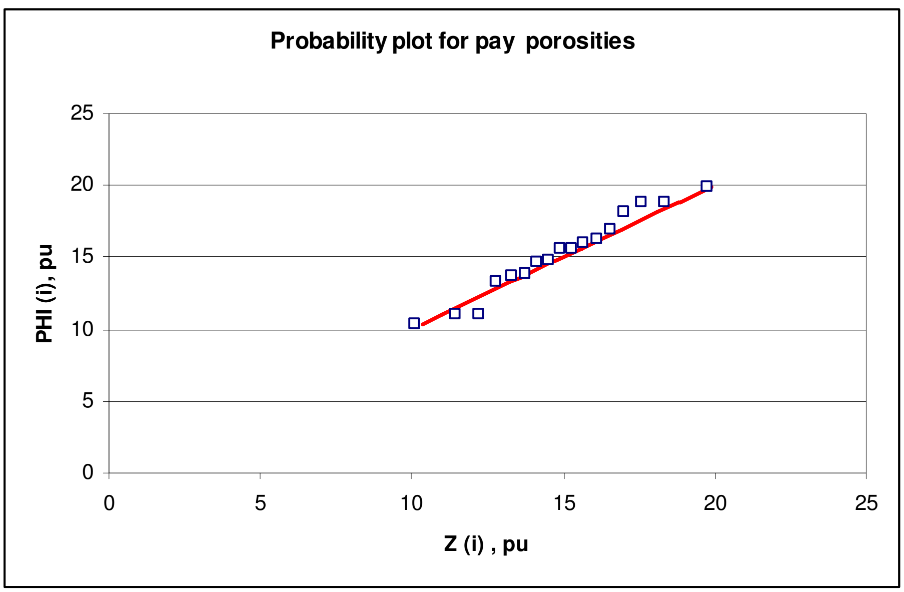
$$Z_i = -\sqrt{2}\, erf^{-1}[1 - 2 * p_i]$$

The values of Z_i may be plotted versus the values of ϕ_i: the more the univariate population deviates from the normal distribution, the less correlated the variates Z_i and ϕ_i are. In this case, the data will significantly deviate from the straight line of slope 1.



The assumption of normality for the non-pay and pay fraction of porosities is tested. 100 samples are generated from the joint normal bivariate population under investigation. A permeability cut-off value of k_c= 1 md is applied so as to segregate the non-pay from the pay fraction of porosities. The p(i) and Z(i) values are calculated from the porosity values and the probability plots for non-pay and pay-fraction are generated:

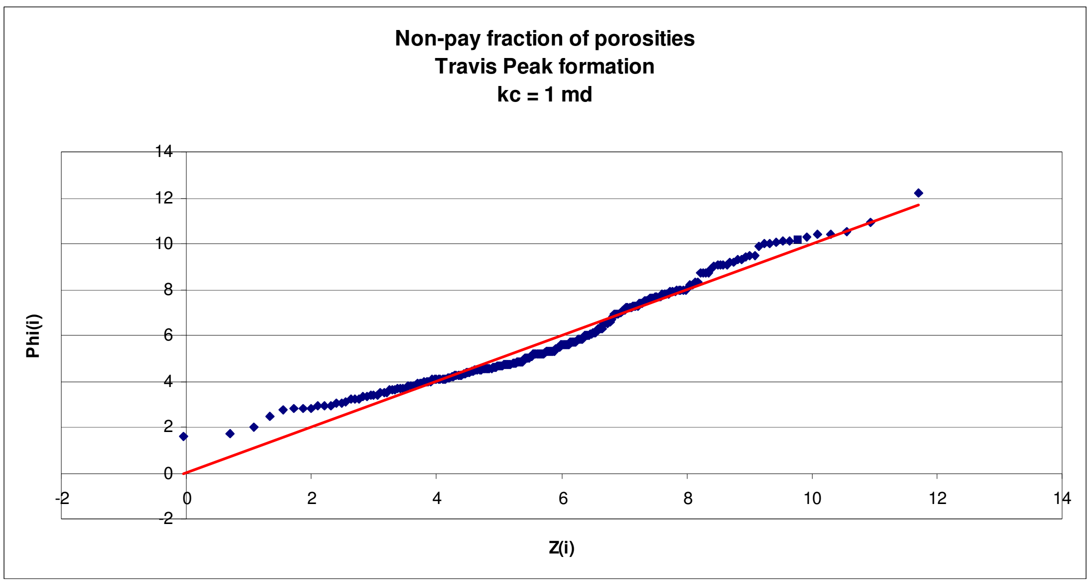




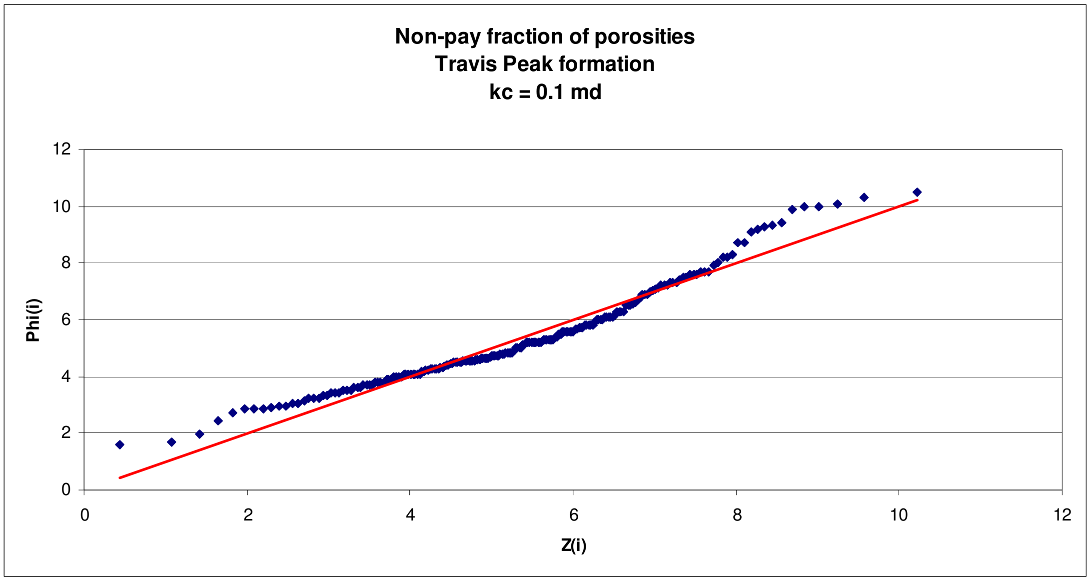
The pay and non-pay fraction of porosities do not deviate significantly from the line with slope 1, meaning they can be considered as normally distributed. The second plot shows that even with a few data, the pay fraction of porosities still has a normal distribution. Probability plots may be created for other degree of sampling: it is expected that the normality of porosities will be more and more clearly expressed as the number of samples increases. The discriminant analysis may be properly used with these datasets generated from the population density under consideration.

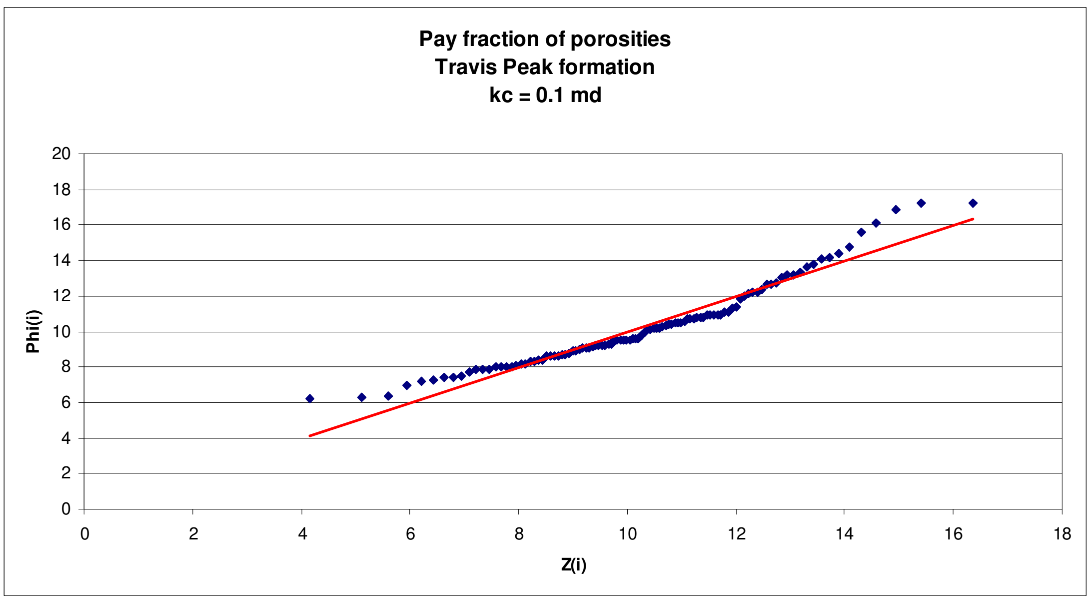
The probability plots are realized for the pay and non-pay fraction of the porosities for the Travis Peak formation.

For the case where k_c = 1 md, the normality of the two sub datasets is not rejected even though the normality is slightly altered for the larger and smaller porosities of the pay fraction.

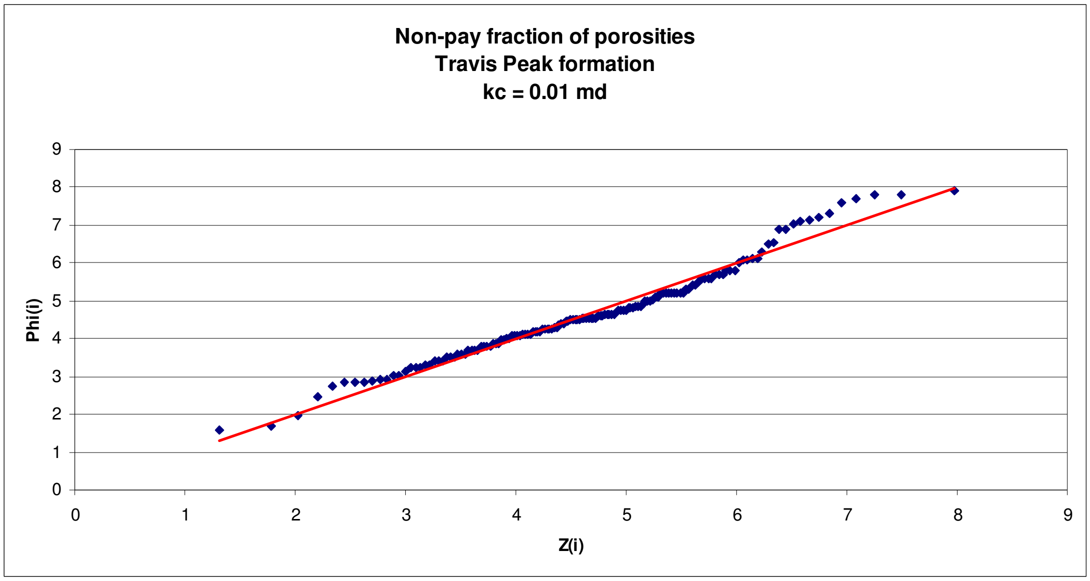


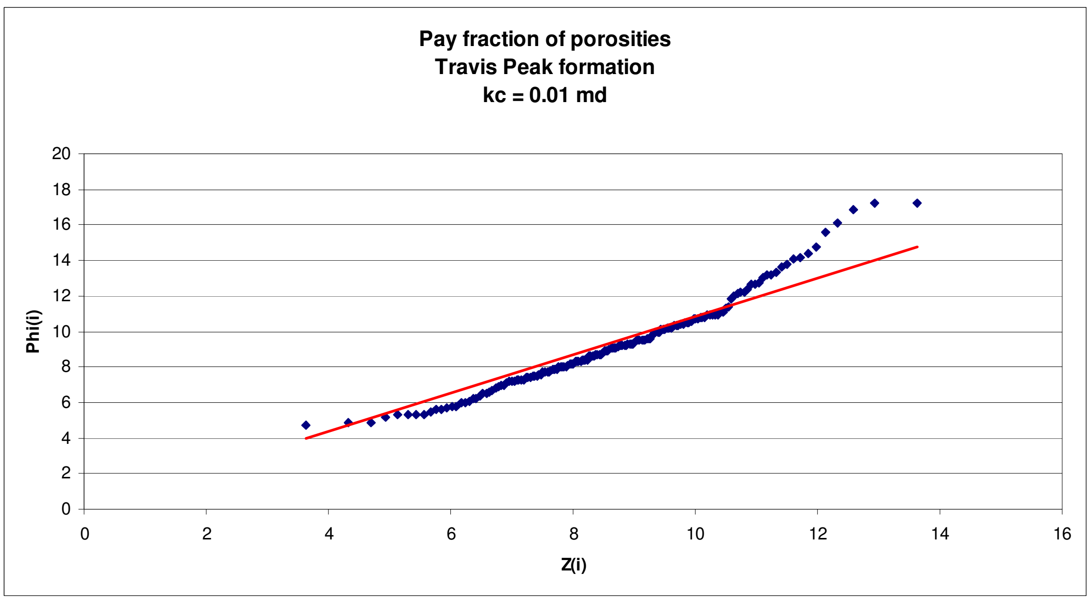
For the case where k_c = 0.1 md, the normality of the two sub datasets is not rejected.





For the case where k_c = 0.01 md, the normality of the non-pay fraction of the dataset is not rejected. However, the probability plot for the pay fraction of porosities clearly shows a deviation to the normality for porosities larger than 11 percent.


Non-pay fraction of porosities
Travis Peak formation
kc = 0.01 md
Phi(i)
Z(i)
0
1
2
3
4
5
6
7
8
9


Pay fraction of porosities
Travis Peak formation
kc = 0.01 md
Phi(i)
Z(i)
0
2
4
6
8
10
12
14
16
18
20

APPENDIX C

BIVARIATE NORMAL DENSITY

Sampling from a bivariate normal distribution, or joint normal distribution may be easily conducted by using a classic algorithm.

The joint normal distribution is defined by the following PDF,

$$f(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left[-\frac{x^{*2}-2\rho x^* y^* + y^{*2}}{2\left(1-\rho^2\right)}\right] \tag{C-1}$$

where $x^* = \dfrac{x-\mu_X}{\sigma_X}$, $y^* = \dfrac{y-\mu_X}{\sigma_Y}$, and $\rho = \dfrac{\sigma_{XY}}{\sigma_X\sigma_Y}$.

So as to sample from the pdf defined by equation (C-1), two uncorrelated, standard normal variates, Z_1 and Z_2 are generated. For instance any random function may be used and then converted into a normal distribution function.

$$Z \rightarrow \Phi\left(\mu = 0, \sigma^2 = 1\right)$$

The correlation is introduced by computing the two variates, X_1 and X_2, as shown by equations (C-3) and (C-4):

$$X_1 = \mu_1 + \sigma_1 Z_1$$

$$X_2 = \mu_2 + \sigma_2\left[Z_1 . \rho + Z_2 . \sqrt{1-\rho^2}\right]$$

The quantities μ_1 and σ_1 are the mean and standard deviation of the variate X_1 whereas μ_1 and σ_1 are the mean and standard deviation of the variate X_2. The parameter ρ corresponds to the coefficient of correlation between the variates X_1 and X_2.

Joint normal bivariate datasets may be therefore generated with various numbers of paired values, i.e. samples of ϕ and k.

APPENDIX D

HYPOTHESIS TEST ON INTERCEPT OF ESTIMATOR'S STANDARD ERRORS EQUATIONS

The objective is to test that the intercepts, β_0, of the equations of the standard errors for the different estimators under consideration are not significantly different from zero. Two-tailed t-tests using the null hypothesis H_0: "$\beta_0=0$". The following results show that in all cases, the t-statistic is within the t-critical range obtained from a table of values for the t-distribution at the 5% confidence level. Therefore, H_0 is accepted and the intercepts are found not to be significantly different from zero.

Y-on-X	No noise, k_c=1.5 md	No noise, k_c=1 md	No noise, k_c=0.5 md	No noise, k_c=0.1 md
β_0	0.021607	-0.0371	0.018489	-0.009
N= n-2	7	7	7	7
T statistic	-1.69852	-1.86093	-2.05524	-1.80515
t critical range (two-tailed test)	[-2.37, 2.37]	[-2.37, 2.37]	[-2.37, 2.37]	[-2.37, 2.37]
Conclusion	**accept Ho**	**accept Ho**	**accept Ho**	**accept Ho**

RMA	No noise, k_c=1.5 md	No noise, k_c=0.5 md
β_0	0.00463	0.003195
v= n-2	7	7
t statistic	-2.31113	-1.59636
t critical range (two-tailed test)	[-2.37, 2.37]	[-2.37, 2.37]
Conclusion	**accept Ho**	**accept Ho**

RMA, k_c=0.1 md	No noise	Corrupted with noise 1	Corrupted with noise 2
β_0	0.0022	0.007277	0.016806
v= n-2	7	6	6
t statistic	-0.7129	0.480951	0.654535
t critical range (two-tailed test)	[-2.37, 2.37]	[-2.447, 2.447]	[-2.447, 2.447]
Conclusion	**accept Ho**	**accept Ho**	**accept Ho**

RMA, k_c=1 md	No noise	Corrupted with noise 1	Corrupted with noise 2
β_0	0.0022	0.00645	0.007855
v= n-2	7	5	6
t statistic	-1.405	0.4495	-0.50925
t critical range (two-tailed test)	[-2.37, 2.37]	[-2.571, 2.571]	[-2.447, 2.447]
Conclusion	**accept Ho**	**accept Ho**	**accept Ho**

Discriminant, k_c=1 md	No noise	Corrupted with noise 2	Corrupted with noise 1
β_0	0.0019	0.075545	-0.075352
v= n-2	5	5	5
t statistic	-0.757	0.361374	0.135364
t critical range (two-tailed test)	[-2.57, 2.57]	[-2.57, 2.57]	[-2.57, 2.57]
Conclusion	**accept Ho**	**accept Ho**	**accept Ho**

Quarter (NGR), k_c=1 md	No noise	Corrupted with noise 1	Corrupted with noise 2
β_0	-0.0784	-0.030411	-0.056807
v= n-2	5	4	4
t statistic	-1.58351	-0.55242	-0.59147
t critical range (two-tailed test)	[-2.57, 2.57]	[-2.776, 2.776]	[-2.776, 2.776]
Conclusion	**accept Ho**	**accept Ho**	**accept Ho**

APPENDIX E

TEST FOR JOINT NORMALITY

The test for joint normality of the log(k)-ϕ dataset from the Travis Peak formation is conducted by using the quantities $\hat{U}_3^2$ and $\hat{U}_4^2$ which are a generalization of the first two non-zero components of the Lancaster's test for univariate normality.

One assumes that (x_{11},x_{12}), (x_{21},x_{22}), ………, (x_{n1},x_{n2}) is a bivariate sample of size n.

For i = 1, 2, 3,…., n,

$$y_{i,1} = (x_{i,1} - \mu_1).\sigma_1^{-1}$$

$$y_{i,2} = ((x_{i,2} - \mu_2).\sigma_2^{-1} - \rho(x_{i,1} - \mu_1).\sigma_1^{-1}).(1 - \rho^{-0.5})$$

where μ_1 and σ_1 represents respectively the average and standard deviation of the variables $x_{i,1}$ and μ_2 and σ_2 represents respectively the average and standard deviation of the variables $x_{i,2}$.

$$\hat{U}_3^2 = n..((m_{21}^2 + m_{12}^2)/2 + (m_{30}^2 + m_{03}^2)/6)$$

$$\hat{U}_4^2 = n.((m_{22} - 1)^2/4 + (m_{31}^2 + m_{13}^2)/6 + ((m_{04} - 3)^2 + (m_{40} - 3)^2)/24)$$

where $m_{rs} = \frac{1}{n}.\sum_{i=1}^{n} y_{i1}^r y_{i2}^s$

The components of $\hat{U}_3^2$ are $m_{21}/\sqrt{2}$, $m_{12}/\sqrt{2}$, $m_{03}/\sqrt{6}$, $m_{30}/\sqrt{6}$. The components of $\hat{U}_4^2$ are $(m_{22} - 1)/2$, $m_{31}/\sqrt{6}$, $m_{13}/\sqrt{6}$, $m_{04}/\sqrt{24}$,($m_{40} - 3)/\sqrt{24}$.

In our case, testing for joint normality we found that $\hat{U}_3^2 + \hat{U}_4^2$ =250.10 with components equal to -0.31, 0.40, 0.33, 0.06, 0.2867, -0.47, 0.048, 0.06 and n = 320. No component does account for most of the quantity $\hat{U}_3^2 + \hat{U}_4^2$. The latter is nonetheless moderately insignificant in spite of the strong influence of the outliers. The hypothesis of joint normality is therefore not rejected.

VITA

Name: Nicolas Bouffin

Permanent Address: 7 Rue de FECAMP, 57140 WOIPPY, FRANCE

Email Address: nbouffin@yahoo.fr

Education: M.Eng, Geology and Geological Engineering, Ecole Nationale Supérieure de Géologie, Nancy (France), 2005

M.S., Geoscience and Reservoir Engineering, IFP School, Paris (France), 2007

M.S., Petroleum Engineering, Texas A&M University, College Station, Texas (U.S.A.), 2007

Professional Experience: Total Exploration and Production, Pau, France – Intern

Total E&P Indonesia, Balikpapan, Indonesia – Intern

BP, Houston, Texas – Reservoir Engineer