PRJ, Vol. 7 (January 1995); P. 67–74, 3 FIGS. 1 TABLE.

PROBABILITY OF THE VALIDITY OF THE ROCK COMPONENT MODEL IN COMPLEX LOG INTERPRETATION

G. Szigeti*

إحتمالية صحة تطبيق نموذج المركبات الصخرية في تفسير سرود الآبار المعقدة

ج. زيقيتي

التقليل من الفارق بين السرود المسجلة والمحسوبة أو النظرية هو أساس التفسير للسرود. وفي حالة تعدد المركبات الصخرية في النموذج المستخدم لا يمكن مقارنة الفارق مباشرة ، لذلك تم إدخال إحتمالية صحة أو تطبيق النموذج المستخدم لغرض المقارنة ، كما تم دراسة تأثير التحديدات على حجم المعادن المحسوبة من النموذج المستخدم.

وتم تعديل توزيع «كاي» النظري واختبار حدود تطبيقاته في وجود هذه التحديدات ، كما تم تطبيق نموذج الإحتمالية لغرض الإيضاح في تفسير السرود المسجلة من تكوين صخور رملية.

ABSTRACT

The minimization of the incoherence (the distance between measured and calculated logs) is a basic principle of complex well log interpretation. If lithological models with different numbers of unknowns (rock components) occur, however, the incoherences cannot be compared direclty.

The probability, α, of validity of rock component models is introduced as the basis of comparison. The effect of restrictions on the mathematically evaluated mineral volume fractions is discussed. The theoretical Chi-distribution is modified and the limits of its application are examined in case of restrictions.

The model probability, α, concept is applied for practical illustration in the interpretation of a well in a sandstone formation.

INTRODUCTION

The interpretation of geophysical data, in general, involves the comparison of the measured field data with that of the geologic geophysical model of the situation under study.

This comparison can be mathematically formulated as a minimization task where the goal is to minimize the distance, called incoherence, between the empirical and the theoretical model as a function of some model parameters. This idea leads to acceptable result for any fixed model but does not give a key, which model to accept. Indeed, in general, the more parameter considered by the model the less becomes the distance between the measured and calculated results. Defining too many parameters for the system of response functions, as might be known from practice, the determination becomes less definite.

On the other hand, as a property of any measuring system the experimental material consists of some degree of unavoidable uncertainty. Because of the presence of this uncertainty the statement about the validity of the interpretation model becomes a probability statement rather than a certain one. Having adequate definition for this probability and mechanism for its calculation we accept the maximum likelihood principle which means the choice of the model which has a higher probability of validity and which is not necessarily the model having the more parameters.

The objective of the present paper is to give a proper probability measure to the simple distance

* Petroleum Research Centre P.O. Box 6431, Tripoli, G.S.P.L.A.J.

concept, where this measure depends on the dimensions of the minimization task as well. The mathematical scheme applied in this work, based on the Fisher-Cochran theory, is close to that of decision making of mathematical statistics. Two specific features, however, influence the way of thinking. First, we have always low number of different logs in a given sample point, and second, the log readings represent a strong statistical average of a repeated observation.

The mathematical scheme developed does not give a key on how to set up rock component models, however, it is a useful tool to decide, when it is possible, which model to choose in a doubtful situation.

SETTING UP THE PROBLEM AND ESTABLISHMENT OF THE DECISION MAKING

Let the measured field data be denoted by ε concrete form, of which is indifferent at the moment, and suppose that two different geological-geophysical models A_1 and A_2 can be set up on the basis of the general knowledge about the situation under study. The question is which model describes better the real geological conditions on the basis of observation ε. Having the answer for two models, of course, we have the answer for any finite number of models.

We suppose at this point, in advance, that the probabilities of the involved events do exist, i.e. can be correctly defined later on. In this sense, $P(A_1)$ means the probability of the event that A_1 is the right model of the studied situation. The observed ε is the only sure thing about the studied situation, thus we need to compare the conditional probabilities

$$P(A_1|\varepsilon) \quad \text{and} \quad P(A_2|\varepsilon) \tag{1}$$

Because of the purely empirical character of ε, conclusion about the nature of the model cannot be drawn directly. A_1 and ε are mutually valid if within the ε event A_1 is valid, or if within the A_1 event ε is valid, this can be described using conditional probabilities by the equation

$$P(A_1\varepsilon)=P(\varepsilon)P(A_1|\varepsilon)=P(A_1)P(\varepsilon|A_1)$$

from which

$$P(A_1|\varepsilon)=\frac{P(A_1)}{P(\varepsilon)}*P(\varepsilon|A_1) \tag{2}$$

This is the simplest form of the Bayes theorem, for one hypothetical model A_1. Substituting (2) in (1) the answer to the problem lays in the comparison of quantities.

$$\frac{P(A_1)}{P(\varepsilon)}\cdot P(\varepsilon|A_1) \quad \text{and} \quad \frac{P(A_2)}{P(\varepsilon)}\cdot P(\varepsilon|A_2) \tag{3}$$

For this comparison $P(\varepsilon)$ does not play any role, $P(A_1)$ and $P(A_2)$ must reflect the knowledge about the area prior to the measurements. Often there is no evidence about the priority of one model over the other (e.g. clean and shaly sand equally possible at a given depth point of a sandstone area) and in this case (3) simplifies to the comparison of

$$P(\varepsilon|A_1) \quad \text{and} \quad P(\varepsilon|A_2) \tag{4}$$

The essential difference between (1) and (4) is that, unlike ε, A_1 (and A_2) is a theoretically defined complete situation, the model of the area under study, which gives the opportunity to define and calculate the probability of occurrence of the measured material ε.

Definition of the Probability of Occurrence of Given Measured Data at a Given Model of Interpretation

Now, the studied situation is the unit volume rock (true formation) which surrounds the bore hole at a given depth point. The measured data ε is the vector $\underline{\mathbf{b}}_M$ which consists of the number m log readings of different physical parameters at the same depth point. The model of the rock is the k number component of the rock (typically one fluid and k-1 solid component) and its linearized response function for each tool is in the form of the zone parameter matrix **Z** having m rows and k columns.

Because the model A_1 is valid (there is conditional probability in (4)) one defined independently from the exploration, but unknown k dimensional vector $\underline{\mathbf{x}}$ belongs to it, which gives the volumes of each components of the rock at the given depth. **Z** and $\underline{\mathbf{x}}$ together generate the ideal response vector $\underline{\mathbf{b}}$ in the form

$$\mathbf{Z}\underline{\mathbf{x}}=\underline{\mathbf{b}} \tag{5}$$

which gives m exact equations, without contradiction. If there is no inner connection between $\underline{\mathbf{b}}$ and $\underline{\mathbf{b}}_M$ then there is no way to know $\underline{\mathbf{x}}$. Consequently we make at this point the strong assumption that $\underline{\mathbf{b}}_M$ is an m-dimensional stochastic vector variable of Gaussian distribution with expectation $\underline{\mathbf{b}}$ and with a unit variation of each coordinates. The last assumption means, in practice, that the variation of each component is known and equation (5) have been already normalized according to this variations. Hence, instead of (5) we have;

$$\mathbf{Z}\cdot\underline{\mathbf{x}}=\underline{\mathbf{b}}_M \tag{6}$$

overdetermined system of equations. The best linear estimation for $\underline{x}$ is $\underline{\alpha}$ given by the Gaussian least square method (see e.g. [1]), which is the solution of the (ordinary) linear equation system

$$Z^T \cdot Z\underline{\alpha} = Z^T \underline{b}_M \tag{7}$$

There is a strong connection between $\underline{x}$ and $\underline{\alpha}$, namely the expectation value of $\underline{\alpha}$ is $\underline{x}$, $E(\underline{\alpha}) = \underline{x}$. In equations (5), (6) and (7) Z, $\underline{x}$ and $\underline{b}$ are deterministic values, while $\underline{b}_M$ is stochastic vector variable and $\underline{\alpha}$ is a linear combination of coordinates of $\underline{b}_M$ is stochastic variable as well.

Now, we are going to examine the distribution of the distance between the two sides of equation (6) around the $E(\underline{\alpha}(\underline{b}_M)) = \underline{x}$ point. Multiplying (5) by Z^T from the left and subtracting from (7), we get

$$(\underline{\alpha} - \underline{x}) = (Z^T Z)^{-1} * Z^T(\underline{b}_M - \underline{b}) \tag{8}$$

The incoherence I between the measured vector and the response function with the best statistical estimation $\underline{\alpha}$, using (5) and (8) can be written in the form

$$I = \| \underline{b}_M - Z \cdot \underline{\alpha} \| = \| (\underline{b}_M - \underline{b}) - Z(\underline{\alpha} - \underline{x}) \|$$

$$= \| (\underline{b}_M - \underline{b}) - Z(Z^T Z)^{-1} \cdot Z^T(\underline{b}_M - \underline{b}) \|$$

$$= \sqrt{\eta_1^2 + \eta_2^2 + \cdots \eta_m^2}, \tag{9}$$

where $\| \ \|$ is the Euclidean distance in the m-dimensions, and $\eta_1, \eta_2, \ldots \eta_m$ are the coordinates of the incoherence, each one is a linear combination of the coordinates of the vector $\underline{b}_M - \underline{b}$. As a matter of our assumption, the coordinates of the stochastic vector $\underline{b}_M - \underline{b}$ are independent of each other, as their mutual distribution is a Gaussian distribution with the zero vector of expectation and with the unit (matrix) variation. The coordinates of the $\underline{\eta} = (\eta_1, \eta_2, \ldots \eta_m)$ incoherence vector are not independent from each other, because there is k number of linear relationship among them, namely,

$$Z^T \cdot \underline{\eta} = Z^T(\underline{b}_M - \underline{b}) - Z^T \cdot Z(Z^T \cdot Z)^{-1} \cdot Z^T(\underline{b}_M - \underline{b}) \equiv 0$$

To suppose more linear relationships among the coordinates of η contradicts with the independence of $\underline{b}_M - \underline{b}$. According to the Fisher-Cochran theorem [1], from this situation it follows for η, examining it in a properly chosen orthonormal basis, that its m-k coordinates vary freely, while the remainder k coordinates equalise each other to zero. From this it follows that the distribution of the right side expression of (9) is the distribution of the distance of m-k number of independent Gaussian distributions, each one with zero expectation and unit variation. In other words we got the theoretical distribution of the incoherence I for the A_1 model (represented by the Z matrix). This incoherence is a stochastic scalar variable, let it be denoted by ξI_{A_1} and its distribution is the Chi-distribution belonging to the m-k parameter, i.e.

$$\xi I_{A_1} = \| \underline{b}_M - Z \cdot \underline{\alpha} \| = \chi_{m-k} \tag{10}$$

The distribution of the Chi random variable χ_n is well known, and for small values of **n** is characteristic for the complex log interpretation application, and can be expressed by the following form [2]

$$P(\chi_n < x) = \int_{-\infty}^{x} h_n(t)\, dt$$

$$h_n(t) = \begin{cases} 0 & t \leqslant 0 \\ g_n(t) & t > 0 \end{cases}$$

$$g_1(t) = \sqrt{\frac{2}{\pi}} \cdot e^{-(t^2/2)}; \qquad g_2(t) = t \cdot e^{-(t^2/2)} \tag{11}$$

$$g_3(t) = \sqrt{\frac{2}{\pi}} \cdot t^2 \cdot e^{-(t^2/2)}; \quad g_4(t) = \tfrac{1}{2} \cdot t^3 \cdot e^{-(t^2/2)}$$

$$g_5(t) = \frac{1}{3}\sqrt{\frac{2}{\pi}} \cdot t^4 \cdot e^{-(t^2/2)}.$$

In this way we established the theoretical distribution of the incoherence as a result of the uncertainty of observation $\underline{b}_M$. This distribution is a function of the only variable $DF = m - k$ degree of freedom which reflects the dimension of the observations and the dimension of the applied model of interpretation.

On the other hand, substituting the only measured value for the vector $\underline{b}_M$ into formula (9) we get the experimental incoherence

$$I_{\underline{b}_M} = \| \underline{b}_M - Z \cdot \underline{\alpha} \|$$

using the $\underline{\alpha}$ statistical estimation of $\underline{x}$ defined by (7).

Now it is clear, the smaller the $I_{\underline{b}_M}$ empirical incoherence the more reason to accept the choice for model A_1, thus we define the probability of validity of the rock component model A_1 as the probability measure of the event that the theoretical distribution is higher than the experimental one, and denoting this probability by α we estasblish

$$\alpha = P(\varepsilon = \underline{b}_M \mid A_1) = P(\xi I_{A_1} \geqslant I_{\underline{b}_M}) \tag{12}$$

Substituting (10) into (12), using (11) and some elementary features of any distribution function, and also changing the notation of the experimental $I_{\underline{b}_M}$ for I_{exp}, we get

$$\alpha = P(\xi I_{A_1} \geqslant I_{\underline{b}_M}) = 1 - P(\xi I_{A_1} < I_{\underline{b}_M})$$
$$= 1 - P(\chi_{m-k} < I_{exp})$$
$$= 1 - \int_0^{I_{exp}} g_{m-k}(t)\, dt \qquad (13)$$

Returning to the original problem (4), the observation $\underline{\mathbf{b}}_M$ out of the models A_1 and A_2 verify that model, and that estimation $\underline{\alpha}$ given by the model, for which the probability α of (13) is higher. In (13) m is the number of logs involved into the interpretation, while I_{exp} and k depend very much on the models A_1 or A_2 used for interpretation purpose. Figure 1 demonstrate, the probability α as a function of DF and I_{exp}.

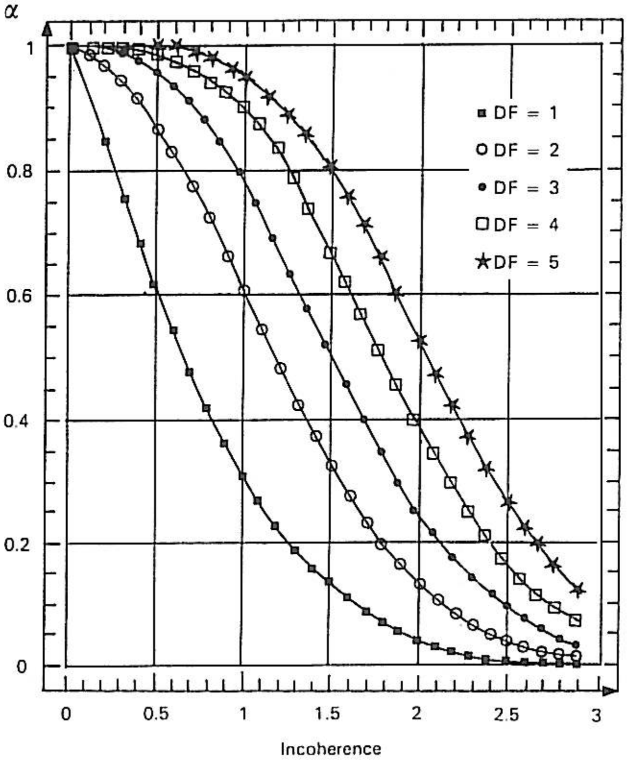


FIG. 1. Probability of validity of rock component model as a function of degree of freedom and incoherence.

Effect of Restrictions on the Degree of Freedom and the Exactness of the Calculation of Probability α

In real practice of complex log interpretation instead of equation system (6) we solve more complete task. In fact, we use one more equation, the material balance equation

$$x_1 + x_2 + \cdots x_k = 1 \qquad (14)$$

and we require all the volume fractions to be non negative, $x_i \geqslant 0$, $i = 1, 2, \ldots k$, having in this way the set of so called logical restrictions [3].

The new equation (14) simply means that in (13) we have to put $m+1$ instead of m in the definition of degree of freedom, since the deterministic equation (14) is a limit of a stochastic equation tending to zero variance on its right side, see [4].

In equations (8) and (9) the $\mathbf{Z}(\underline{\alpha} - \underline{\mathbf{x}}) \in \mathbf{R}^m$ vector represents the projection of the $\underline{\mathbf{b}}_M - \underline{\mathbf{b}} \in \mathbf{R}^m$ vector to the k dimensional hyper plain $\kappa(\underline{\mathbf{z}}_1, \underline{\mathbf{z}}_2, \ldots \underline{\mathbf{z}}_k)$ of $\mathbf{R}^m$, defined by the all possible linear combinations of vectors $\underline{\mathbf{z}}_1, \underline{\mathbf{z}}_2, \ldots \underline{\mathbf{z}}_k$ which are the column vectors of matrix $\mathbf{Z}$. Until the projection of the decisive mass of the m dimensional Gaussian distribution of $\underline{\mathbf{b}}_M - \underline{\mathbf{b}}$ is inside of the domain allowed by restrictions i.e. the projection is given in the form $\underline{\mathbf{z}}_1 x_1 + \underline{\mathbf{z}}_2 x_2 + \cdots + \underline{\mathbf{z}}_k x_k$, where all $x_i \geqslant 0$, then the theoretical distribution of the incoherence is undisturbed and is the Chi-distribution as given by (10). However, when the projection of the only measured $\underline{\mathbf{b}}_M$ vector is outside or inside but close to the boundaries of the allowed domain then the projection we calculated for the theoretical distribution in (9) is not completely inside the allowed domain and so in formula (9) $\underline{\alpha}$ loses its meaning and, in this way, our definition for ξI_{A_1} fails.

Now we show that having the solution, as in [3], for the minimising task with restrictions we can avoid the case when the ordinary projection of the measured $\underline{\mathbf{b}}_M$ vector (which is the solution (7) of the minimization task without restrictions) is outside the allowed domain. Indeed, suppose that the exact solution of the minimising task with restrictions is $\hat{\underline{\mathbf{x}}} \in \mathbf{R}^k$ contain n_0 number of zeros ($n_0 < k$). In this case we redefine the interpretation model by leaving off the zone parameters belonging to the zero values in $\hat{\underline{\mathbf{x}}}$, having in this way the $\mathbf{Z}'$ matrix of m rows and $k - n_0$ columns. Now the solution with restrictions for the $\mathbf{Z}'$ model is evidently given by $\hat{\underline{\mathbf{x}}}$ (without the zeros in it) because the minimum is given by $\hat{\underline{\mathbf{x}}}$ even on a wider set. From this it follows that for the real geophysical task with logical restrictions and with the material balance equation (14) the correct definition of the degree of freedom in the formula (13) for the probability of the model validity is

$$DF = (m+1) - (k - n_0) \qquad (15)$$

where n_0 is the number of zeros in the solution of the minimization task with logical restrictions.

Analysing the case when the ordinary projection of the measured $\underline{\mathbf{b}}_M$ is inside the allowed domain but close to the boundary, we confine ourselves to giving estimation for the minimum distance which guarantees that we can use formula (13) with degree of freedom given by (15) for calculating the probability α with a satisfying accuracy. We construct this estimation only for the most important, (i.e. most frequent) case in practice, when the projection is near to only one boundary, given by the $x_i \geqslant 0$ condition, and far from all the others. As long as the projection of $\underline{\mathbf{b}}_M - \underline{\mathbf{b}}$ at the calculation of the theoretical distribution is inside the allowed domain, vector $\underline{\alpha}$ keeps its original meaning, and formula (9) is valid. This projection is still a Gaussian distribution around $\underline{\mathbf{b}}$ with a unit variation and with $k-n_0$ independent coordinates. Let us denote the distance of $Z\underline{\alpha}$ from the i-th boundary by μ_b. In this case it can be shown easily that $\Phi(\mu_b)$ part of the distribution is undistrubed by the i-th boundary, where Φ is the error function, the distribution function of the one dimensional Gaussian distribution. For example at $\mu_b=2$, $\Phi(\mu_b)=0.9772$ that is, for this part of the distribution $\underline{\alpha}$ still represents a model point, so the calculation of the probability α in (13) is satisfactorily accurate.

So we need to calculate the distance of the $\mathbf{Z}\hat{\underline{\mathbf{x}}}$ point from the i-th boundary, where $\hat{\underline{\mathbf{x}}}$ is the solution of the minimization task with logical restrictions (which is the same as $\underline{\alpha}$ calculated after the dimension reduction described in (15)). We can avoid repeating the calculation for each $\hat{\underline{\mathbf{x}}}$ solution at different depth points, noticing that if we calculate only once the distance D_i of $\underline{z}_i$, the vertex of the allowed simplex domain, from the $\kappa(\underline{z}_1, \dots \underline{z}_{i-1}, \underline{z}_{i+1}, \dots \underline{z}_{k-n_0})$ hyper plane then the distance we are looking for becomes x_i part of D_i, where x_i is the i-th co-ordinate of $\underline{\mathbf{x}}$. This statement is a direct consequence of the fact that $t\underline{z}_i+\kappa(\underline{z}_1, \dots \underline{z}_{i-1}, \underline{z}_{i+1}, \dots \underline{z}_{k-n_0})$ represents parallel planes, distance of which is regulated linearly by the t parameter alone. In this way the formula (13) remains valid, despite of the restriction conditions, with accuracy governed by μ_b if

$$x_i>\frac{\mu_b}{D_i} \tag{16}$$

for $i=1, 2 \dots k-n_0$.

To calculate the D_i distances we have to find the projection of $\underline{z}_i$ for the $\kappa(\underline{z}_1, \dots \underline{z}_{i-1}, \underline{z}_{i+1}, \dots \underline{z}_{k-n_0})$ hyper plane, which is again given by the best statistical estimation of the

$$(Z_{i-})\cdot\underline{y}=\underline{z}_i \tag{17}$$

overdetermined system of equation, where Z_{i-} is the Z matrix after neglecting its i-th column, and the unknown $\underline{y}$ has $k-1$ coordinates. Thus if we express the projection of vector $\underline{z}_i$, as in (7), then we get the distance

$$D_i=\|\underline{z}_i-Z_{i-}(Z_{i-}^T Z_{i-})^{-1}\cdot Z_{i-}^T\underline{z}_i\| \tag{18}$$

where $\| \ \|$ is the m-dimensional Euclidean distance.

If we have a component in $\hat{\underline{\mathbf{x}}}$ different from zero volume component which does not satisfy inequality (16), that occurs for a fairly defined zone parameter matrix only at a lower percentage of all the depth points of interpretation, then for the calculation of α, instead of the Chi-distribution of (11), we use more sophisticated boundary distributions, which is not the subject of the present paper.

For the sake of conciseness let us introduce besides the degree of freedom in (13) the symbolical degree of freedom as follows:

If DF′ and DF″ are two positive integers and β' and β'' two real numbers, then the statement that we apply (13) with the symbolical degree of freedom

$$DF=\beta'\cdot DF'\oplus\beta''\cdot DF'' \tag{19}$$

means that

$$1-\alpha=\beta'\int_0^{I_{exp}} g_{DF'}(t)\,dt+\beta''\cdot\int_0^{I_{exp}} g_{DF''}(t)\,dt \tag{20}$$

Now, suppose that for all the $i=1, 2, \dots k-n_0$ (16) is satisfied except one i for which x_i tends through positive numbers to zero. In this case one restriction effects the determination of the theoretical distribution of incoherence. Using the symmetry of this special case it is not difficult to realise that we get the proper calculation for the model validity probability by applying (13) with the symbolic degree of freedom

$$\overline{DF}=\tfrac{1}{2}\cdot DF\oplus\tfrac{1}{2}(DF+1) \tag{21}$$

where the meaning of multiplication and 'plus' is given by (19) and (20), and DF is given by (15).

In the case, when (16) is satisfied for all $i=1, 2, \dots k-n_0$ except of two for which x_i and x_j tend to zero, and the angle defined in the m-dimensional Euclidean space by the $\kappa(\underline{z}_1, \dots \underline{z}_{i-1}, \underline{z}_{i+1}, \dots \underline{z}_{k-n_0})$ and $\kappa(\underline{z}_1, \dots \underline{z}_{j-1}, \underline{z}_{j+1}, \dots \underline{z}_{k-n_0})$ hyper planes is φ, then the probability α is given by (13) with the symbolic degree of freedom

$$\overline{\overline{DF}}_\varphi=\frac{\varphi}{2\pi}\cdot DF\oplus\tfrac{1}{2}(DF+1)\oplus\frac{\pi-\varphi}{2\pi}(DF+2) \tag{22}$$

When in (22) φ tends to π, then $\overline{\overline{DF}}_\varphi$ tends $\overline{DF}$ given by (21), since in this case the two hyper planes are in coincidence.

ILLUSTRATION OF THE USE OF THE MODEL VALIDITY PROBABILITY IN SANDSTONE FORMATION

Our general experience related to the usage of the probability concept above shows that depending on the geological circumstances, on the quality of the field measurements, and on the applied models of interpretation very different relations can be obtained among the probability curves. Accordingly, the intention here is just to give some illustration about the practical application of those curves, for the determination of mineral composition of drilled through formation. We have the density, neutron porosity, acoustic, photoelectric index and the thorium and potassium spectral gamma ray logs ($m=6$) of one well in Hamada North area for complex interpretation. For the geophysical modelling of the drilled sandstone we define a set of models in the following way. Suppose that the basic part of the matrix is given by sand (Sd) and by dolomite (Dol) while the pore volume is filled up with interstitial water (Por). Additionally the matrix contains or may contain silt (Sil), different type of clays like illite (Ill) or kaolinit (Kao) and/or some iron bearing minerals like limonite (Lim), hematite (Hem). The response values for the measurements of this materials are given by laboratory data of the studied formation and by general tables, based on calculations and on laboratory measurements [5]. We can include a "mineral" into our mineral composition model only in the case if it has at least one significantly different value among its zone parameters from the zone parameter values of the sand and dolomite and from all the other included materials as well. Using this mineral set we define two component models {Por, Sd} and {Por, Dol}, for the sake of completeness, as the possible simplest formation; the basic clean sandstone without clay component {Por, Sd, Dol}; models for shaly sandstone {Por, Sd, Dol, Ill} or the same but Kao instead of Ill; models with iron hydroxides or oxides as {Por, Sd, Dol, Lim} or the same but Hem instead of Lim. We also define a larger number of five component models, combining the clay and iron bearing minerals. Thus we have a model set with different number of components; $k=2, 3, 4$ or 5, and according to formula (15) $DF=5, 4, 3$ or 2. Comparing for example the {Por, Sd, Dol} and the {Por, Sd, Dol, Ill} models often we experience that in a given depth point for the volumetric fraction of illite we get zero, $x_4=0$, and in this way as $n_0=1$, expression (15) reduces DF by one and the latter model compeletely simplifies to the three component model {Por, Sd, Dol} with exactly the same numerical value for α. But this possibility does not mean that the three component model is superfluous. Indeed, frequently we get that x_4 does exist ($x_4>0$) in the four component model but the probability α belonging to this model is less than that of the three component model. In this situation we conclude that even though the very existence of the illite component in the geological reality at the given depth can not be excluded, the illite component does not exist in a statistically significant quantity at the given depth.

Selecting the most appropriate model for interpretation we have a double aim. Our basic parameter for the oil industry and the most robust one is the porosity and we try to determine its value with the highest possible accuracy. Often we experience that different sandstone models give porosity percentages very near to each other while the probabilities belonging to the models differ significantly. This emphasizes again the high stability of porosity estimation from logs. In this situation we have enough reason to say that using a wider set of logs ($m=5, 6, 7$) we can determine not only the volumetric percentages of given minerals, but also decide certain alternatives related to the composition of the rock, therefore giving additional useful information about the formation for the reservoir engineer.

We note that in the calculation of α typical boundary distances defined by (16) and (18), belonging to the practically yet acceptable $\mu_b=1$ value, are $x_i=0.024$ for the porosity, $x_2 \cong x_3=0.23$ for sand and dolomite, $x_4=0.044$ for shale and $x_5=0.03$ for iron oxides.

In Fig. 2 we demonstrate the model validity probability for the above models as a function of the

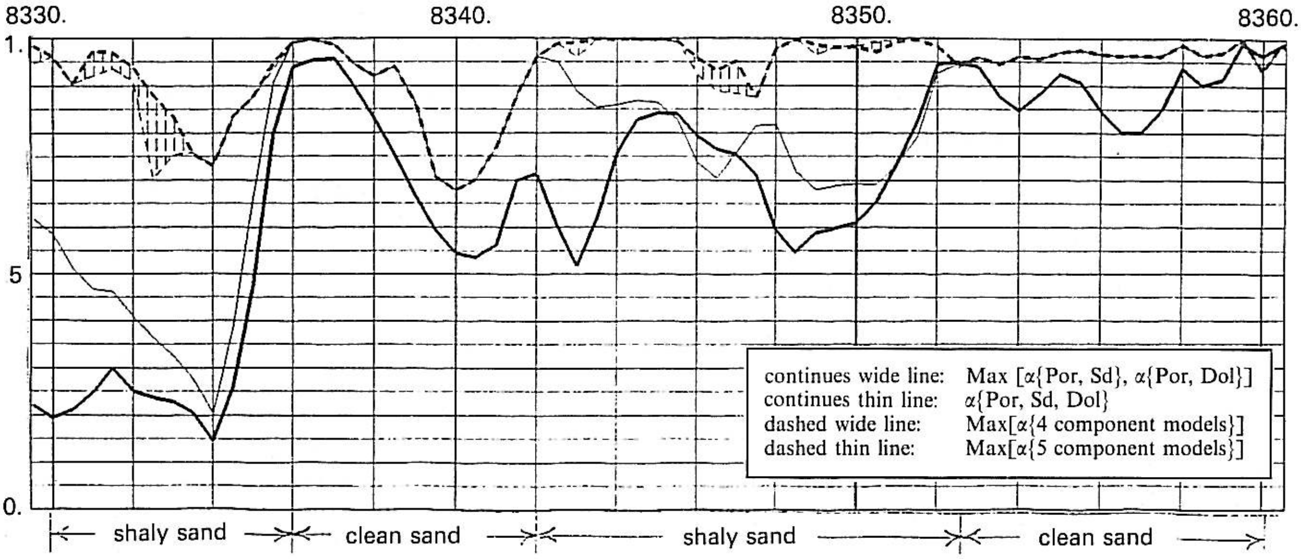
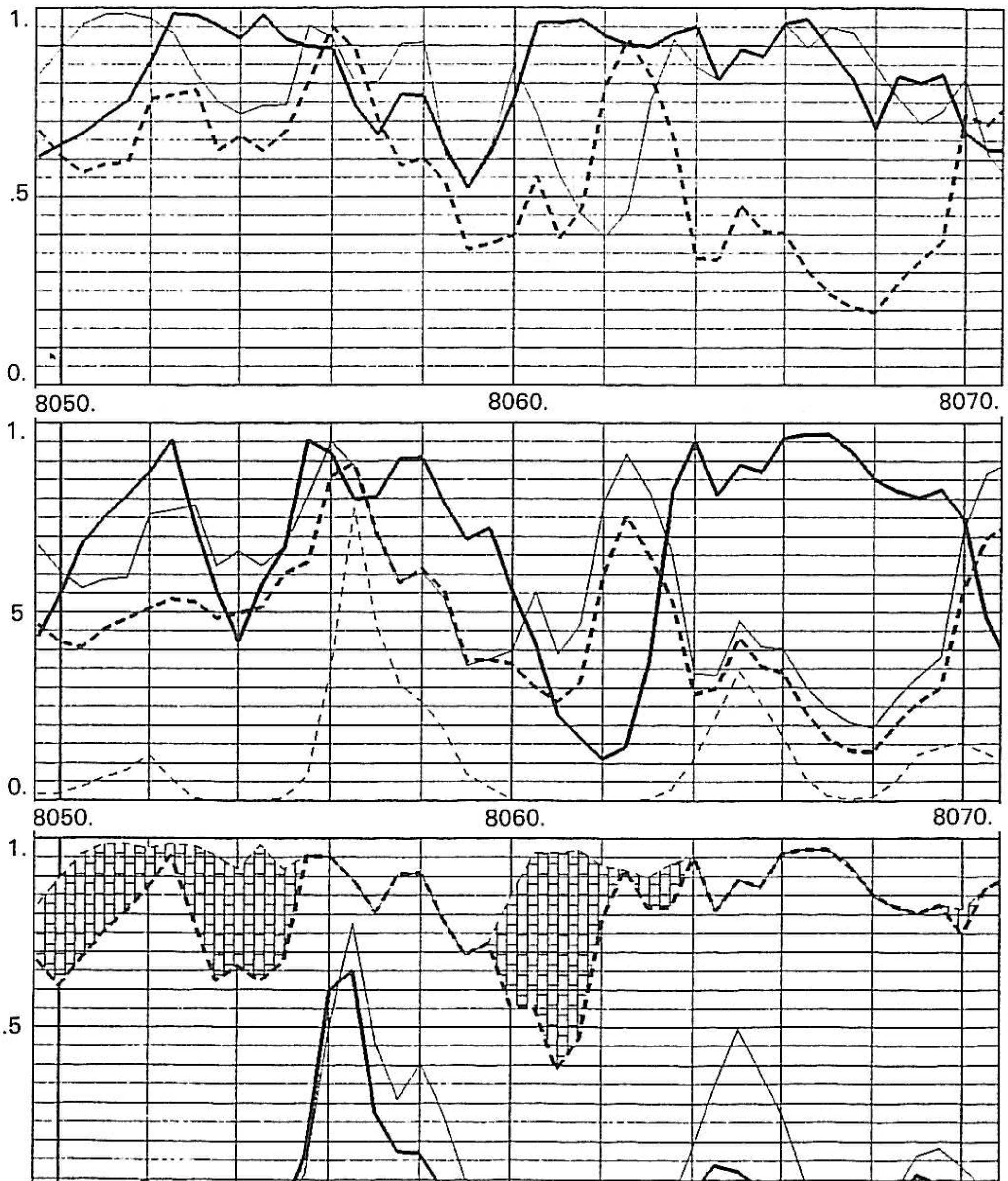


FIG. 2. Probability of validity of different sandstone models as a function of number of mineral components involved into the model in a well of sandstone formation.

number of components establishing the curves of maximum α value of all the two, four and five component models respectively, and using the only one three component model as well. In this interval the best two component model already gives relatively high probabilities, but almost everywhere the {Por, Sd, Dol} basic model works better. In the 8336.–8342. feet and in the 8352.5–8360. feet intervals any four or five component model gives the same probability as the basic model (because, in fact, they simplify to the basic model) which clearly indicates that in this intervals we have clean sand formation. In the two remaining intervals the best four component models give much higher probabilities than the basic model, thus here we have definitely shaly sand.

In the 8330.5–8333.5 feet and 8345.5–8347.5 feet intervals (shaded area in the figure) any fifth component definitely reduces the probability, because here the fifth component does not vanish but statistically is not significant.

More complex rock development is indicated in Fig. 3, where the two and three component models are failing. From 8050.0–8055.5 and from 8059.5–8064, feet (shaded area) without five components we cannot give complete explanation of our measurements, while on the remaining sections four component models give this explanation. In the mentioned five component intervals Kao plus Hem and Kao plus Sil appear in the formation. In the intervals 8057.–8059.5 and 8064.–8070. feet Kao is the type of the clay in formation, while in the interval 8055.5–



c, five component models: the basic model plus Kao, Sil for the continues wide line; the basic model plus Kao, Hem for the continues thin line; the basic model plus Ill, Hem for the dashed thin line.

b, four component models: the basic model {Por, Sd, Dol} plus Kao for the continues wide line; the basic model plus Ill for the continues thin line; the basic model plus Sil for the dashed wide line; and the basic model plus Hem for the dashed thin line

a, same curves as in Fig. 2

FIG. 3. Probability of validity of different sandstone models in the same well.

8057.0 the presence of Kao, Ill or Sil cannot be distinguished.

Table 1 shows in details the determined volumes

Table 1. Volume Fractions and the Probability of Validity Given by Different Sandstone Models at 8054 feet

Por	Sd	Dol	Kao	Ill	Si	Hem	Alfa
0.0971	0.6204	0.1028	0.1798	×	×	×	0.42
0.1581	0.3195	0.3438	×	×	0.1789	×	0.5
0.0923	0.6688	0.0358	×	0.2031	×	×	0.66
0.0923	0.6688	0.0358	×	0.2031	×	0	0.66
0.0975	0.7111	0	0.1718	×	×	0.0144	0.72
0.1251	0.4466	0.2236	0.0983	×	0.1064	×	0.92

and probabilities belonging to different models; at 8054 feet here the Kao, Sil combination alone give the definite maximum probability on a three feet long section, thus the presence of those minerals are proved; their presence at this depth effects the porosity determination relatively strongly.

CONCLUSION

- As far as we work with models of different dimensions for complex log interpretation, we have to develop the minimum incoherence concept into the maximum model validity probability concept.
- For the calculation of the model validity probability, despite of the presence of restriction conditions, for the most depth points we can use the Chi-distribution with a modified degree of freedom. However in certain cases more sophisticated boundary distributions are necessary for the same calculations.
- Using the model validity probability concept we can increase the accuracy of porosity estimation and decide certain alternatives related to the rock development as well.

ACKNOWLEDGEMENT

I would like to thank the management of the Petroleum Research Centre for the support of the present research. The author highly acknowledge the numerous consultations with Prof. Zoltan Barlai.

REFERENCES

[1] Vincze, I., 1968, Matematikai statisztika ipari alkalmazasokkal [Mathematical Statistics with Industrial Applications] Muszaki Konyvkiado Budapest (Technical Book Company).

[2] Renyi, A., 1966, Valoszinusegszamitas [Theory of Probability] Tankonyvkiado Budapest (Text Book Company).

[3] Szigeti G., 1994, Restriction Conditions in Well Log Interpretation, Petroleum Research Journal, Vol. 6., Tripoli, G.S.P.L.A.J.

[4] Szigeti, G., 1993, Improvement of the Accuracy of Statistical Well Log Interpretation Using the material Balance Equation, Petroleum Research Journal Vol. 5, Tripoli, G.S.P.L.A.J.

[5] Barlai, Z., Zilahi-Sebess, L., 1993, Report on the Sheet Silicates and their Effects on the Basic Reservoir Parameters of Shaly Sandstone and Barrier Shales, Petroleum Research Centre Internal Report, Tripoli, G.S.P.L.A.J.