Fast Porosity Estimation by Principal Component Analysis

István Elek
Hungarian Hydrocarbon Institute
Budapest, Hungary

Summary

Based on experience with several wells, we supposed that the first principal component of porosity logs (perhaps taking SP and GR logs into consideration) is related to effective porosity. This brief paper outlines the steps we followed in examining the first principal component, and presents material from one well processed by Principal Component Analysis (PCA is described briefly in Elek, 1988, and Morrison, 1979). The well and the field are well-known in Hungary (shaly sand).

We supposed that the first principal component of porosity logs (perhaps taking SP and GR logs into consideration) is related to effective porosity. We based this supposition on experience, having examined several wells and seeing it certified in each case. Because showing material from each well we processed would make this paper far too long, we introduce only one well processed by Principal Component Analysis (PCA is described briefly in Elek, 1988, and Morrison, 1979). The well and the field are well-known in Hungary (shaly sand).

Briefly outlined, the investigation includes the following steps:

- Selecting the appropriate logs (it would be unreasonable to consider caliper, temperature, or other logs that have little to do with porosity.)
- Computing the first principal component of the input logs. We examine the correlation matrix of the input logs, the eigenvalues and eigenvectors of the correlation matrix. By this means the mutual dependence of the input logs can be established, even the variance content of the input logs and a common background variable (if it exists).
- Examining the dependence of the results achieved by a traditional method, the porosity (and shale volume) and the first principal component with cross-plots and regression analysis.
- Using the coefficients obtained from the regression to estimate porosity (and clay content). The results are compared graphically to the porosity (and clay content) estimated by traditional methods. The coefficients of the regression line can be substituted with the porosity values from two known depth points (i.e., we should know the porosity at least in two depth points. If we know porosity in more points the situation is more accurate.) In this way the first principal component can be transformed into porosity range (similar to a calibration process).

Example

The following section examines these steps in more detail.

We used the following logs :

Well.

Measured logs.

Density (DEL).
Sonic (ATL).
Neutron (FINL).
Gamma ray (TG).
Spontaneous potential (SP).

Interpretation results (traditional).

Effective porosity (PHIE).
Clay content (VCL).

Logs from PCA.

1st principal component	C1
2nd principal component	C2
3rd principal component	C3
4th principal component	C4
5th principal component	C5

Effective porosity from C1 - PHIEC1.
Clay content from C1 - VCLC1.

The measured logs can be seen in Figure 1.

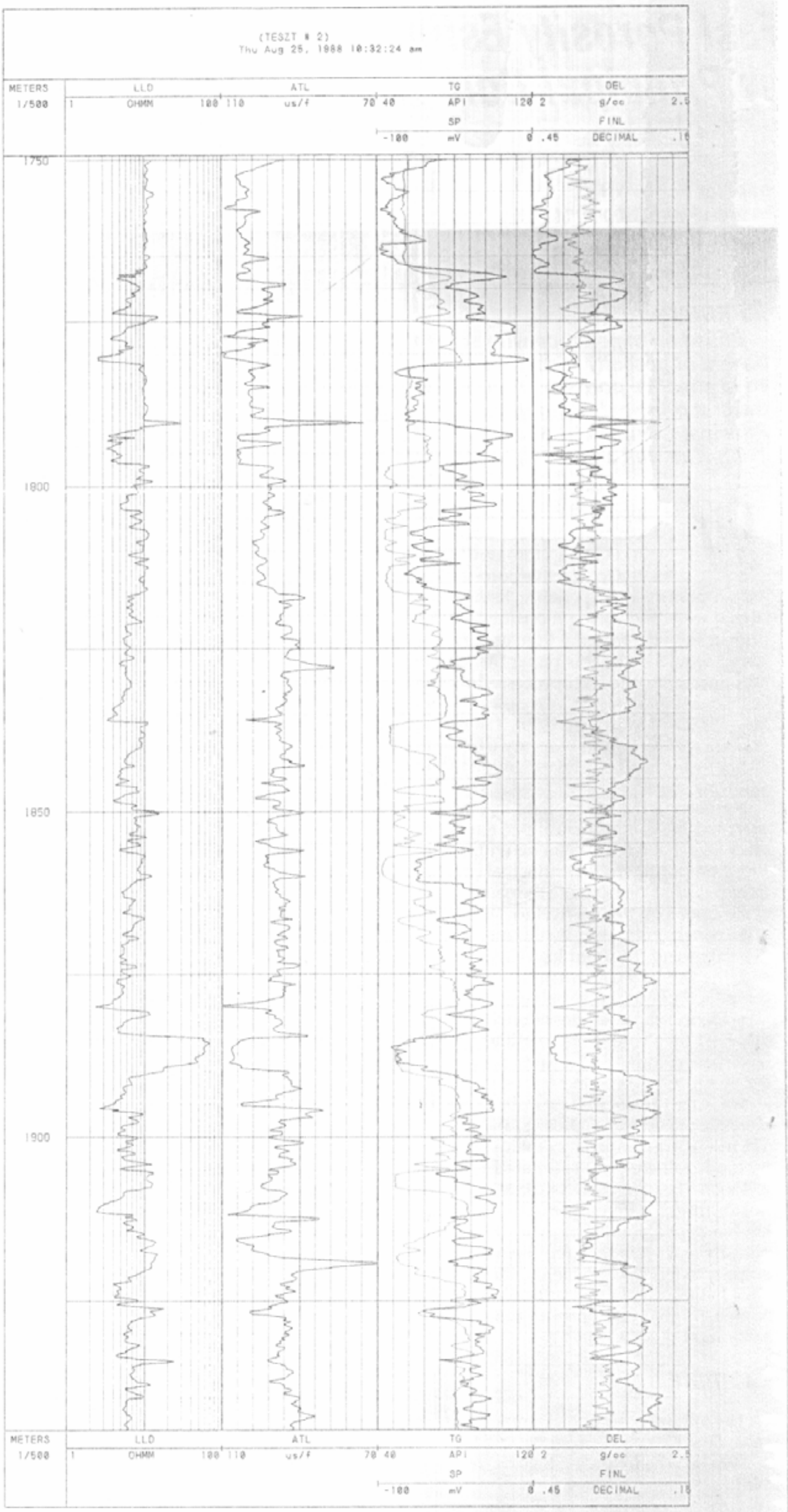
Following the steps outlined above, we carried out PCA of the chosen logs, and examined the correlation matrix of our logs (Figure 2). Figure 3 shows the principal components of the input logs. The first principal component concentrates a considerable proportion of variances.

The eigenvectors (U1, U2,...), which give the coefficients of the weighted average, show the e.g. concerning the first principal component (U1 eigenvector belongs to it) the contribution of which log is the most important in the first principal component. (We can obtain similar information examining the factor matrix).

An interesting fact can be seen: DEL, GR, and SP logs have the greatest roles, while the contribution of the FINL and ATL logs seems to be less important in the first principal component (In the second principal component, there is an entirely opposite situation.)

It is not surprising that the DEL, TG, and SP logs have the most important roles, since this corresponds to the interpreters' experience; in shaly sands almost "everything" can be determined

Figure 1: Measured logs used in the investigation.



INPUT CURVES	OUTPUT CURVES
DEL	C1
FINL	C2
ATL	C3
TG	C4
SP	C5

DATA NUMBER: 976

CORRELATION MATRIX:

1.0000	-.4006	-.6884	.7410	.6566
-.4006	1.0000	.5524	-.0538	-.1697
-.6884	.5524	1.0000	-.3488	-.2652
.7410	-.0538	-.3488	1.0000	.6751
.6566	-.1697	-.2652	.6751	1.0000

EIGENVALUES:

LAMBDA(1)=	2.897457
LAMBDA(2)=	1.213836
LAMBDA(3)=	.493214
LAMBDA(4)=	.262114
LAMBDA(5)=	.133382

VARIANCIAPART OF FIRST PRINCIPAL COMPONENT: 57.949%

EVIGENVECTORS:

U1	U2	U3	U4	U5
.5558	.0291	-.1793	.0360	.8104
-.2971	.6742	-.5633	-.3672	.0712
-.4356	.4452	.5437	.4104	.3848
.4639	.4399	-.1800	.6305	-.4018
.4448	.3911	.5680	-.5459	-.1692

FACTORMATRIX:

.9461	.0321	-.1259	.0184	.2960
-.5057	.7428	-.3956	-.1880	.0260
-.7415	.4904	.3818	.2101	.1405
.7896	.4846	-.1264	.3228	-.1468
.7572	.4309	.3989	-.2795	-.0618

Figure 2: *Correlation matrix, Case A.*

from the DEL, TG, and SP logs. Our experience can be explained in another way: the first principal component contains the most important factors that characterize shaly sands.

Next, we compared the effective porosity (PHIE) to the first principal component (C1) and the clay volume (VCL) to (C1) by means of cross-plots. Figure 4 shows the PHIE-C1 cross-plot and linear regression line. The equation of this line is:

$$PHIE=AO + AI*CI \quad (1)$$

where AO = 0.13679, AI= -0.063678
The correlation coefficient: r = 0.93299.

Figure 5 shows the VCL - C1 cross-plot and the linear regression line. The equation of this line is:

$$VCL = BO + BI*CI \quad (2)$$

where BO = 0.55177, BI = 0.16759
The correlation coefficient: r = 0.91106

The figures show that the dependence is very close between CI and PHIE (or VCL).

It can be seen that the first principal component follows the effective porosity. It should be noted that we do not consider PHIE or VCLs log etalons because these basically come from density-neutron cross-plots. These logs were computed without taking every log into consideration. In the first principal component, all the logs are included. Therefore the role of the PHIE and VCL logs is mainly to help us decide whether the first principal component could follow the porosity (and clay volume).

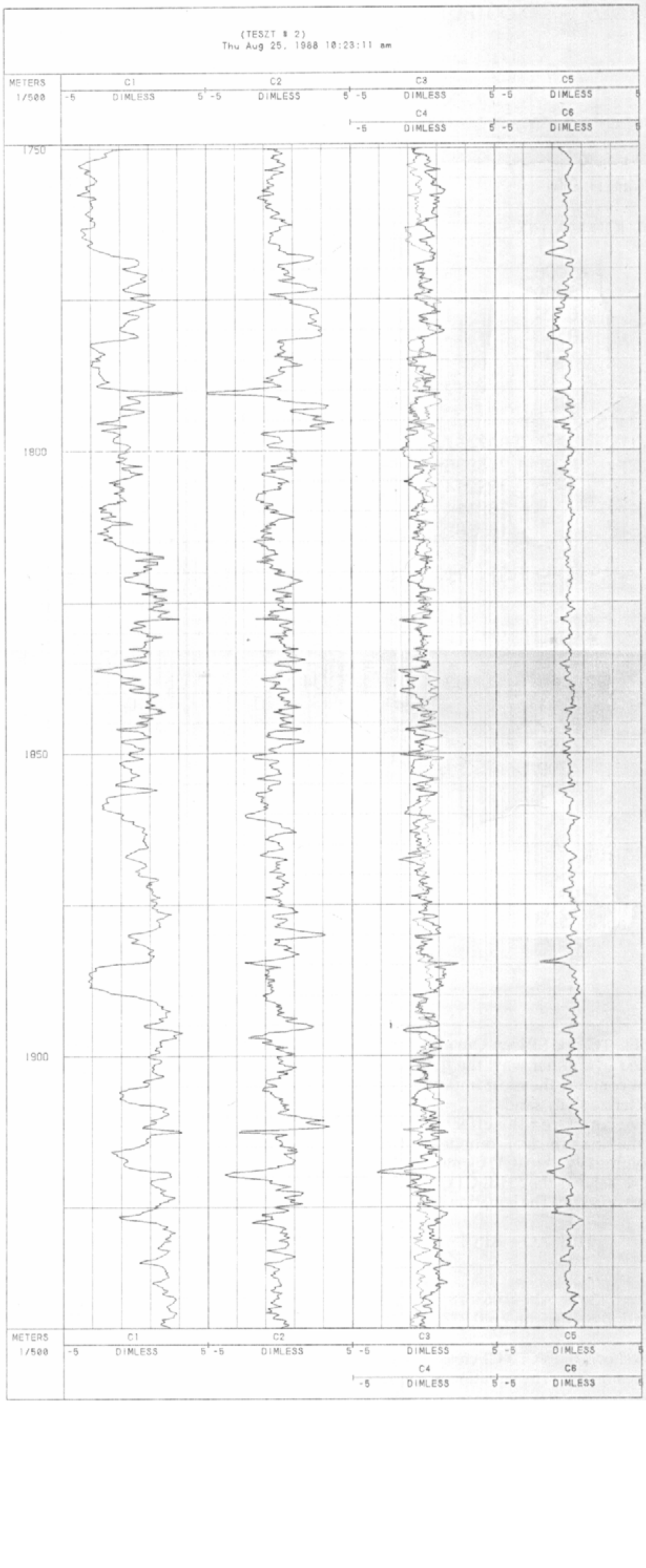


Figure 3: Principal components of the input logs.

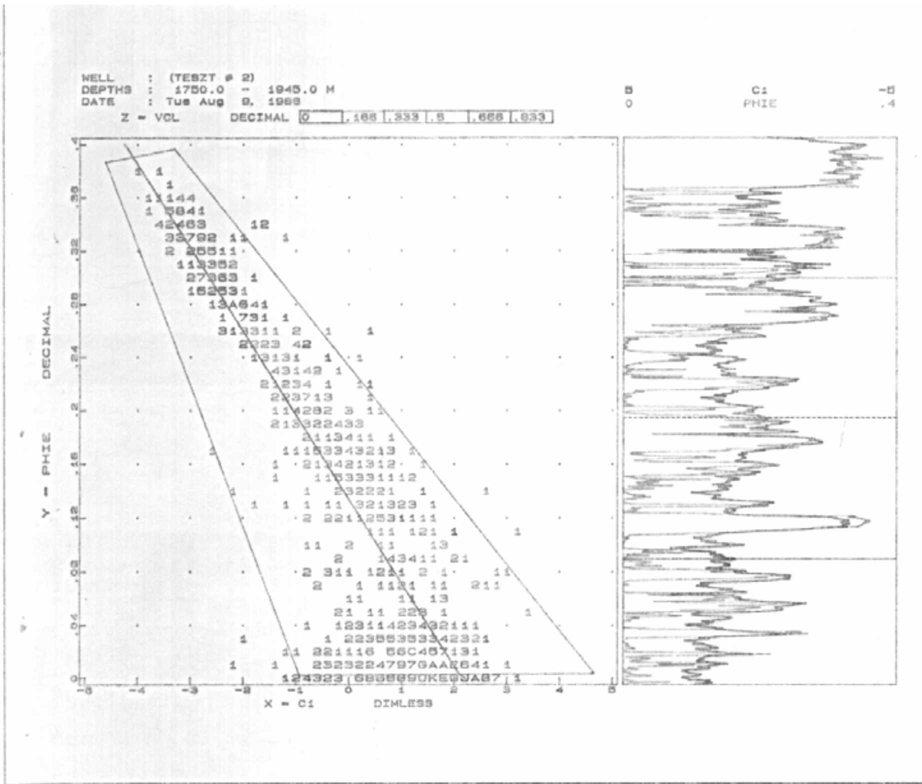


Figure 4: PHIE-C1 cross-plot and linear regression line.

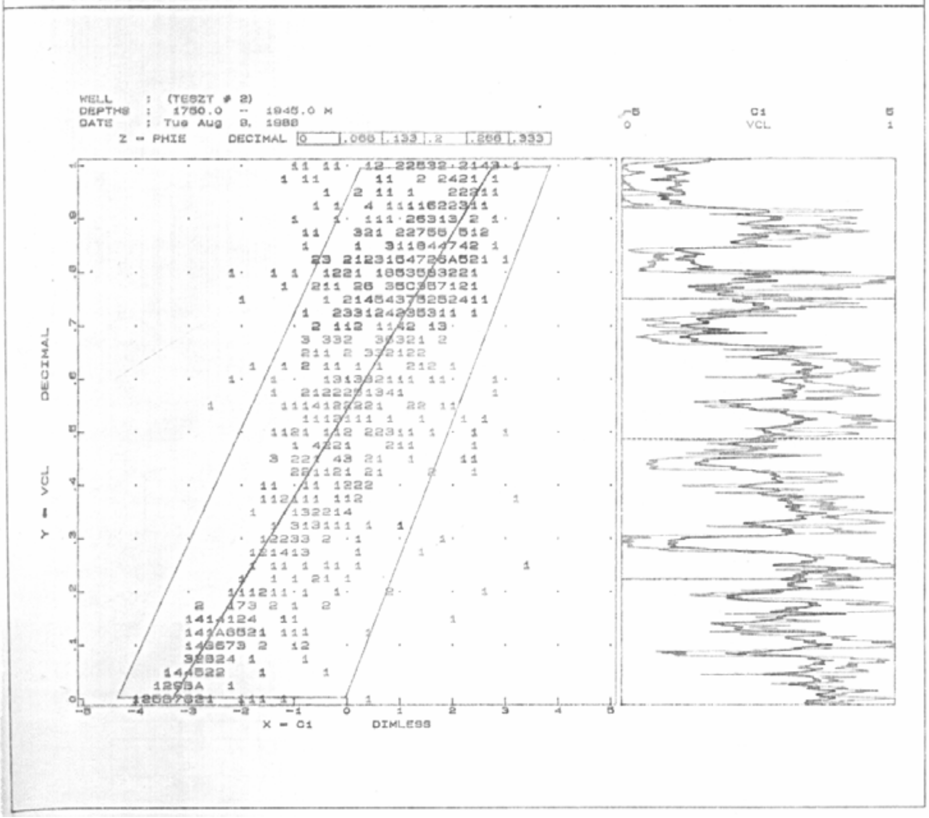


Figure 5: VCL - C1 cross-plot and the linear regression line.

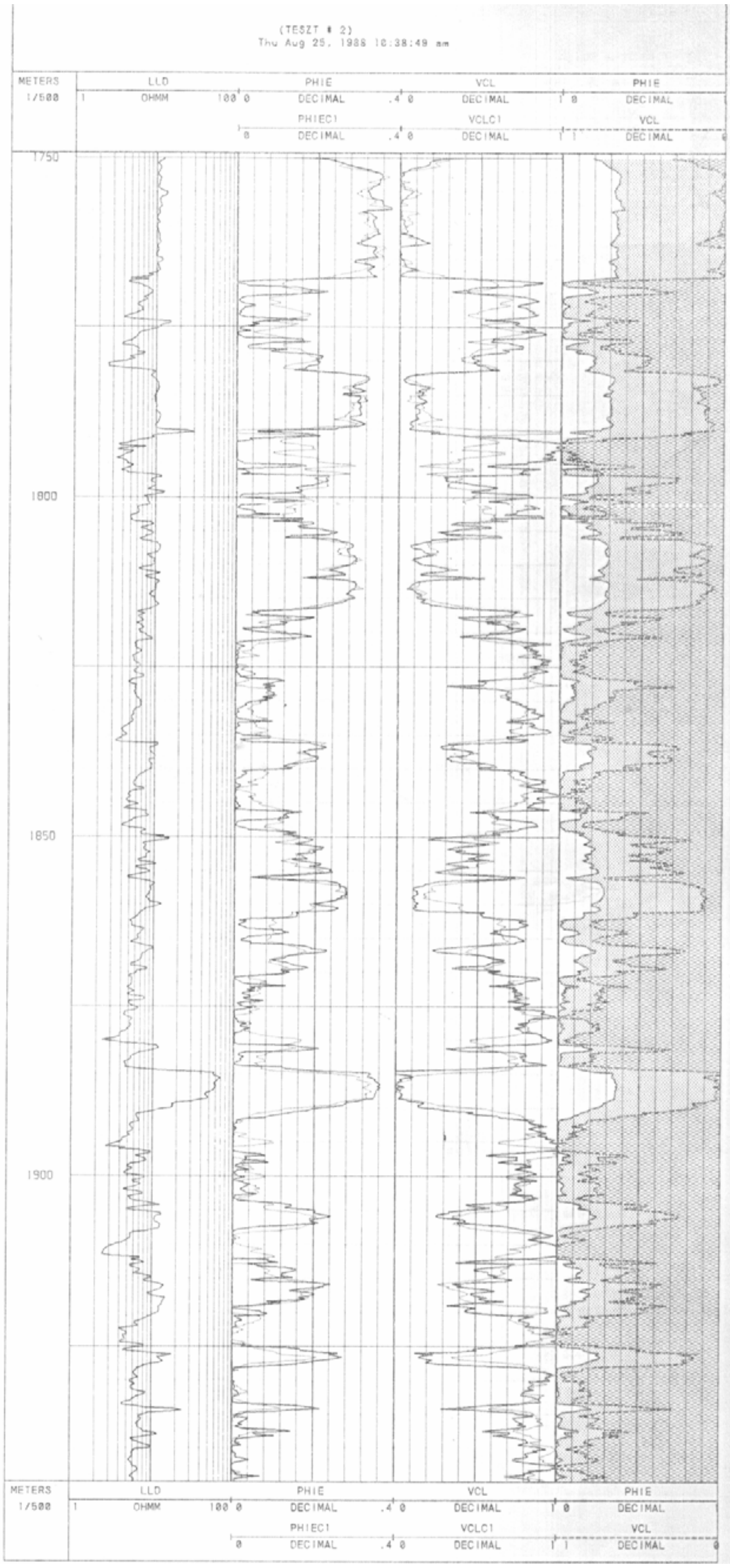


Figure 6: *Graphic comparison of PHIE and PHIECI (or VCL and VCLCI).*

We may compute an estimated porosity from Equation 1 and a clay volume from Equation 2.

$$PHIECI = AO+AI*CI \quad (3)$$

$$VLCCI = BO+BI*CI \quad (4)$$

Comparing PHIE and PHIECI (or VCL and VCLCI) graphically (Figure 6) shows that the similarity is rather closed.

Figure 7 shows the PHIE-VCL cross-plot. The correlation between these logs closed.

We examined a simple case using only density and neutron logs. We carried out the principal component analysis of density and neutron logs (C1', C2'). Having known the geometrical meaning of PCA, the elongation of first principal component is porosity-directed while the second principal component is "lithology"-directed. A neutron-density cross-plot can be shown on the Figure 8. C1'-C2' cross-plot can be shown in Figure 9. The rotation of PCA can be seen. Our experience is that the first principal (C1') is porosity-following, but not so good as the previous case where every log was used.

Figure 10 shows correlation matrix, eigenvectors, and factor matrix. Figure 11 shows the PHIE-C1' cross-plot.

AO = 0.152226 AI = -0.091152 r = 0.8595

In Figure 12, VCL-C1' cross-plot can be seen.

BO = 0.54367 B1 = 0.23026 r = 0.7851

Comparing Case A to Case B illustrates that the first principal component coming from the porosity-following logs + natural gamma ray + spontaneous potential is better suited for porosity estimation than Case B.

Discussion

The correlation between the first principal component and effective porosity is rather closed. If the effective porosity is known at least in two depth-points (from cores or other way), the first principal component can be converted into effective porosity. (This is a calibration procedure). If we know the effective porosity in more than two depth points, the calculation is more accurate. The clay volume can be obtained in a similar fashion.

This method is good for fast porosity estimation in a field. If the field engineer can establish the porosity in two points, he or she can obtain the effective porosity log (and clay content). The computation time is some seconds on an IBM PC/AT. Figure 13 shows the result of this calculation. We tried this method for more than 20 wells, but only in shaly sand. We are going to extend this method for more complicated reservoirs, e.g., metamorphits or fractured reservoirs.

References

Elek, I., 1988, Some applications of principal component analysis: Well-to-well correlation, zonation: *Geobyte*, May, v.3, n.2.

Morrison, D.F., 1979, Multivariate statistical methods; McGraw-Hill, New York.

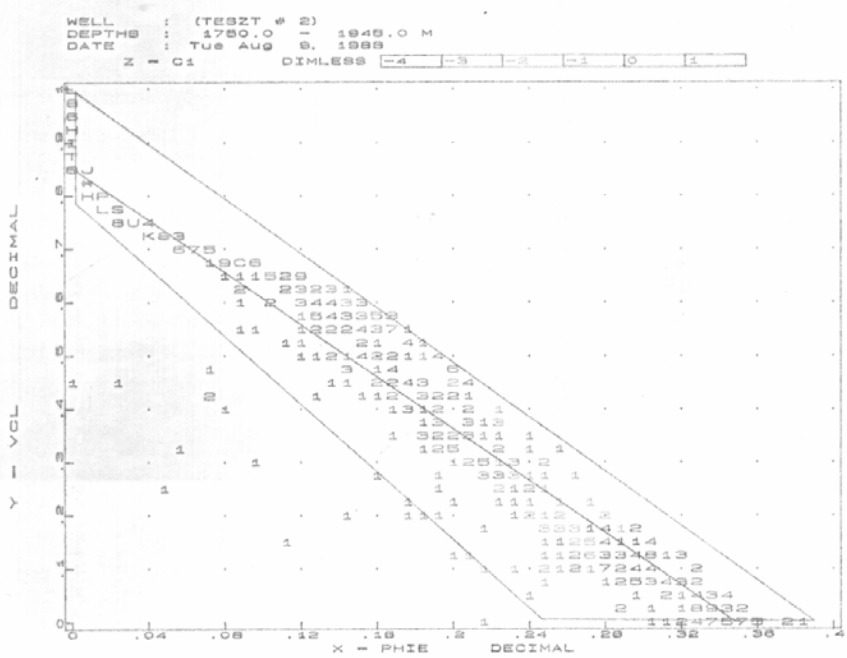


Figure 7: PHIE-VCL cross-plot.

Figure 8: Neutron-density cross-plot.

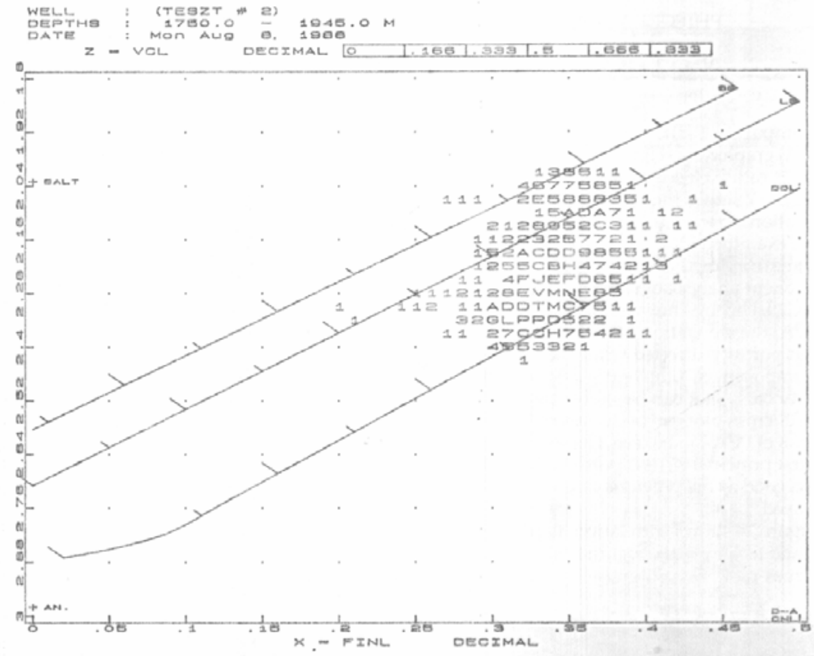
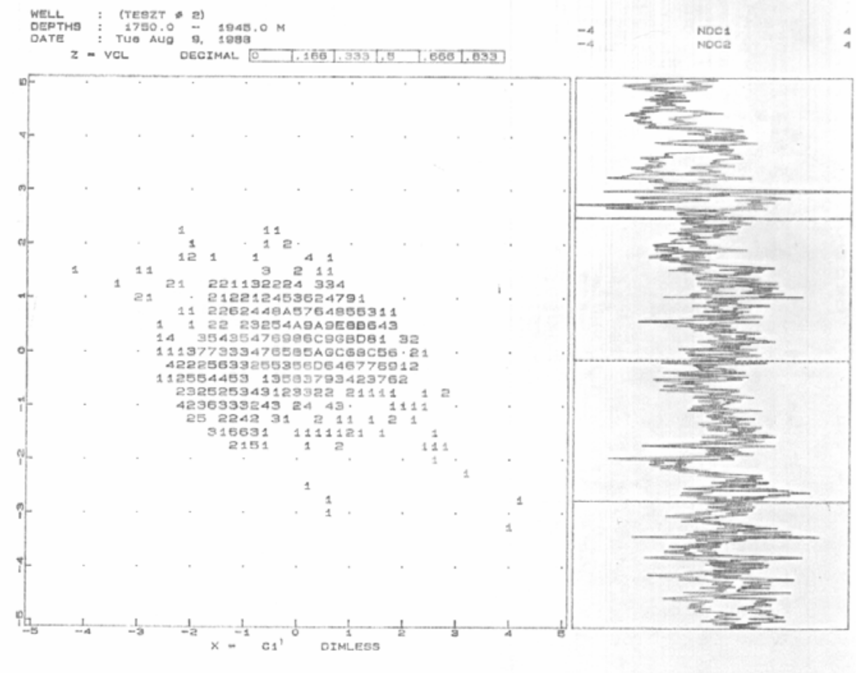


Figure 9: C1'-C2' cross-plot.



INPUT CURVES	OUTPUT CURVES
DEL	NDC1
FINL	NDC2

DATA NUMBER: 976

CORRELATION MATRIX:

1.000	-.4006
-.4006	1.0000

EIGENVALUES:

LAMBDA(1)= 1.400605
LAMBDA(2)= .599396

VARIANCIAPART OF FIRST PRINCIPAL COMPONENT:
70.030%

EIGENVECTORS:

U1	U2
.7071	.7071
-.7071	.7071

FACTORMATRIX

.8368	.5474
-.8368	.5474

Figure 10: *Correlation matrix, Case B, using only density and neutron logs.*

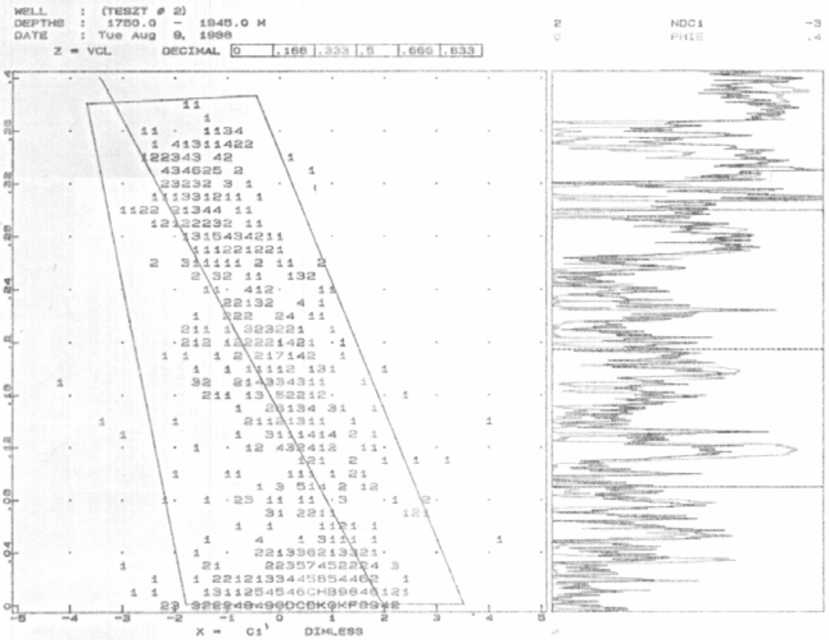


Figure 11: *PHIE-C1' cross-plot.*

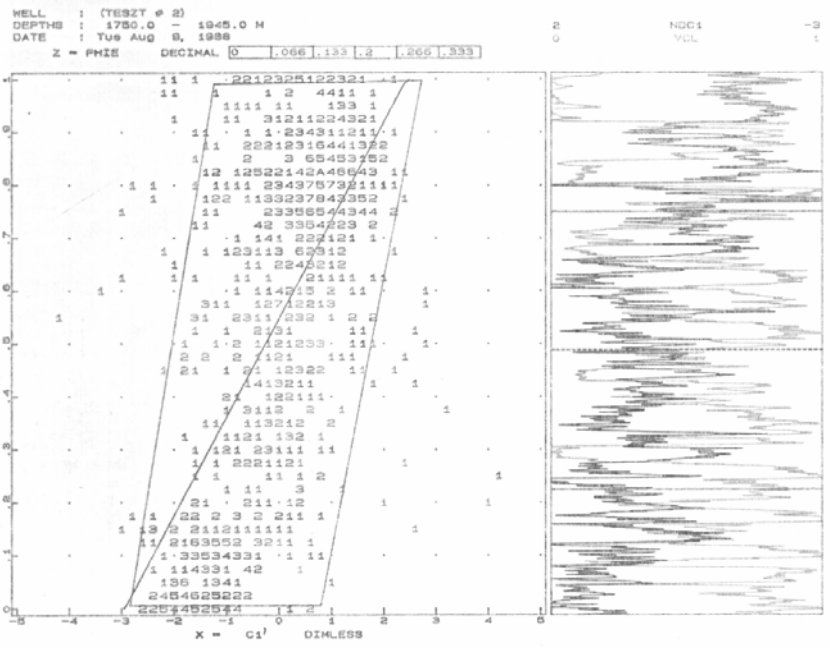


Figure 12: *VCL-C1' cross-plot.*

Figure 13: *Results of using the calculation.*