KGS
OF
97-85

DATA ANALYSIS FOR PETROPHYSICISTS

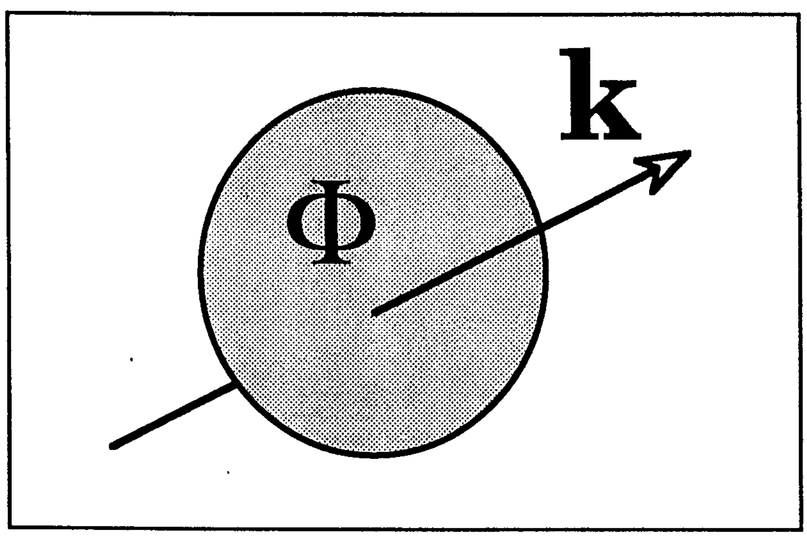
John H. Doveton

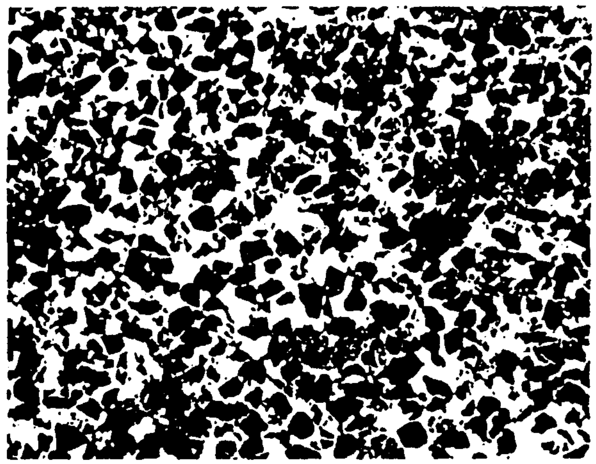
Kansas Geological Survey
Open-file Report

Disclaimer

CONTENTS

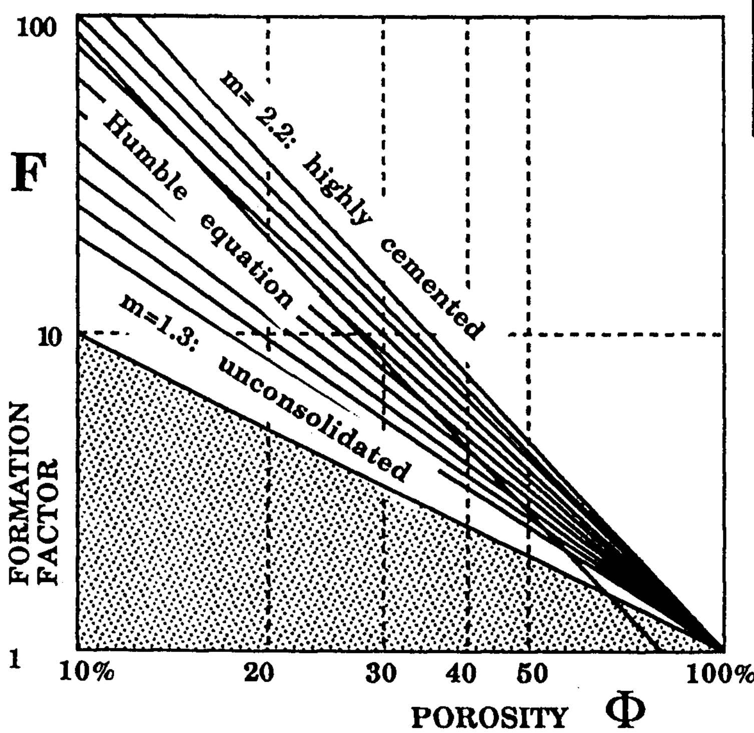
POROSITY & PERMEABILITY IN SIMPLE AND COMPLEX LITHOLOGIES





2-dimensional view of pore network, as seen in thin-section. Bartlesville Sandstone, Kansas

3-dimensional metallic pore cast from a consolidated sandstone (from Collins, 1961).



FLOW OF ELECTRICAL CURRENT THROUGH CLEAN SANDSTONES

Archie equation :

$$F = 1/\Phi^m$$

where :

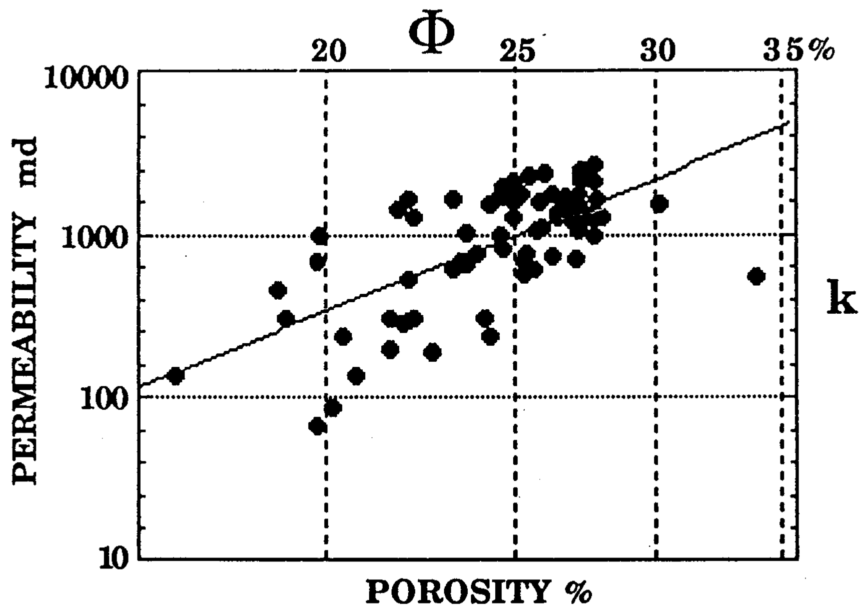
$$F = R_0 / R_w$$

Possible relationship between formation factor and permeability :

$$\text{e.g. } k = \Phi D^2/(32T) \propto D^2/F$$

where T is tortuosity and D is a characteristic length

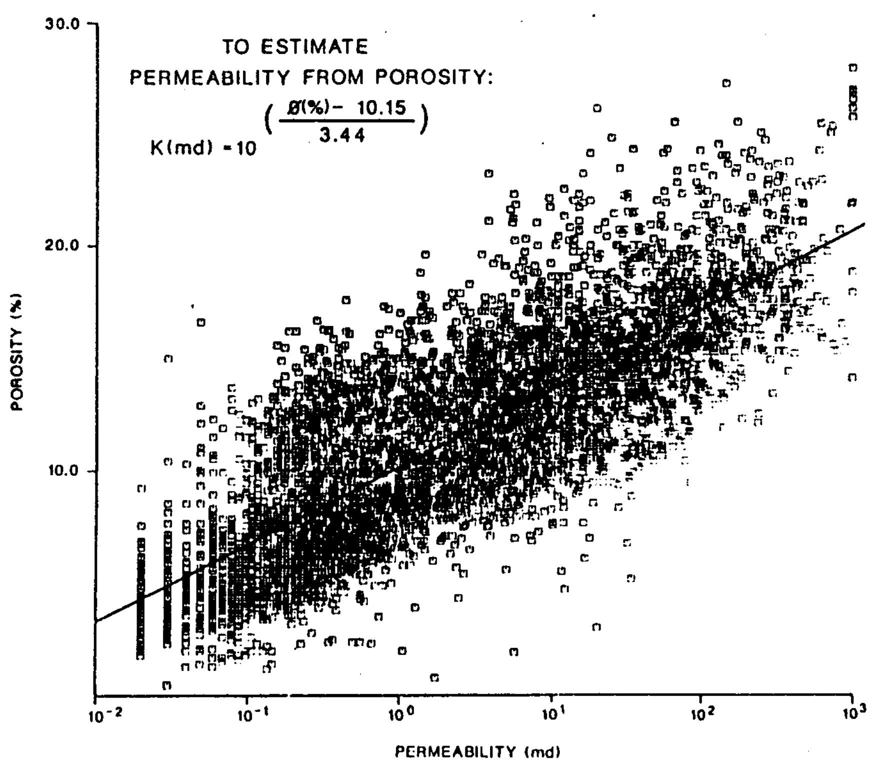
(Kimminau and Schwartz, 1987)



Simple relationship :

$$k = a\Phi^{b} \quad \therefore \quad \log k = \log a + b \log \Phi$$

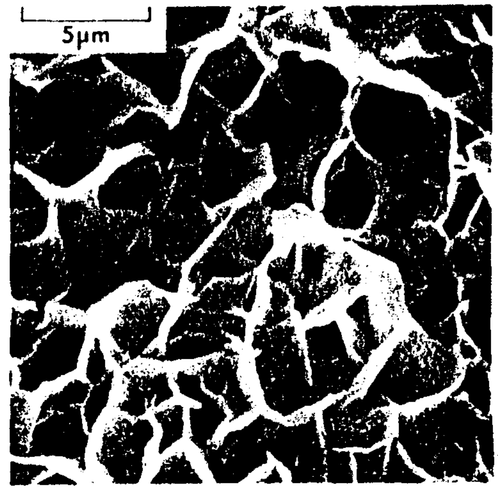
Dakota Formation, Kansas (core data from Preston, 1959)



Porosity - permeability relationship for 5419 core samples of Norphlet sandstone (Upper Jurassic), Alabama.
from Dixon et al (1989)

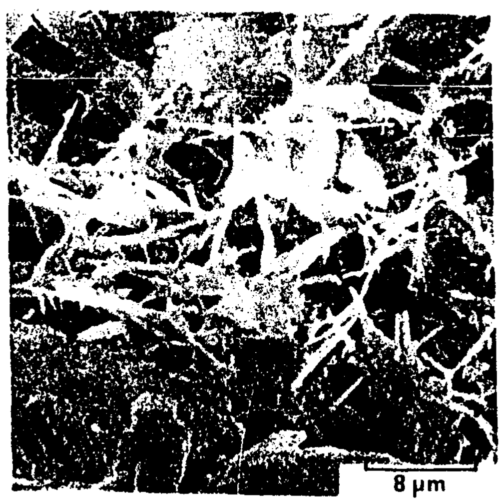
LITTLE SHOP OF HORRORS: CLAY MINERALS & THEIR ROLE IN FORMATION DAMAGE

KAOLINITE
patchy discrete particles
Migration of fines which can block pore throats



SMECTITE
pore - lining / bridging
Highly sensitive to fresh water, resulting in clay swelling and microporosity

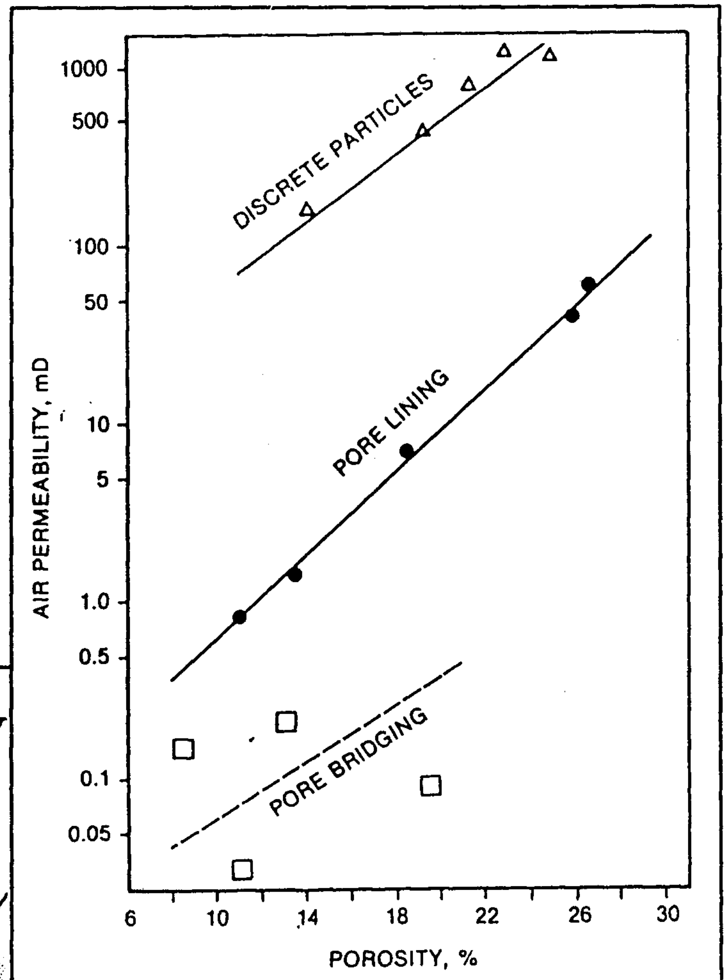
CHLORITE
pore -lining / bridging
Very acid - sensitive, with production of iron hydroxide precipitate to fill pore throats



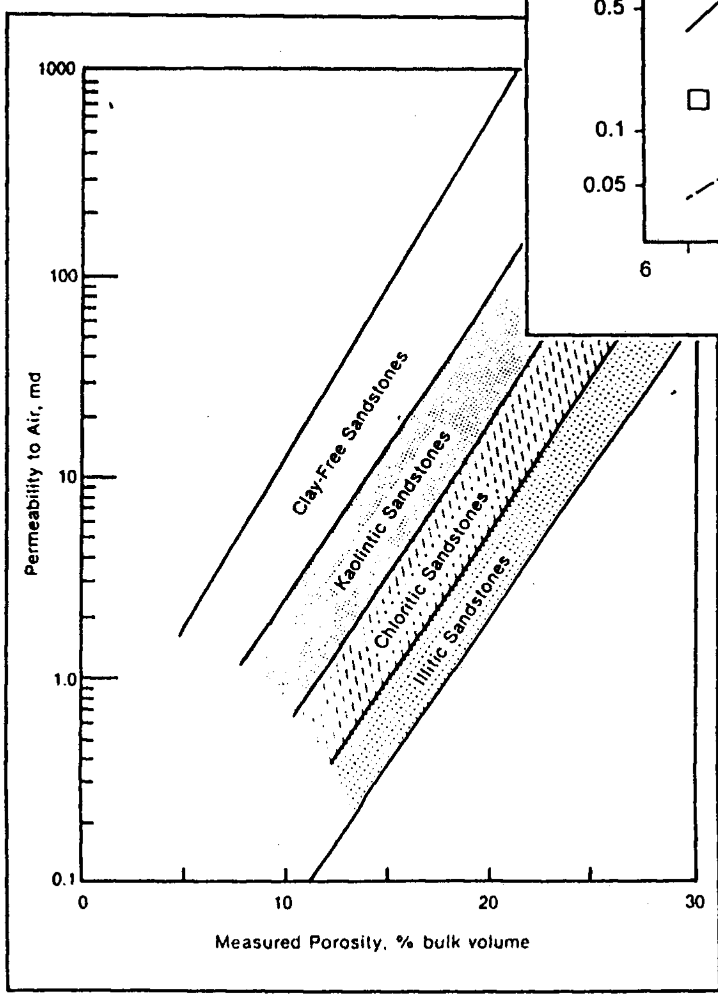
ILLITE
Pore - lining / bridging
Microporosity formation, some migration of fines

SEM photos by Syed Ali in Bigelow (1985)

INFLUENCE OF CLAY MINERALS / CLAY MORPHOLOGIES ON POROSITY - PERMEABILITY RELATIONSHIPS IN SANDSTONES

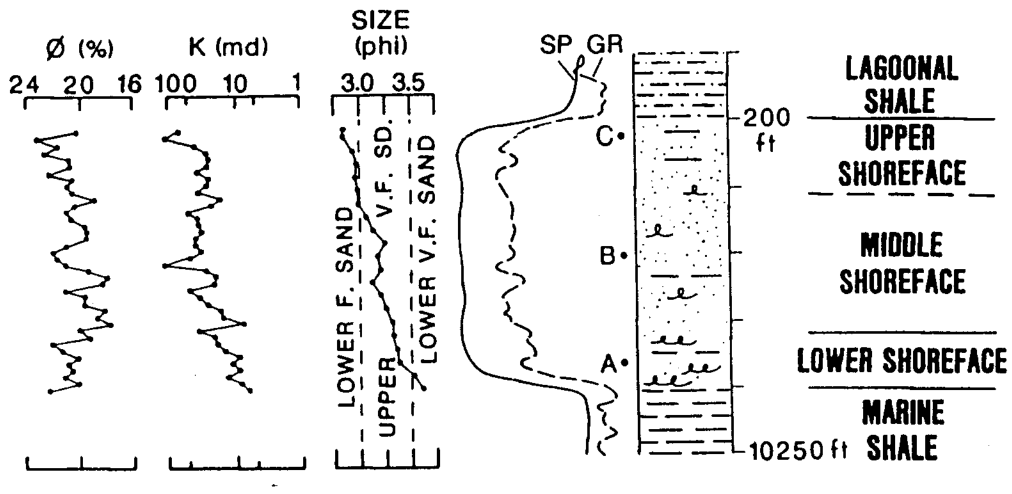


(Neashan, 1977)

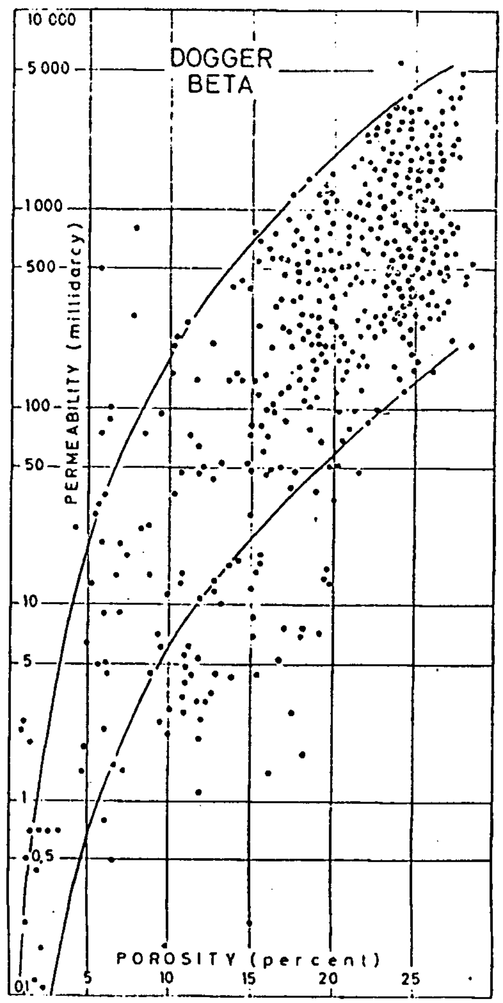


(Wilson, 1982)

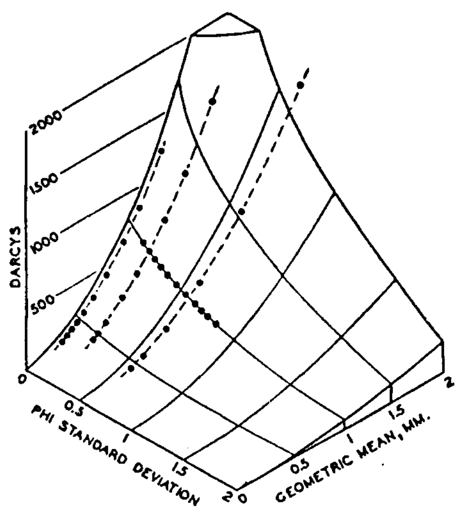
INFLUENCE OF CLASTIC GRAIN SIZE ON PERMEABILITY IN SANDSTONE RESERVOIRS



Porosity, permeability and grain-size variation in a Lower Eocene bar sandstone, Louisiana (Self et al, 1986)



Kozeny-Carman specific -surface contours as indicators of grain size on $\Phi - k$ plot (from Fuchtbauer, 1967)



Grain size and sorting control of permeability in sands (from Krumbein and Monk, 1942)

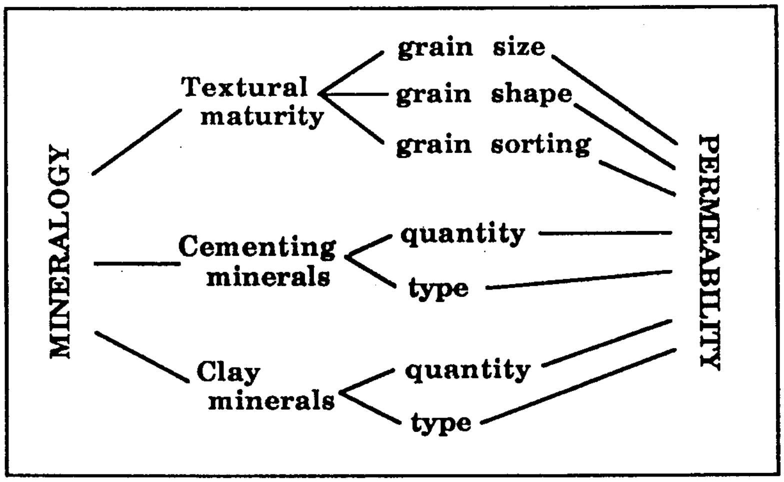
Kozeny - Carman equation :

$$k = \frac{A\Phi^3}{(1-\Phi)^2 S^2}$$

where A is a constant and S is the specific surface area i.e. ratio of surface area to volume of framework solid
For a sphere, S = 6/d
where d is the sphere diameter

HERRON MODEL FOR PERMEABILITY ESTIMATION IN CLASTICS BASED ON POROSITY AND MINERALOGY

Kozeny - Carman equation : $$K = \frac{A\,\Phi^3}{(1-\Phi)^2 S^2}$$



Herron equation : $$K = \frac{a_f\,\Phi^3}{(1-\Phi)^2}\exp\left(\sum B_i M_i\right)$$

Rewritten :

$$\log K = A_f + 3\log\Phi - 2\log(1-\Phi) + \sum B_i M_i$$

where A_f is the textural maturity term, M_i is the abundance of the ith mineral and B_i is a constant

Based on data from Venezuela, California and Oklahoma :

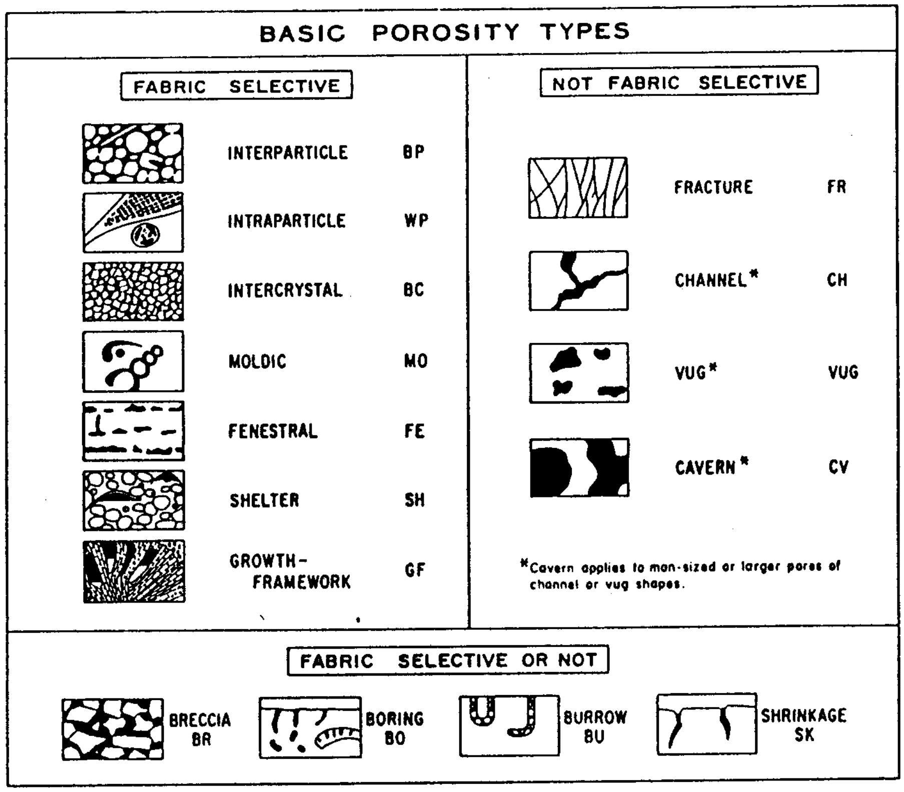
$$A_f = 4.9 + 2\,\text{Feldspar}_{max}$$

B_i values :

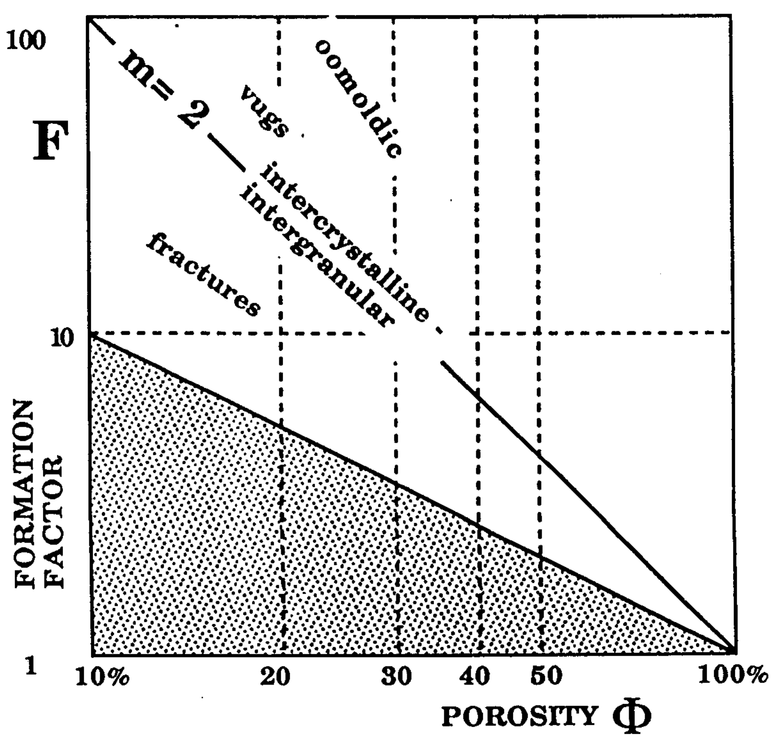
CLAYS		CEMENTS		FRAMEWORK MINERALS	
Kaolinite	-4.5	Calcite	-2.5	Quartz	0.1
Illite	-5.5			Feldspars	1.0
Smectite	-7.5				

Reference : Herron (1987)

POROSITY AND RESISTIVITY IN CARBONATES



from Choquette and Pray, 1970



FLOW OF ELECTRICAL CURRENT THROUGH CARBONATES

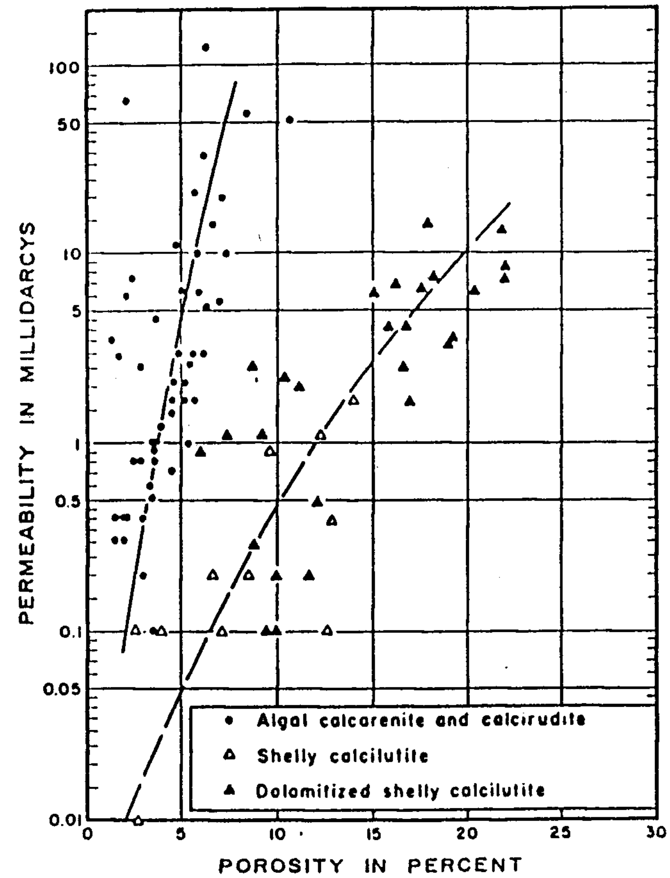
Archie equation :

$$F = 1/\Phi^{m}$$

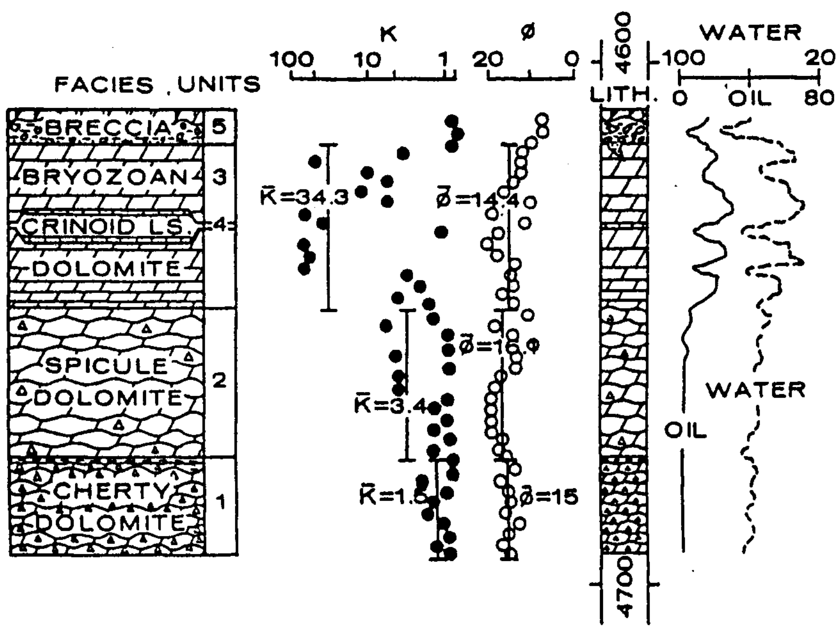
where :

$$F = R_o / R_w$$

CONTROL OF POROSITY - PERMEABILITY PATTERNS IN CARBONATES BY SEDIMENTARY FACIES AND DIAGENESIS



after Choquette and Traut, 1963



from Ebanks, Euwer and Nodine - Zeller, 1977

Blank

CASE STUDIES

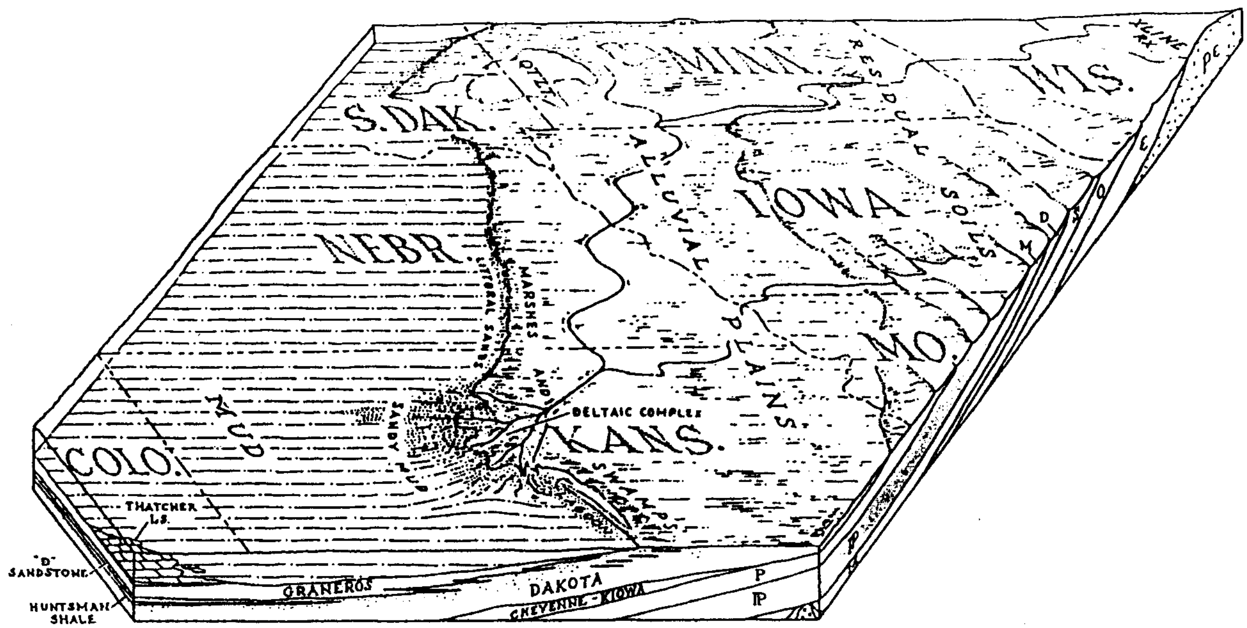
* CLASTIC CASE STUDY LOGS
  Cretaceous and Lower Permian
  KGS Braun #1 NENENE 30-12S-18W
  Ellis County, Kansas

* CARBONATE CASE STUDY LOGS
  Chase Group (Lower Permian)
  Mobil Brown #1-2 CNW 11-35S-37W
  Stevens County, Kansas

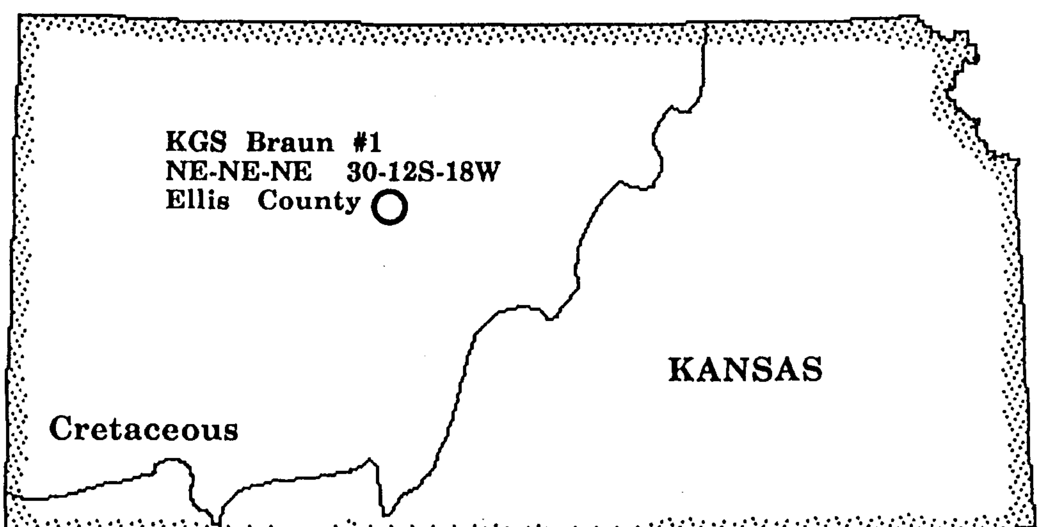
Blank

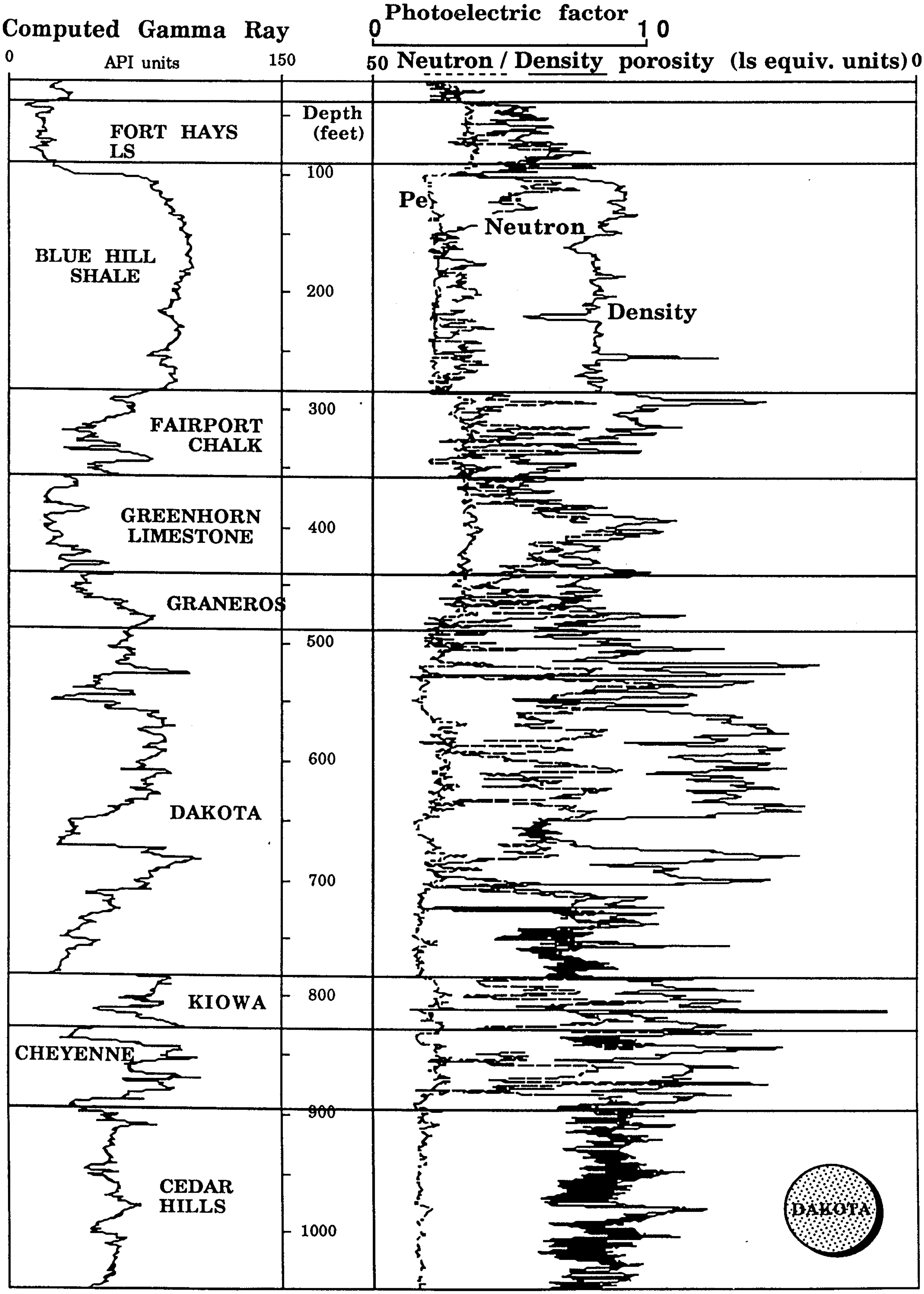
DAKOTA

CLASTIC CASE STUDY
Cretaceous -Lower Permian sandstones and shales (with some limestones)

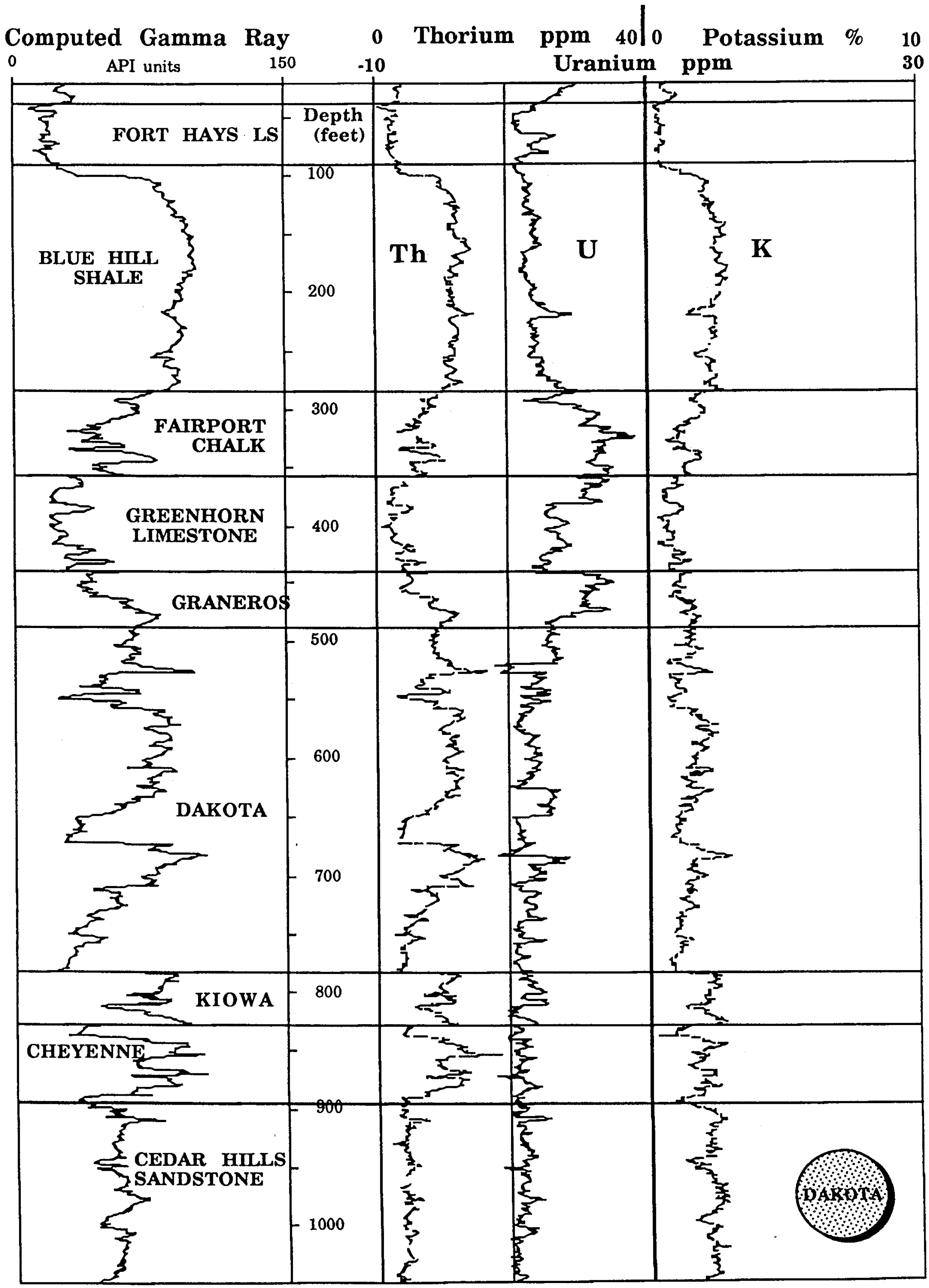


(from Hattin, 1967)





Computed Gamma Ray
0
API units
150
Photoelectric factor
0
10
50 Neutron / Density porosity (ls equiv. units) 0
Depth (feet)
100
200
300
400
500
600
700
800
900
1000
FORT HAYS LS
BLUE HILL SHALE
FAIRPORT CHALK
GREENHORN LIMESTONE
GRANEROS
DAKOTA
KIOWA
CHEYENNE
CEDAR HILLS
Pe
Neutron
Density
DAKOTA

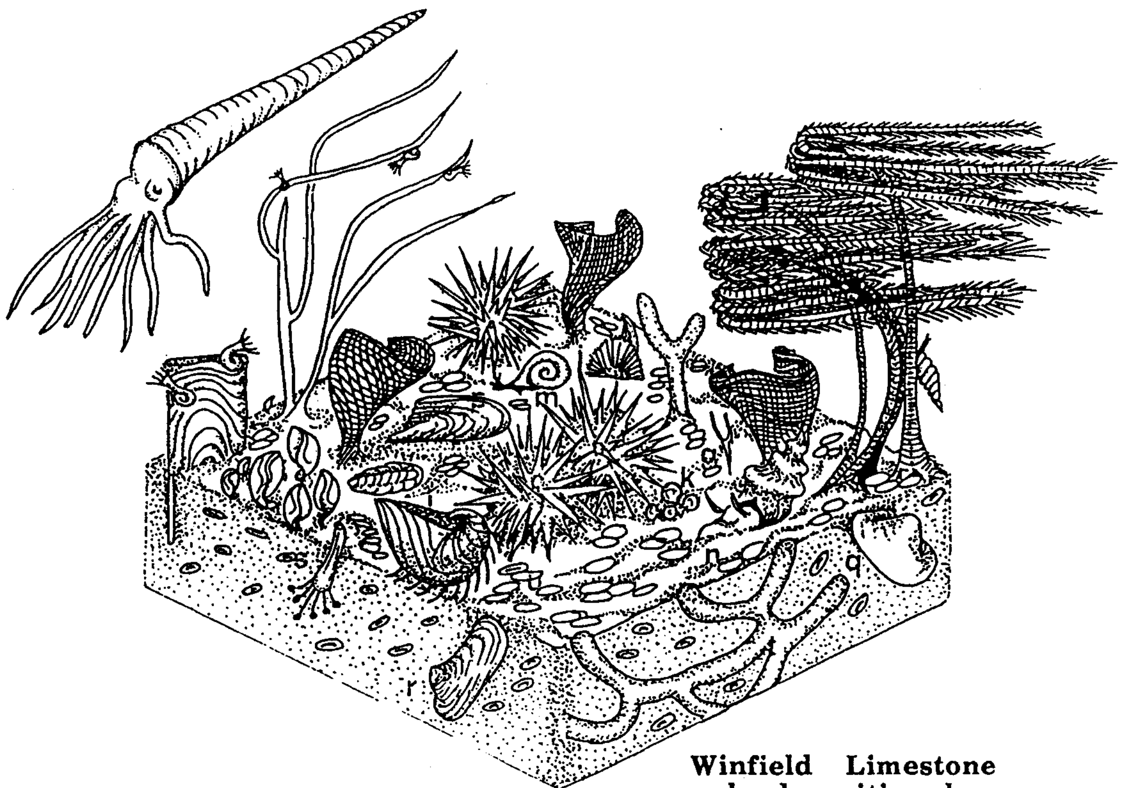

Computed Gamma Ray
0
API units
150
0
Thorium ppm
40
-10
Uranium ppm
30
0
Potassium %
10
Depth (feet)
100
200
300
400
500
600
700
800
900
1000
FORT HAYS LS
BLUE HILL SHALE
FAIRPORT CHALK
GREENHORN LIMESTONE
GRANEROS
DAKOTA
KIOWA
CHEYENNE
CEDAR HILLS SANDSTONE
Th
U
K
DAKOTA

Blank

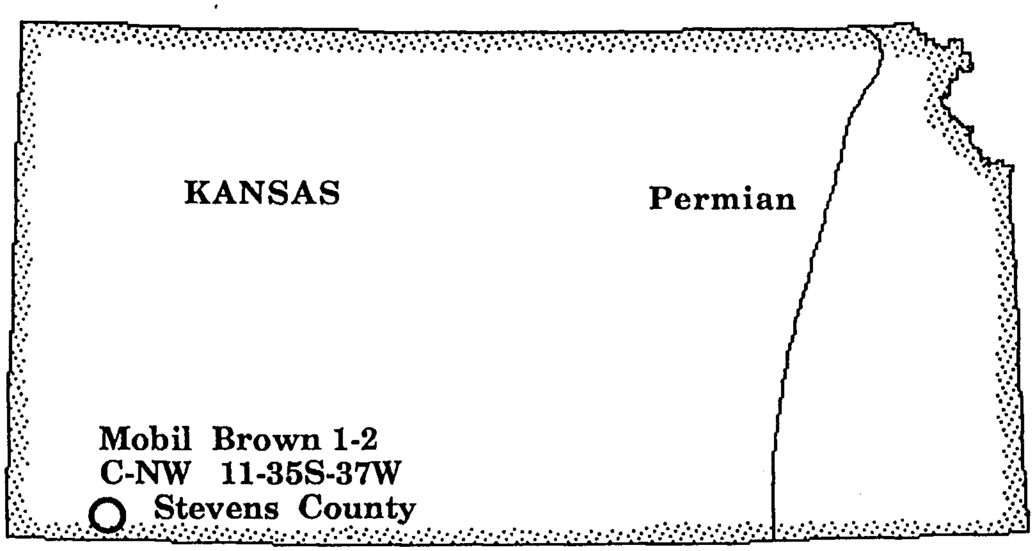
CHASE

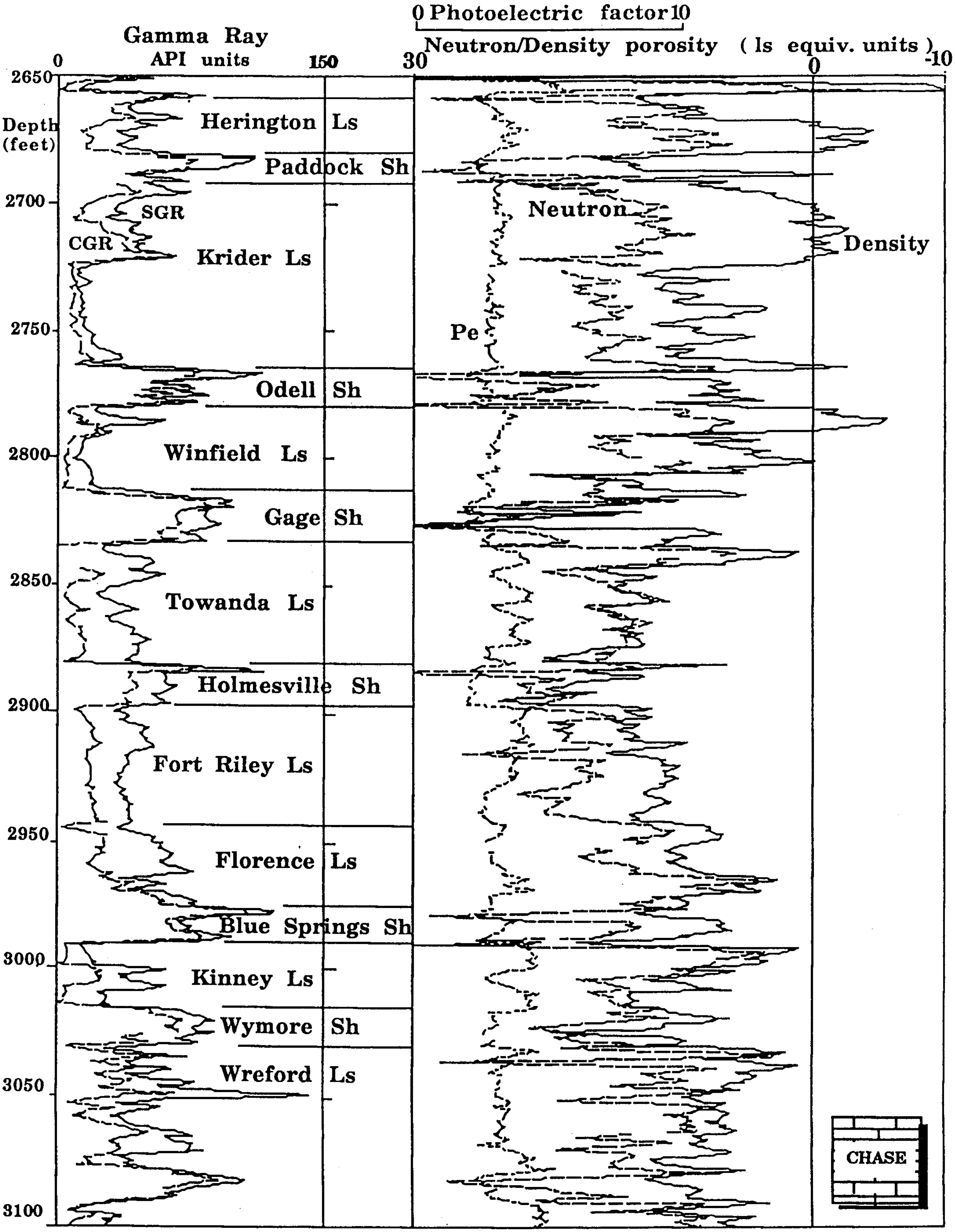
CARBONATE CASE STUDY

Lower Permian dolomites, cherty limestones, anhydrite and shales

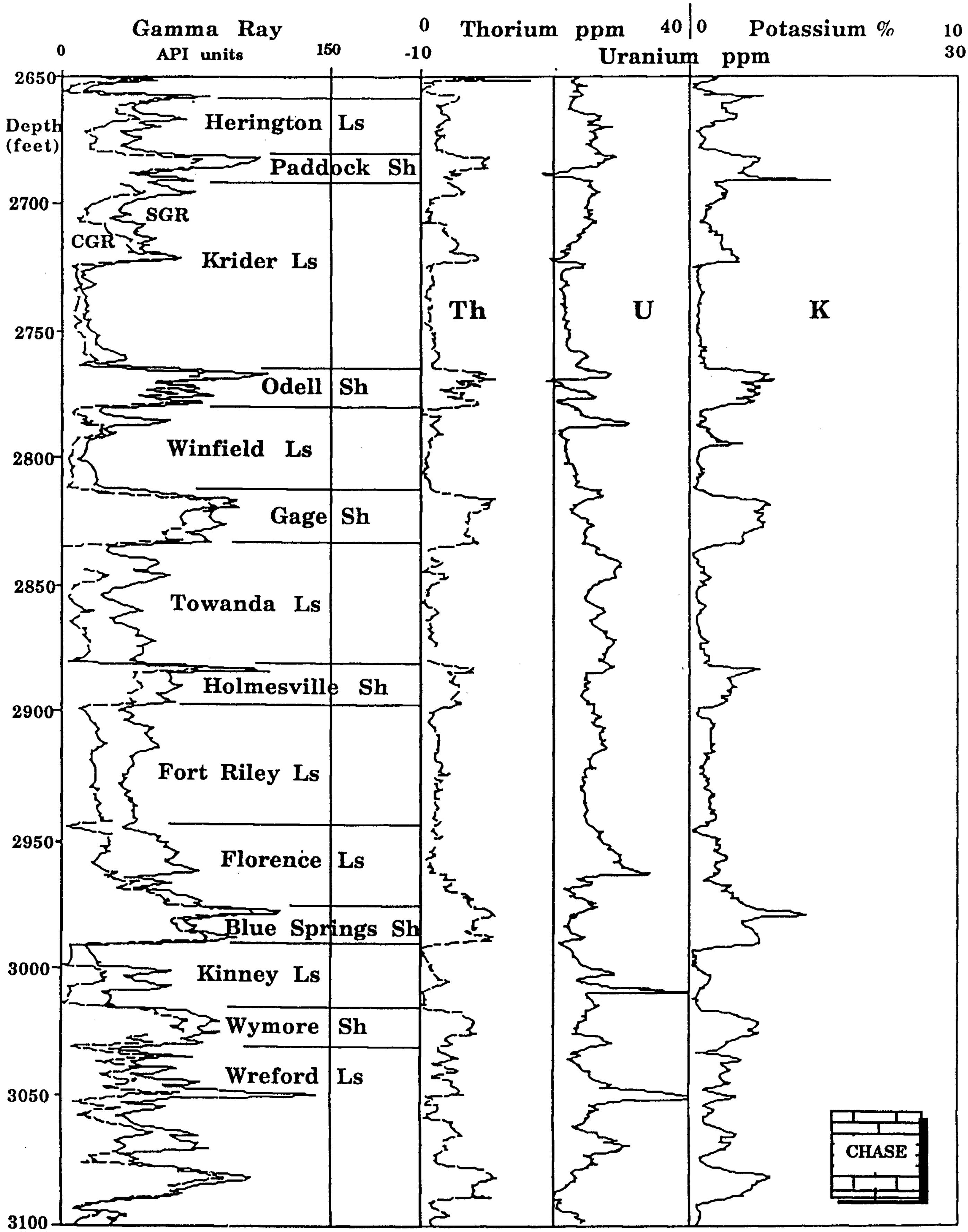


Winfield Limestone paleodepositional community (from Toomey & Mitchell, 1986)



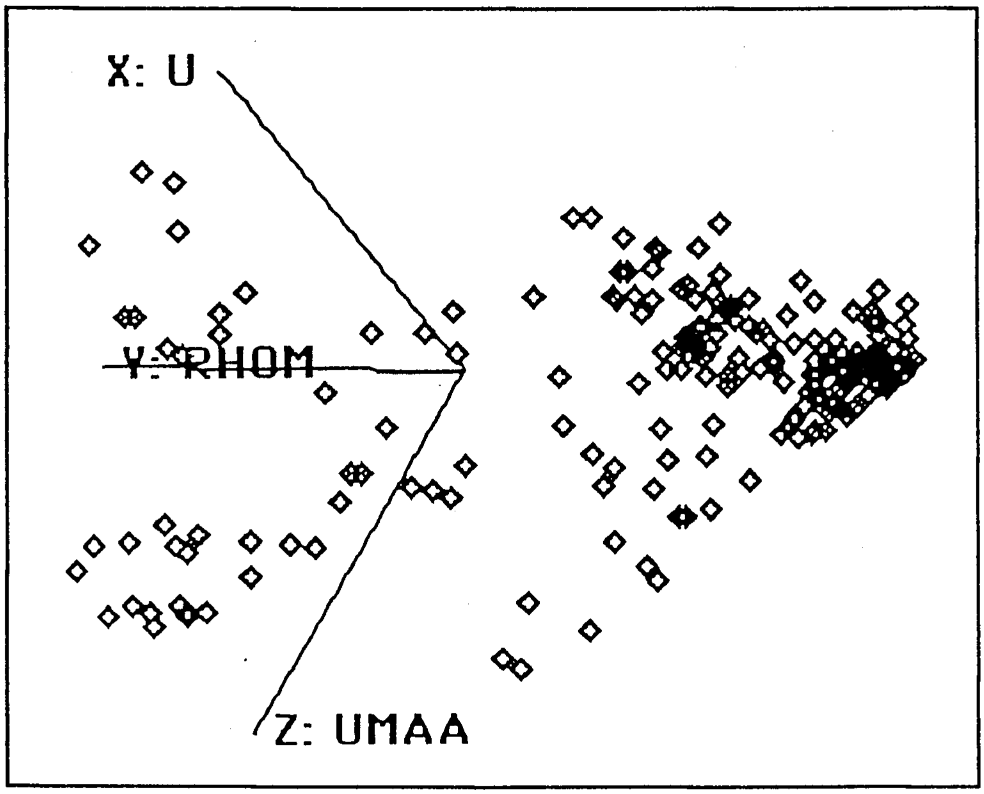


0 Photoelectric factor 10
Gamma Ray
API units
0
150
30
Neutron/Density porosity (ls equiv. units)
0
-10
Depth (feet)
2650
2700
2750
2800
2850
2900
2950
3000
3050
3100
Herington Ls
Paddock Sh
Krider Ls
SGR
CGR
Odell Sh
Winfield Ls
Gage Sh
Towanda Ls
Holmesville Sh
Fort Riley Ls
Florence Ls
Blue Springs Sh
Kinney Ls
Wymore Sh
Wreford Ls
Neutron
Density
Pe
CHASE

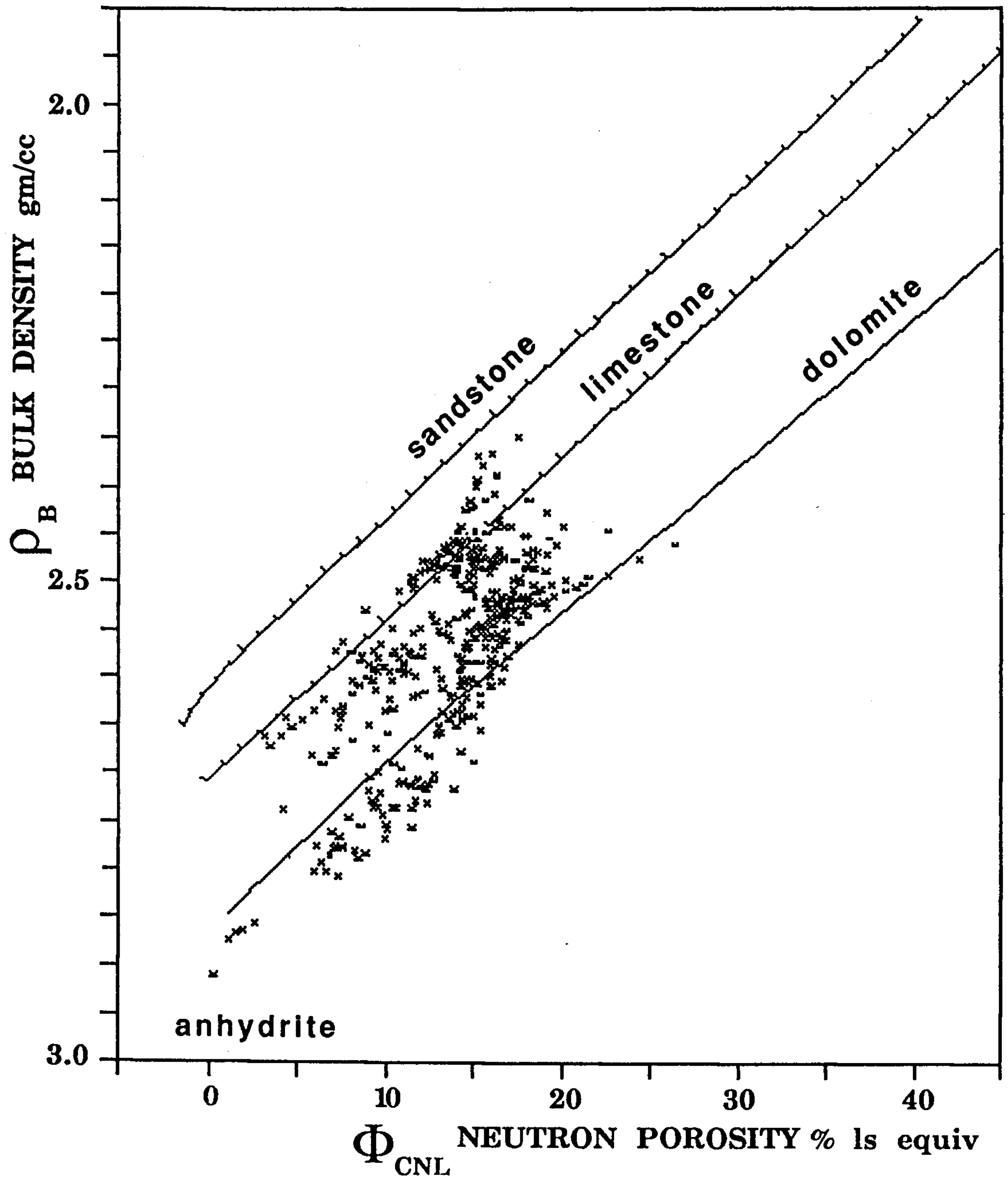

Gamma Ray
0
API units
150
0
Thorium ppm
40
-10
Uranium ppm
30
0
Potassium %
10
Depth (feet)
2650
2700
2750
2800
2850
2900
2950
3000
3050
3100
Herington Ls
Paddock Sh
SGR
CGR
Krider Ls
Th
U
K
Odell Sh
Winfield Ls
Gage Sh
Towanda Ls
Holmesville Sh
Fort Riley Ls
Florence Ls
Blue Springs Sh
Kinney Ls
Wymore Sh
Wreford Ls
CHASE

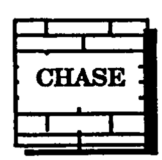
Blank

PATTERN RECOGNITION TECHNIQUES

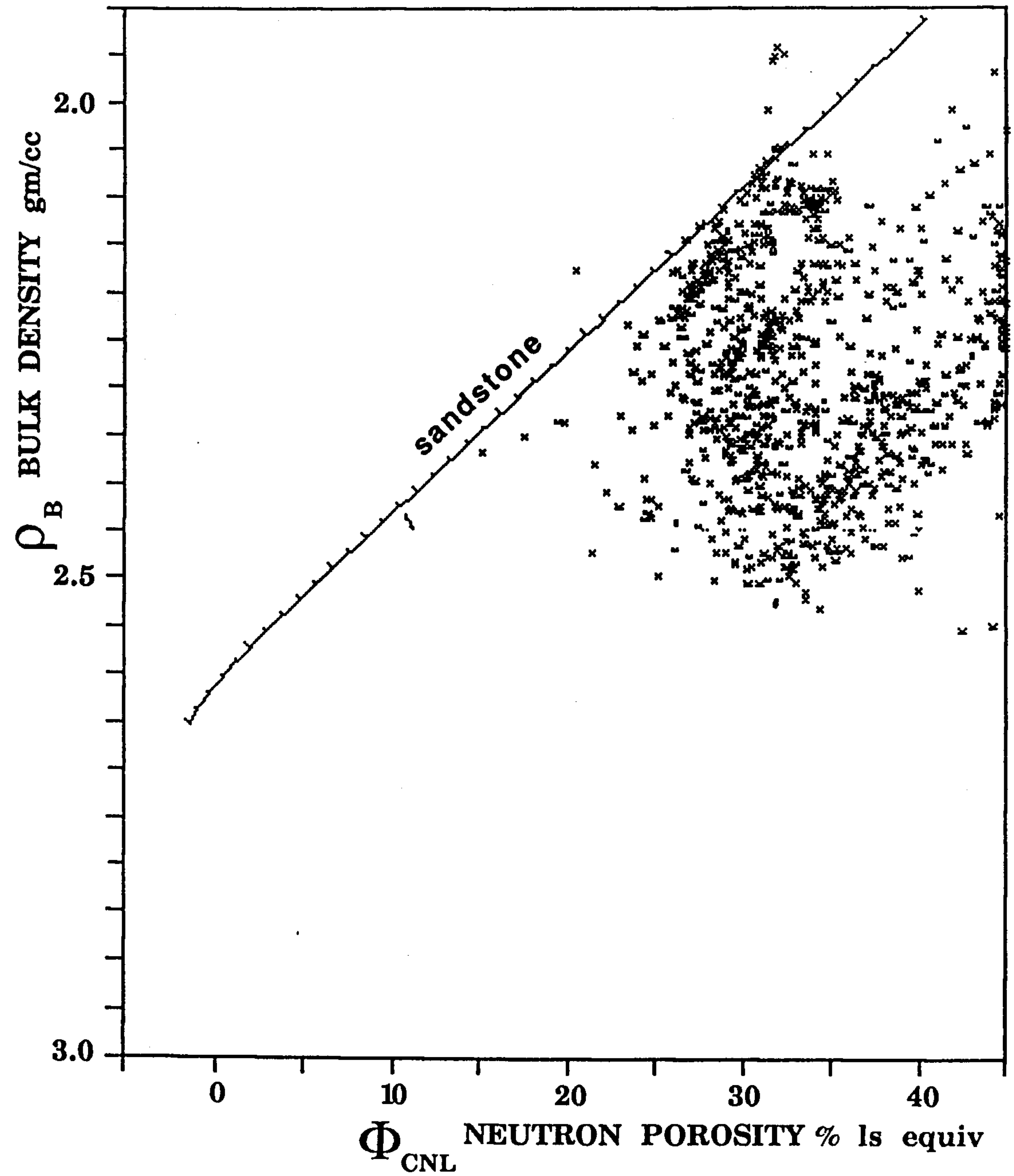


COMPLEX CARBONATE NEUTRON - DENSITY CROSSPLOT
Zones with gamma ray values < 20 API units
Mineral types : calcite, dolomite, chert, anhydrite





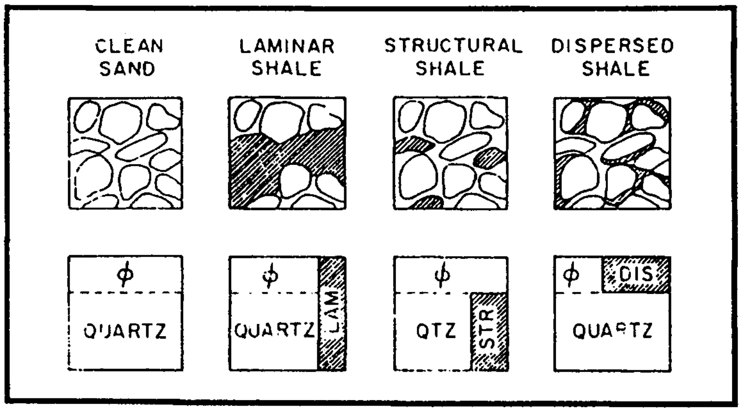
SANDSTONE - SHALE NEUTRON - DENSITY CROSSPLOT



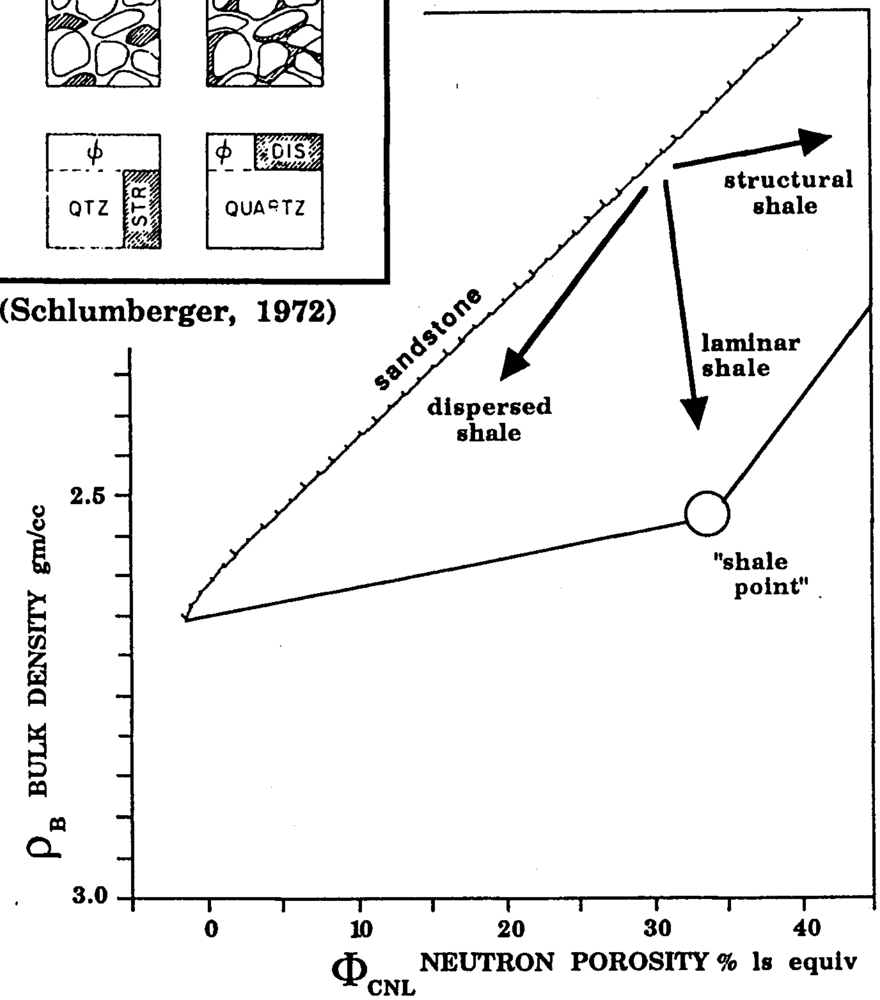
DATA RANGE:
475 - 925 feet depth

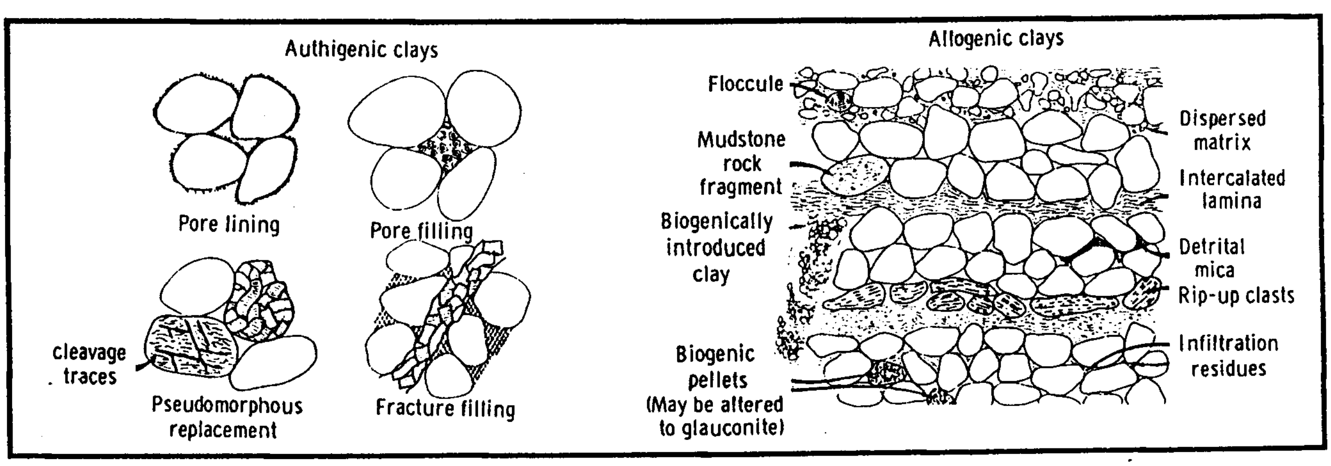
DAKOTA

CLASSIC LOG ANALYSIS MODEL OF SHALE AND CLAY DISTRIBUTION IN SHALY SANDSTONES



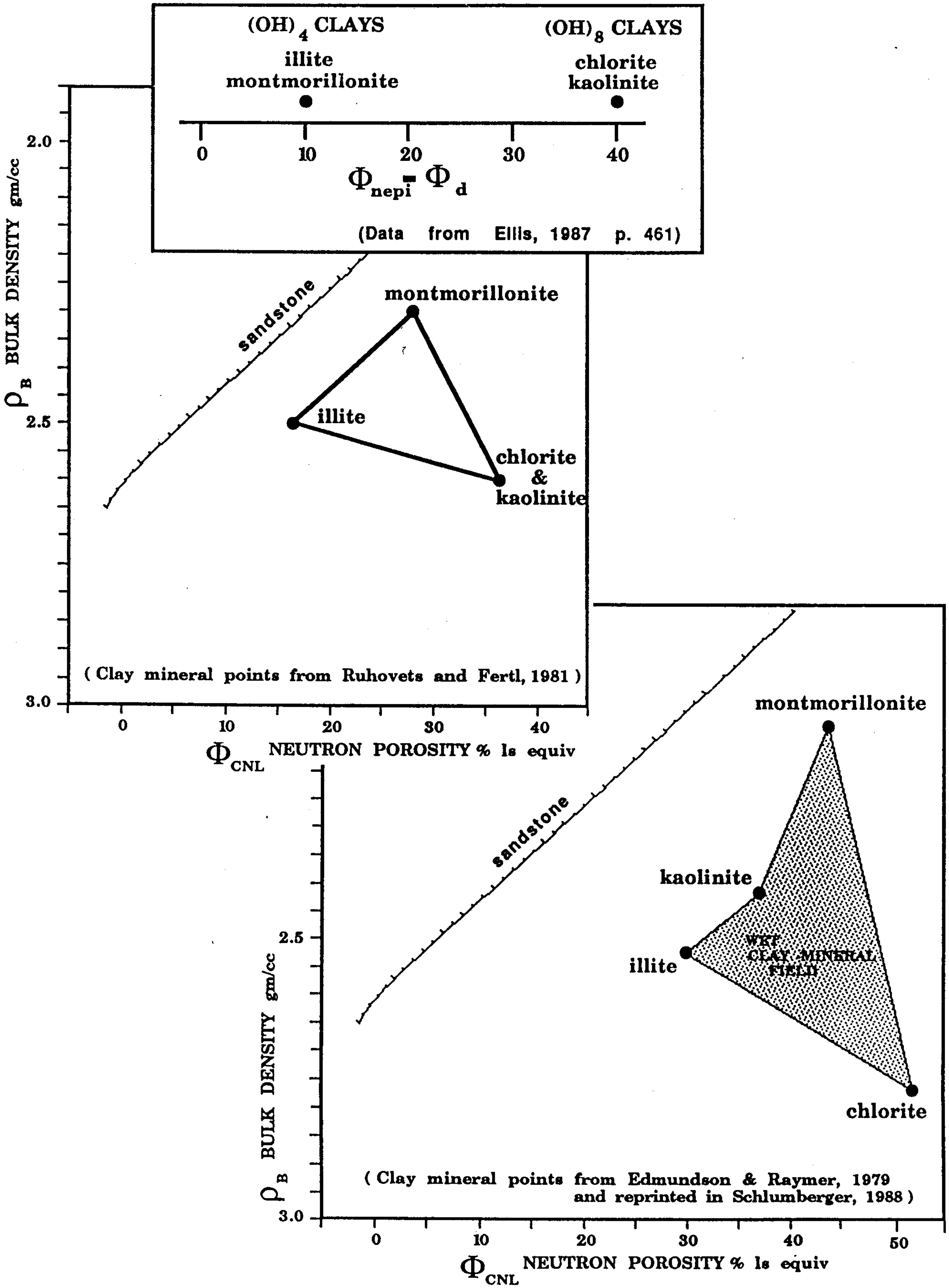
MORPHOLOGY (Schlumberger, 1972)





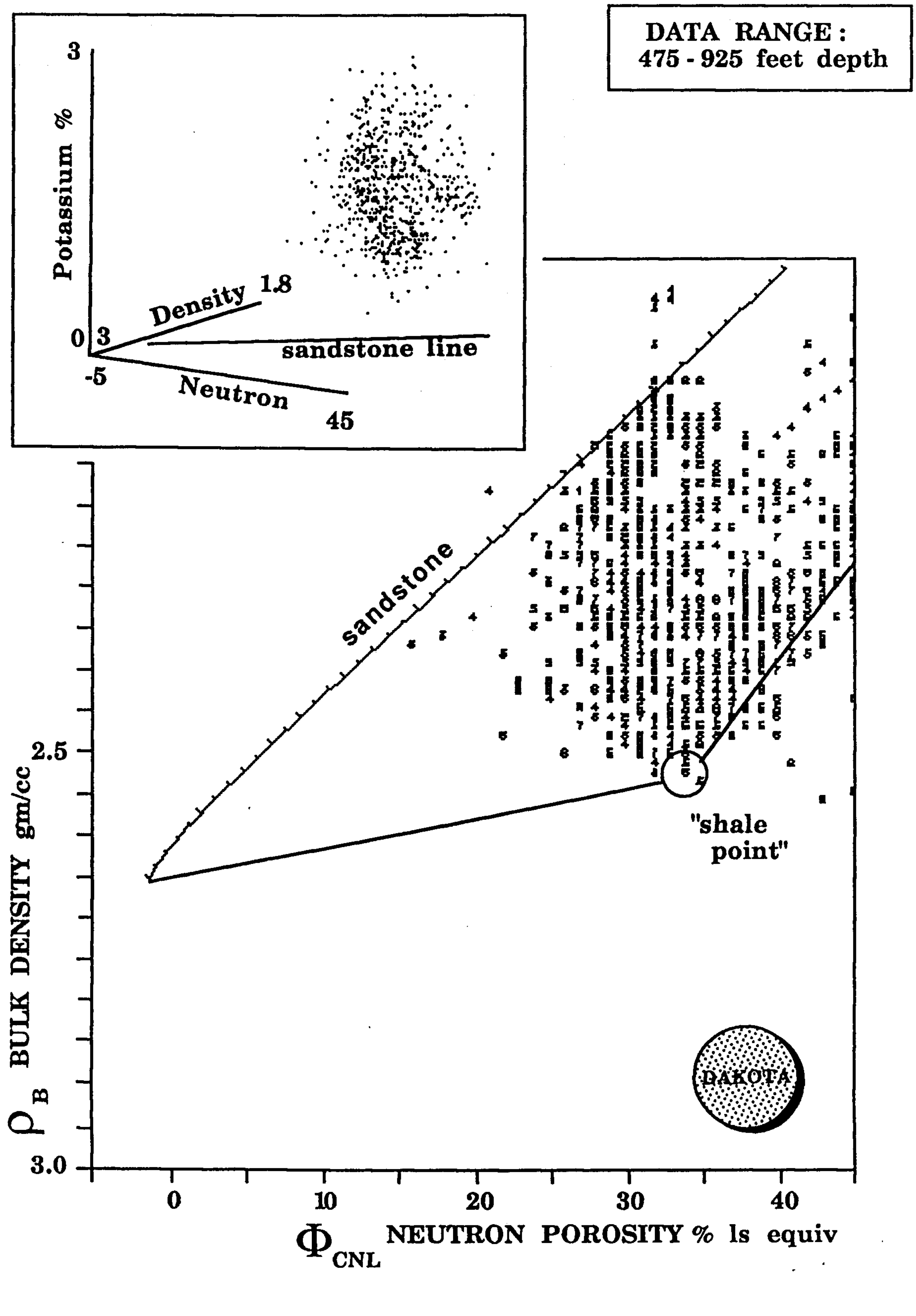
GENESIS (Wilson and Pittman, 1977)

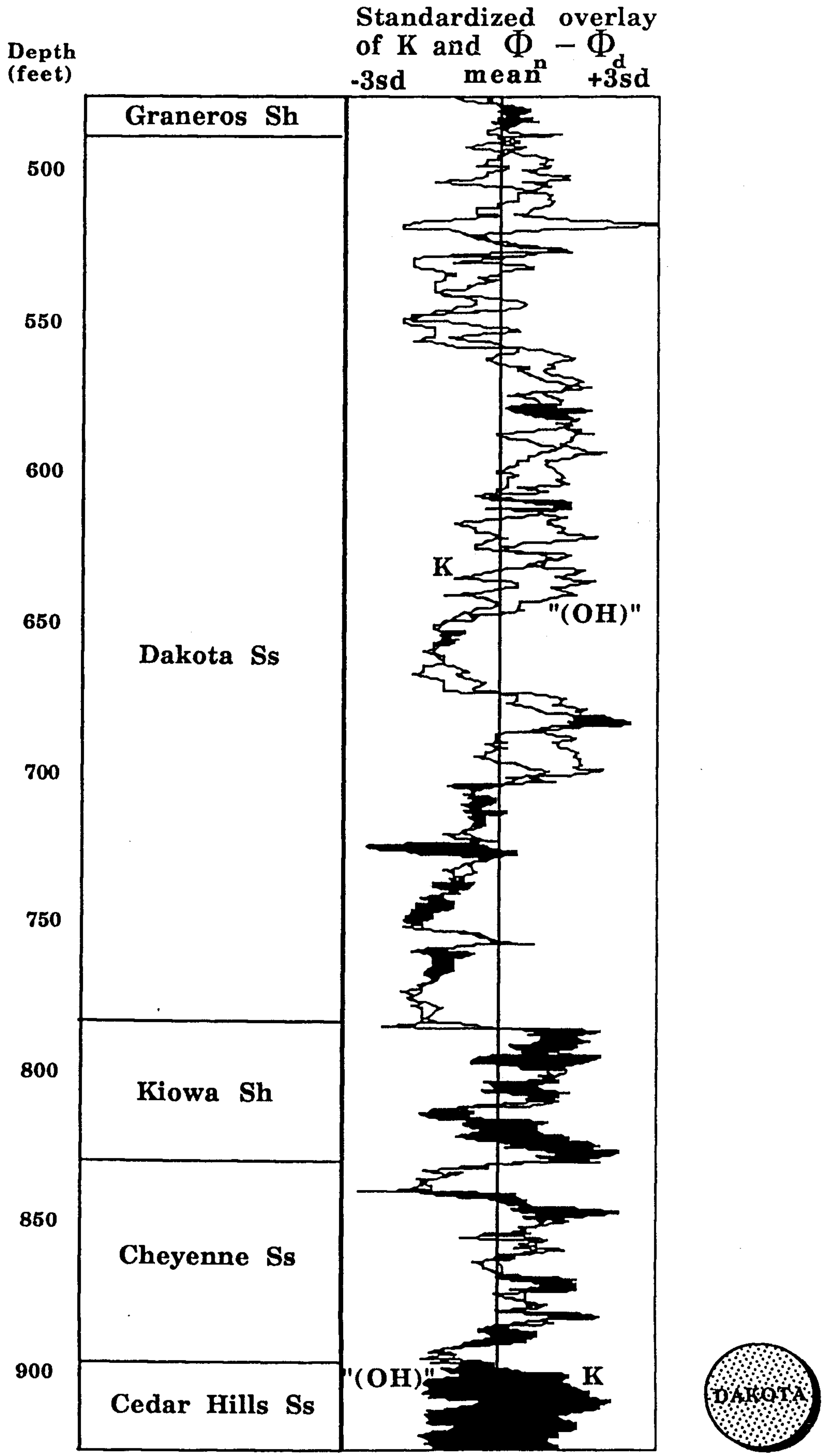
MODERN SHALY SANDSTONE MODELS BASED ON CLAY MINERAL COMPOSITIONS



SANDSTONE - SHALE NEUTRON - DENSITY Z- PLOT

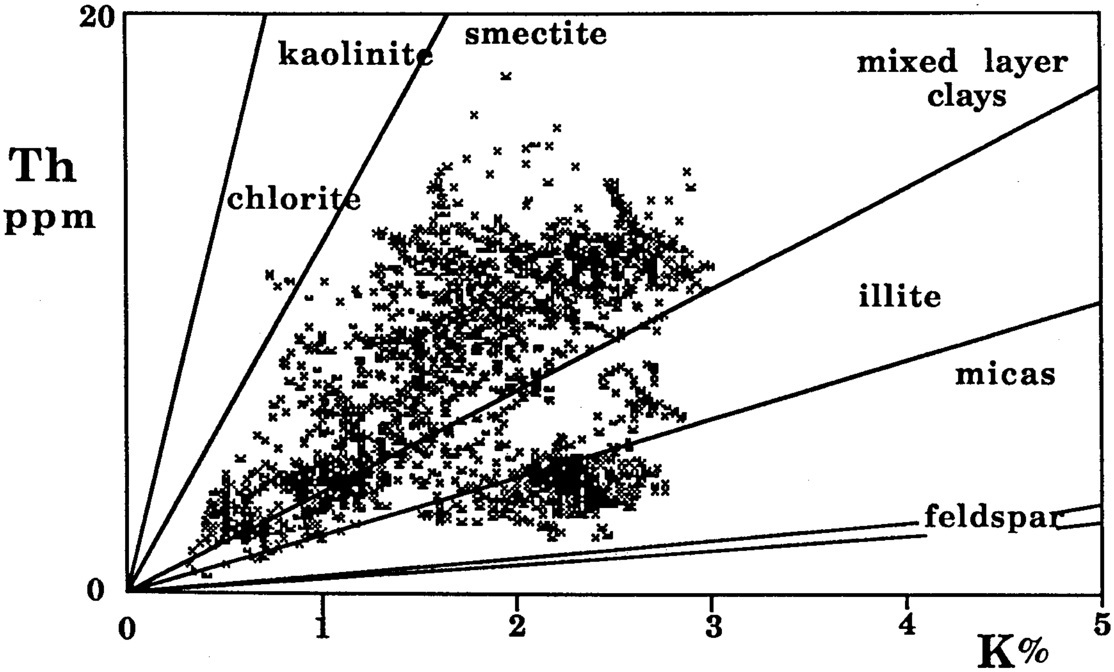
Z - variable = Potassium %
Integer range 0 - 9 for 0 - 3% K

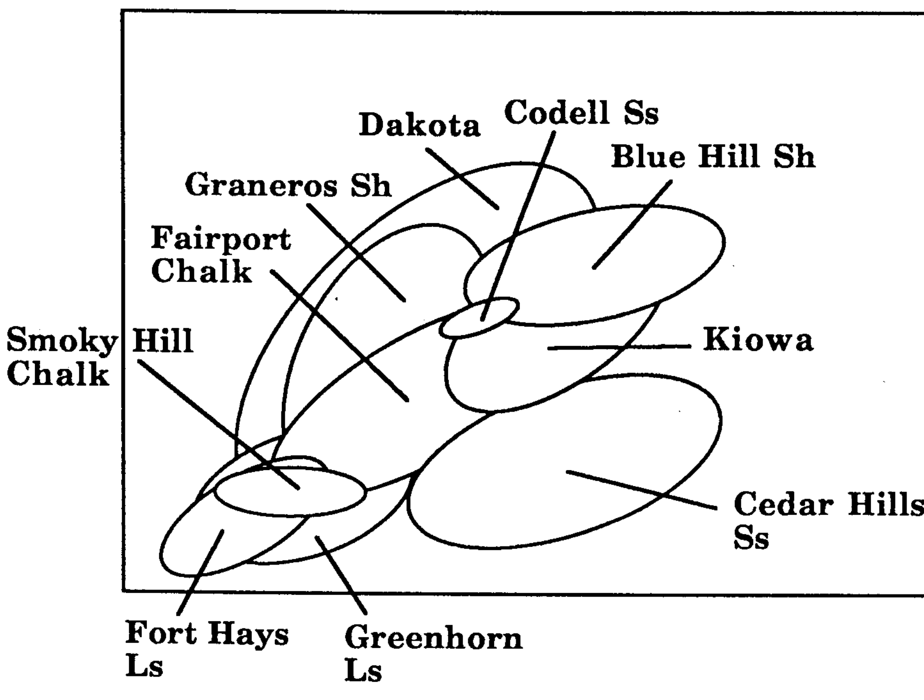




Standardized overlay of K and $\Phi_n - \Phi_d$
Depth (feet)
-3sd
mean
+3sd
Graneros Sh
500
550
600
650
700
750
800
850
900
Dakota Ss
Kiowa Sh
Cheyenne Ss
Cedar Hills Ss
K
"(OH)"
"(OH)"
K
DAKOTA

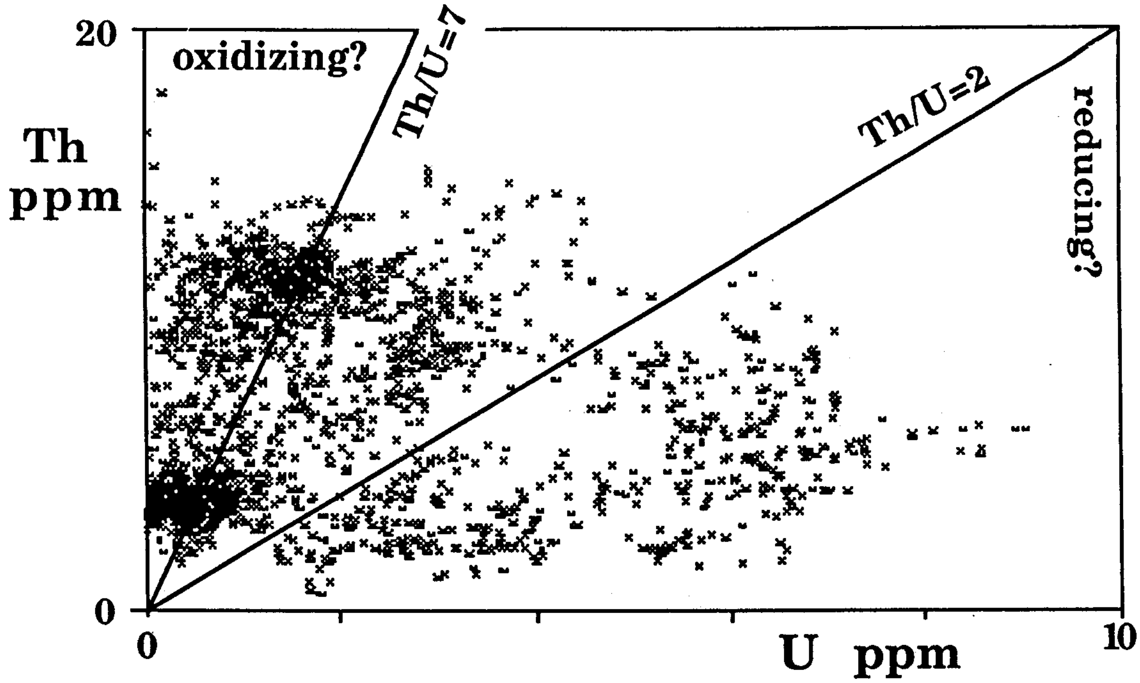
SANDSTONE - SHALE (AND SOME LIMESTONE) SPECTRAL GAMMA-RAY CROSSPLOTS

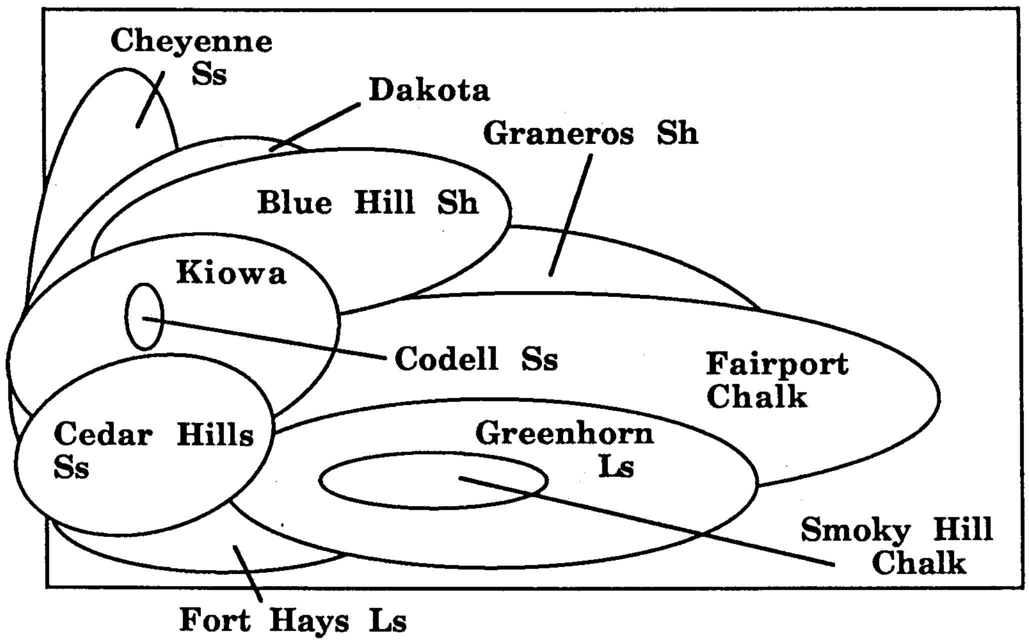


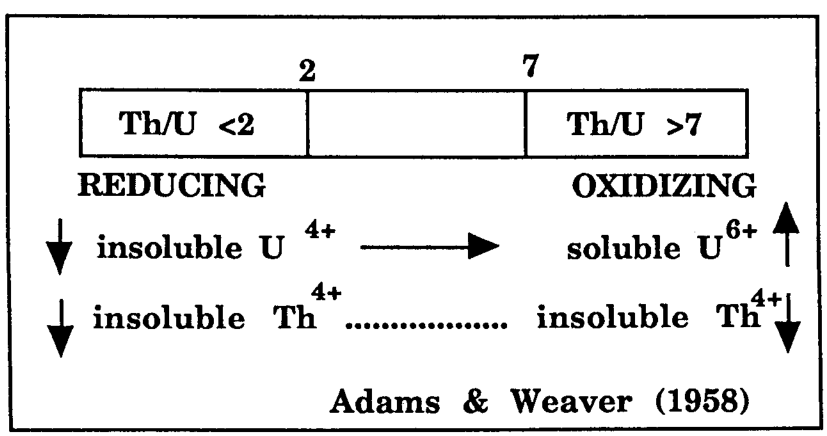


Crossplot of potassium and thorium (this page) and uranium versus thorium (facing page) in Permian and Cretaceous formations . The top figures show a composite of all data ; the lower figures graph elliptical templates for location of cloud subdivisions matched with each stratigraphic unit.

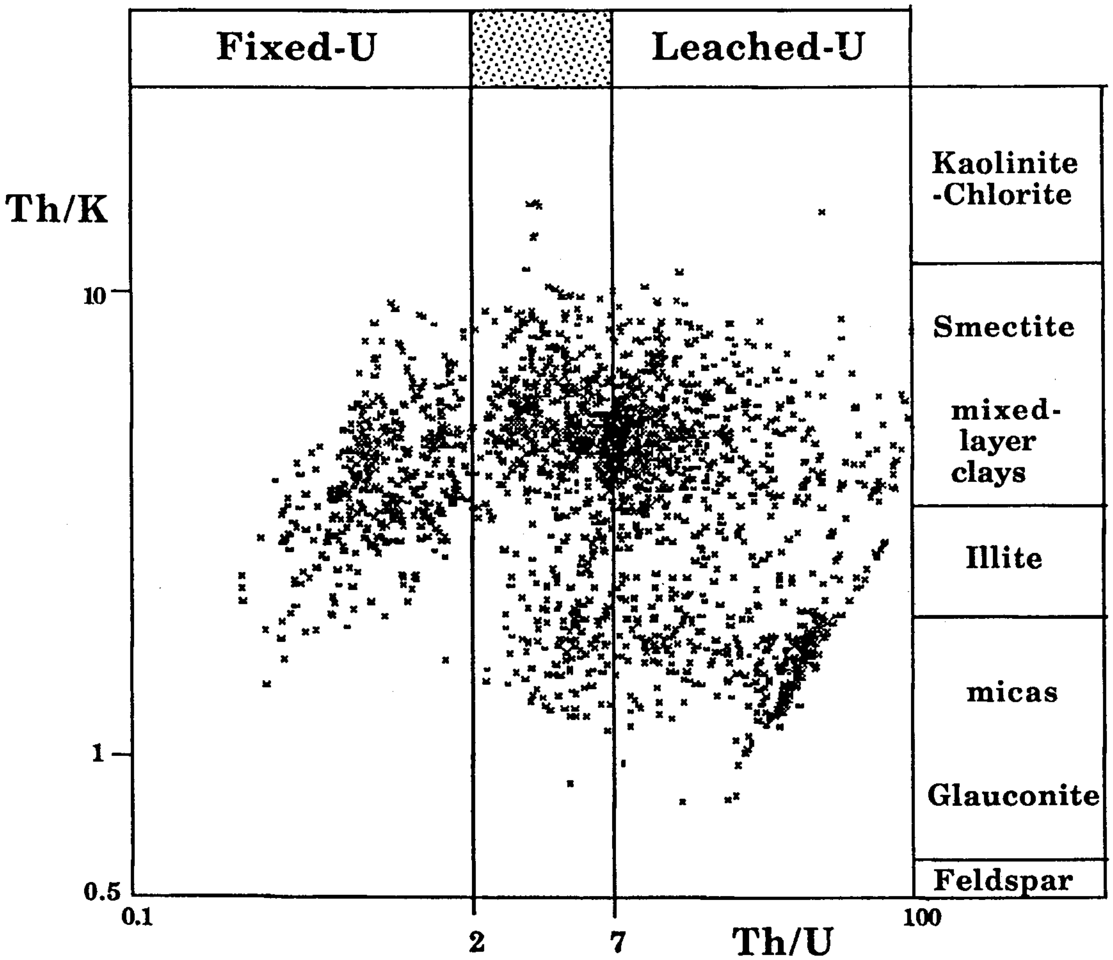
DAKOTA

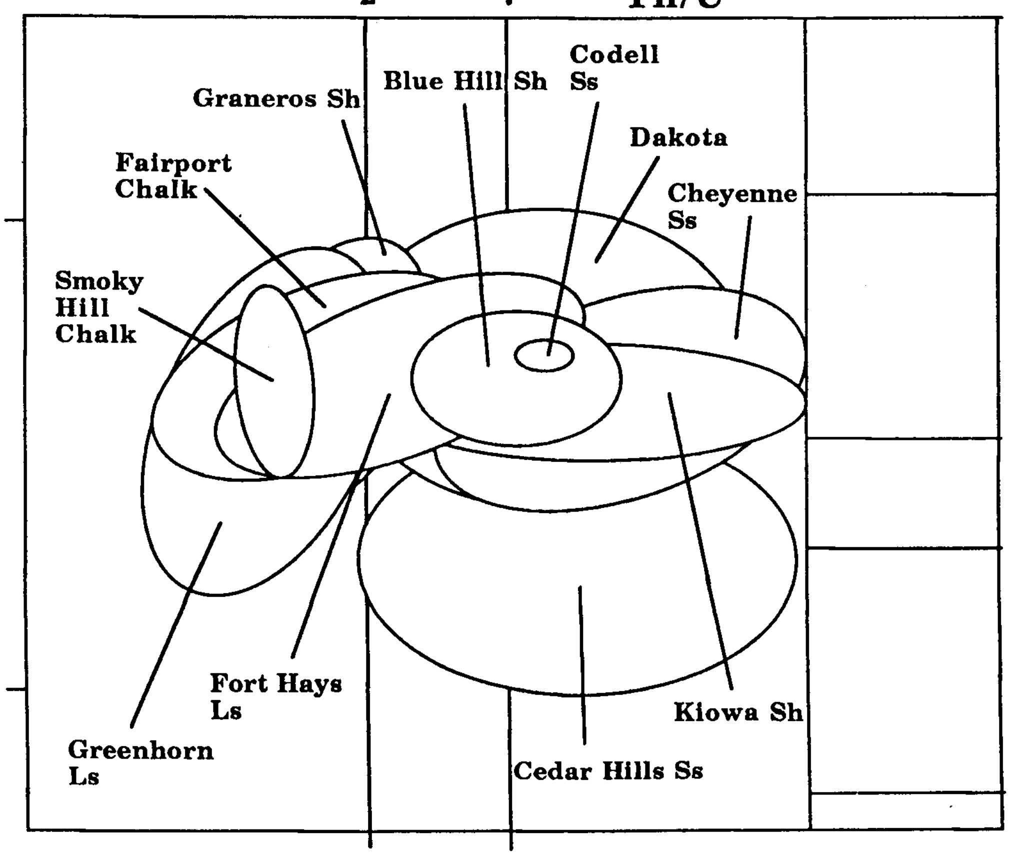

20
0
Th
ppm
0
U ppm
10
oxidizing?
Th/U=7
Th/U=2
reducing?

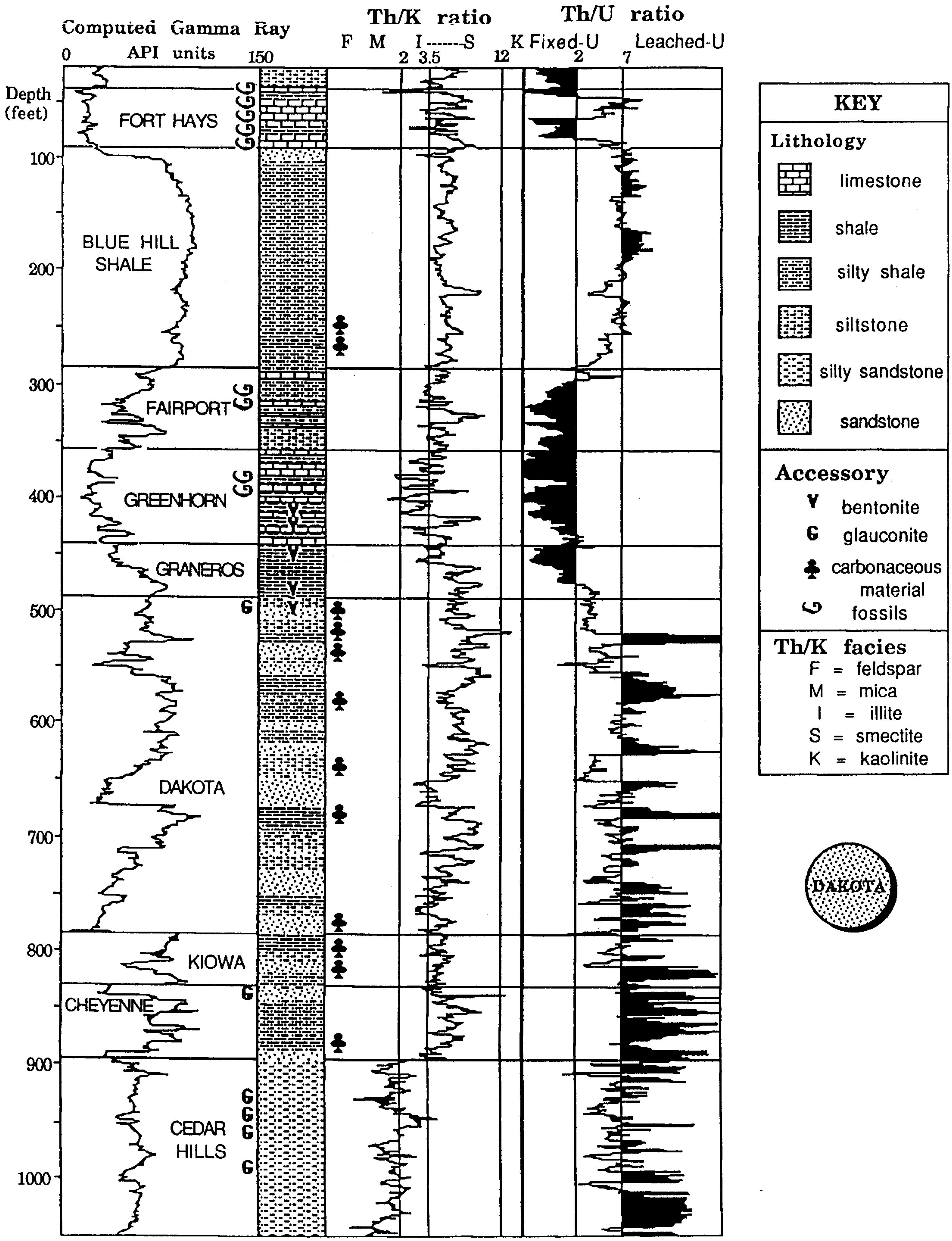

Cheyenne
Ss
Dakota
Graneros Sh
Blue Hill Sh
Kiowa
Codell Ss
Fairport
Chalk
Cedar Hills
Ss
Greenhorn
Ls
Smoky Hill
Chalk
Fort Hays Ls


2
7
Th/U <2
Th/U >7
REDUCING
OXIDIZING
insoluble U 4+
soluble U 6+
insoluble Th 4+
insoluble Th 4+
Adams & Weaver (1958)

Th/K versus Th/U Crossplot

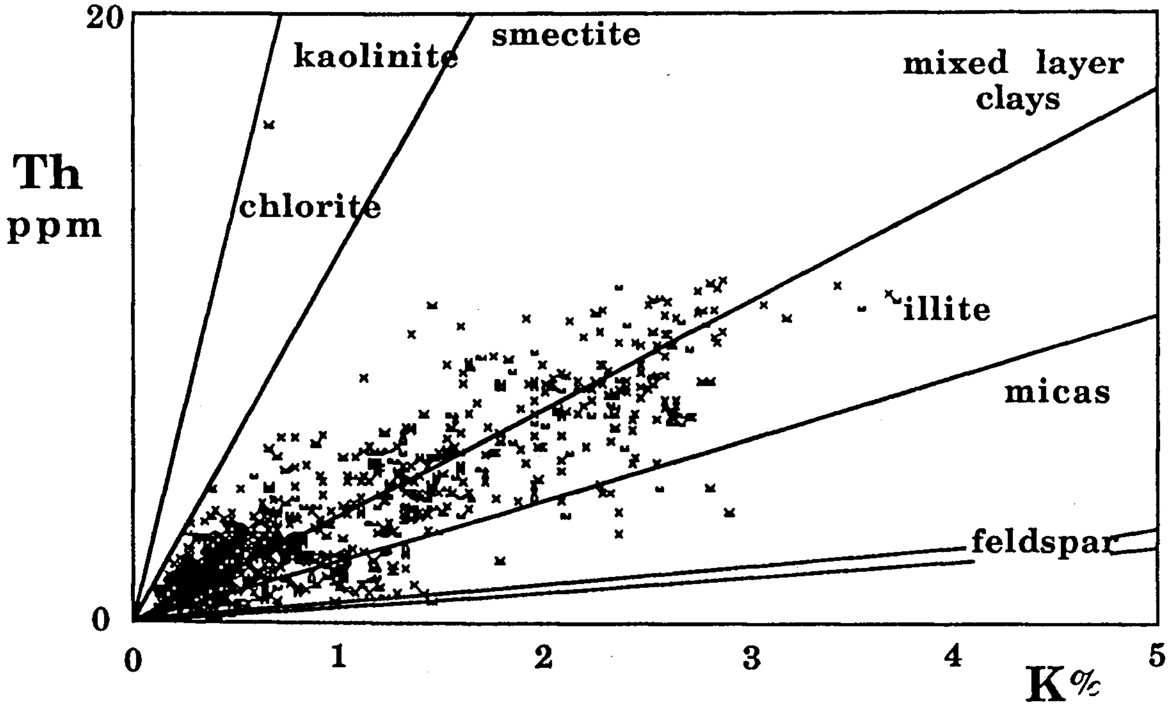




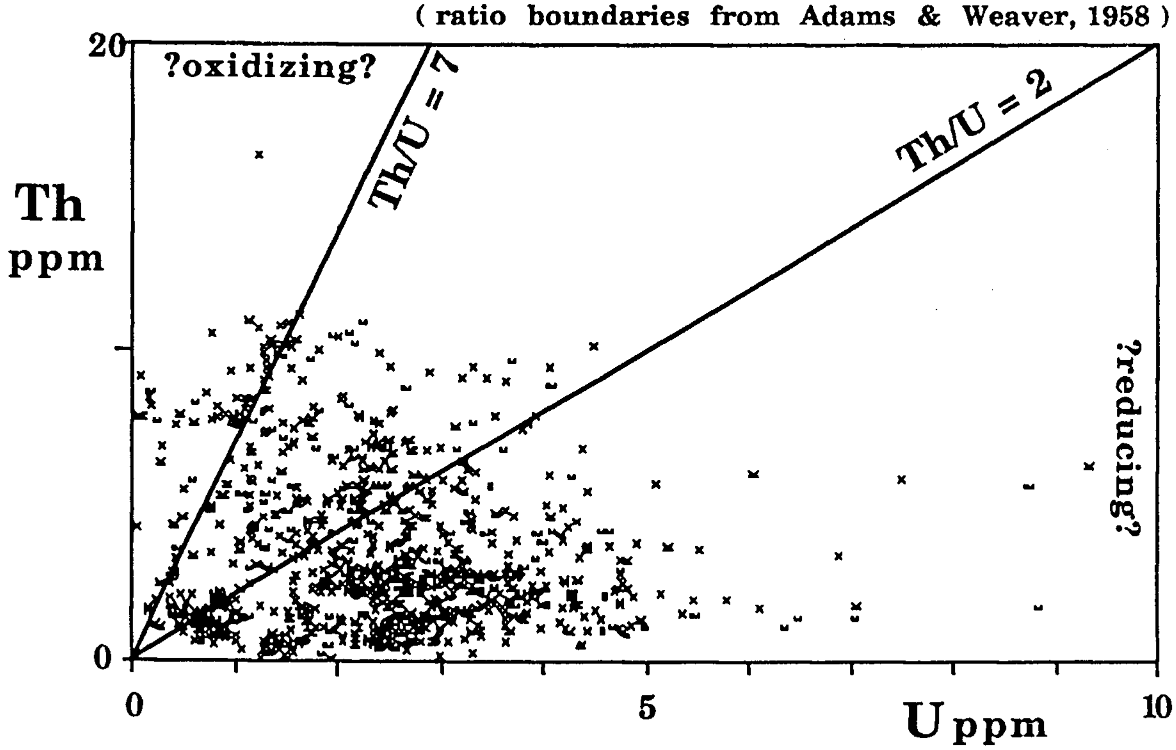


Computed Gamma Ray
0 API units 150
Th/K ratio
F M 2 I 3.5 S 12 K
Th/U ratio
Fixed-U 2 7 Leached-U
Depth (feet)
100
200
300
400
500
600
700
800
900
1000
FORT HAYS
BLUE HILL SHALE
FAIRPORT
GREENHORN
GRANEROS
DAKOTA
KIOWA
CHEYENNE
CEDAR HILLS
KEY
Lithology
limestone
shale
silty shale
siltstone
silty sandstone
sandstone
Accessory
bentonite
glauconite
carbonaceous material
fossils
Th/K facies
F = feldspar
M = mica
I = illite
S = smectite
K = kaolinite
DAKOTA

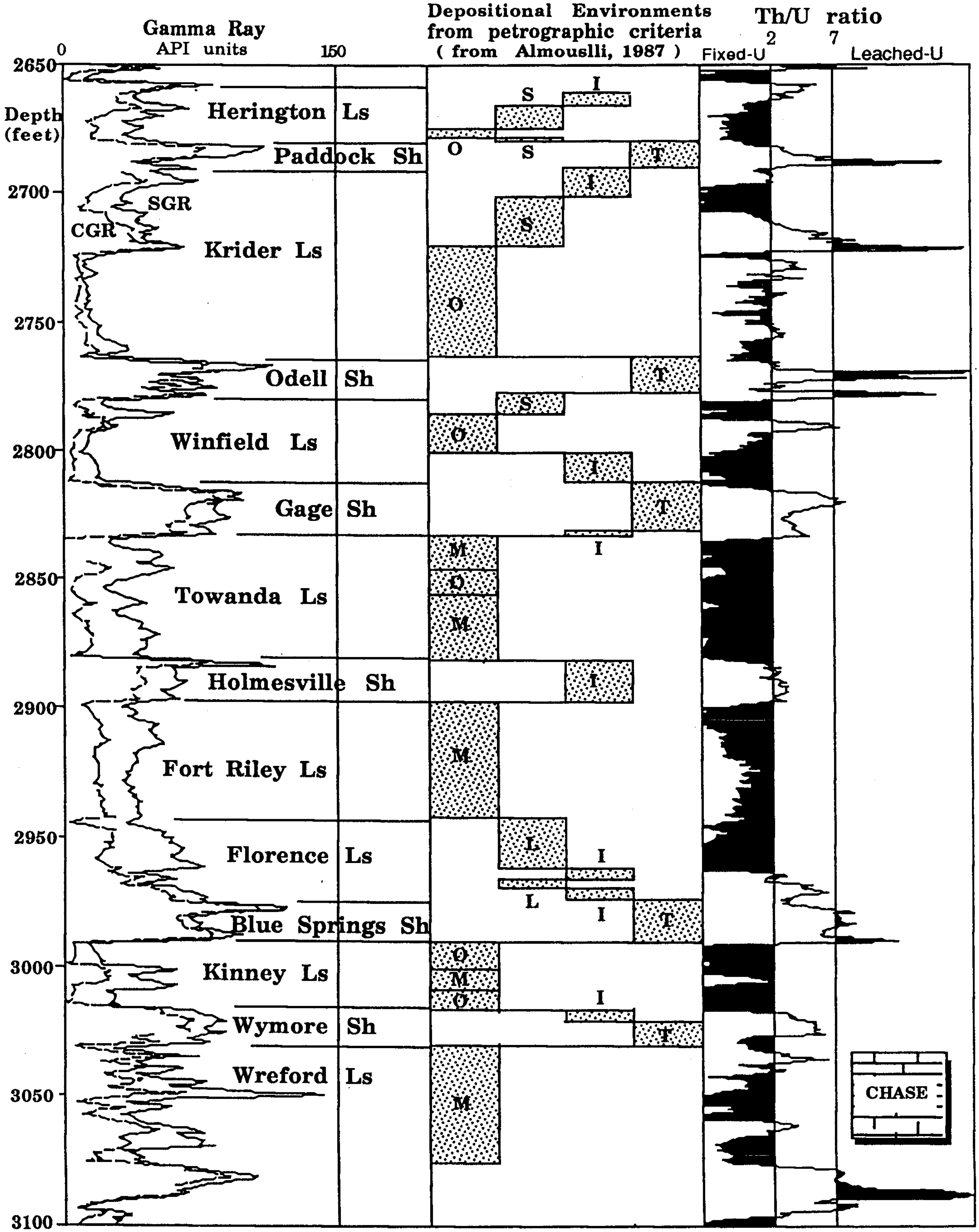
COMPLEX CARBONATE SPECTRAL GAMMA-RAY CROSSPLOTS



(ratio boundaries from Adams & Weaver, 1958)



CHASE



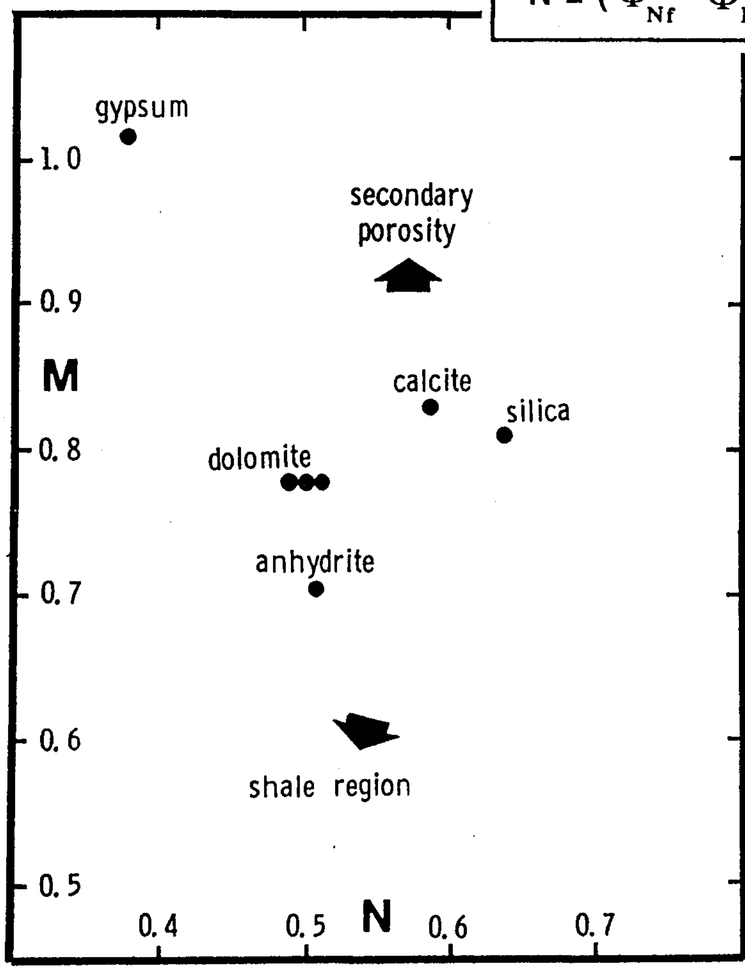
M = marine; O = off-bank; S = subtidal; L = lagoonal; I = intertidal; T = supratidal

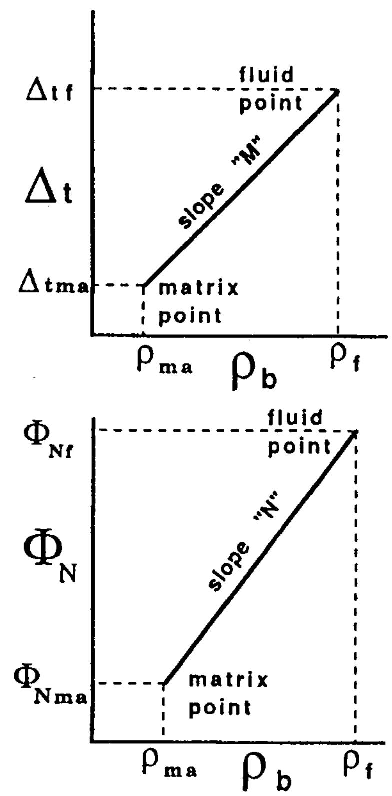
THE M - N PLOT :

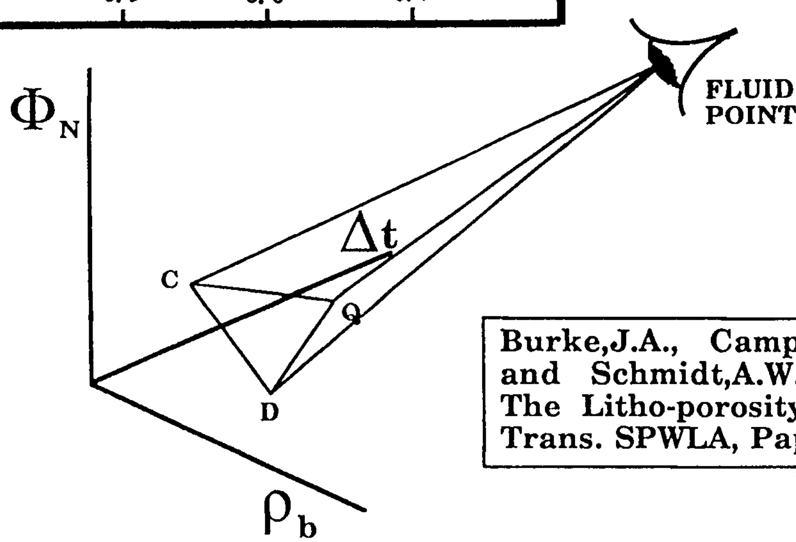
A PROJECTION IN THREE POROSITY LOG SPACE AS SLOPES ONTO A 2-DIMENSIONAL PLOT

$$M = (\Delta t_f - \Delta t)/(\rho_b - \rho_f) * 0.01$$

$$N = (\Phi_{Nf} - \Phi_N)/(\rho_b - \rho_f)$$



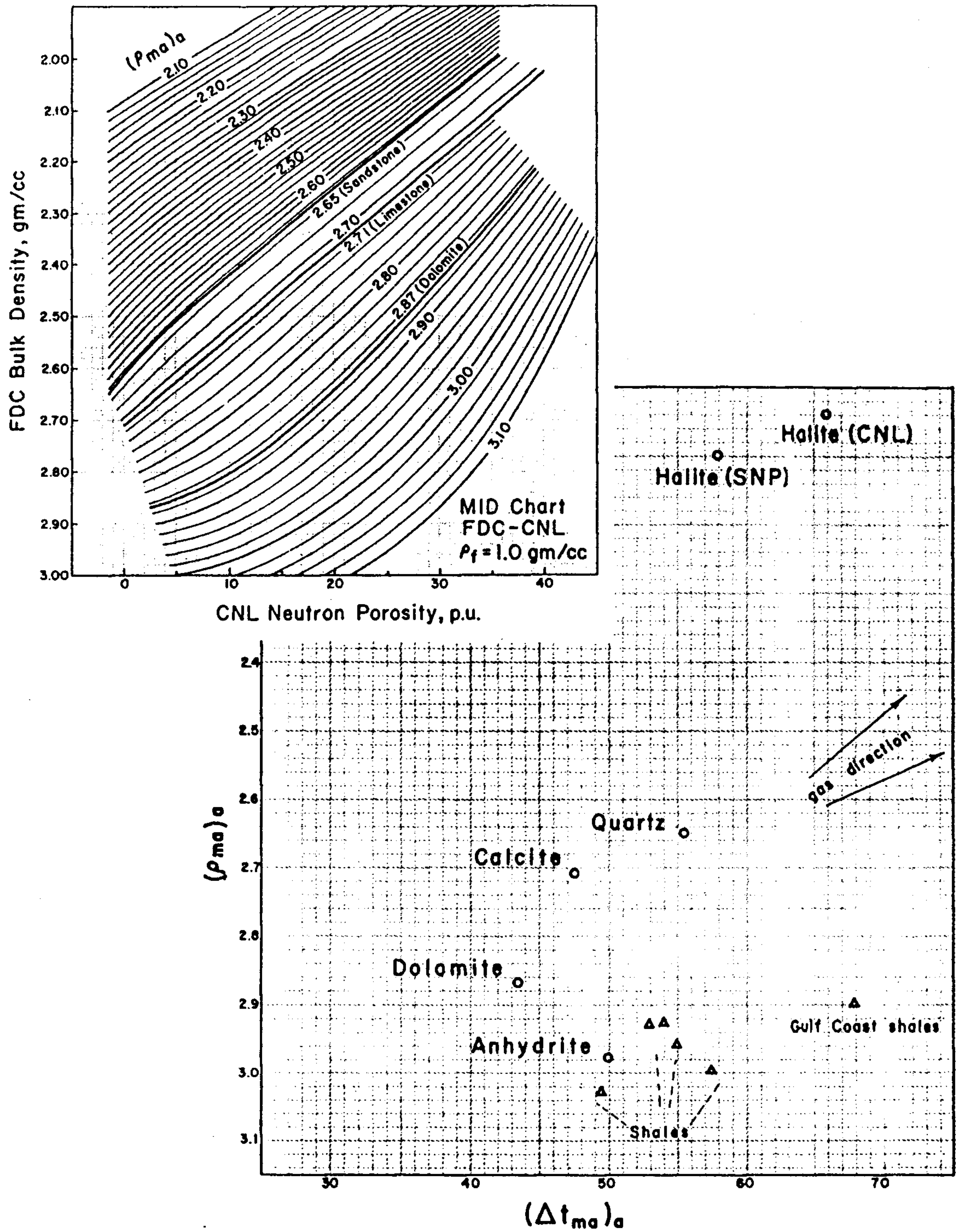




Burke,J.A., Campbell,R.L.,Jr, and Schmidt,A.W., 1969, The Litho-porosity Cross Plot: Trans. SPWLA, Paper Y, 29 pp.

THE MID PLOT :

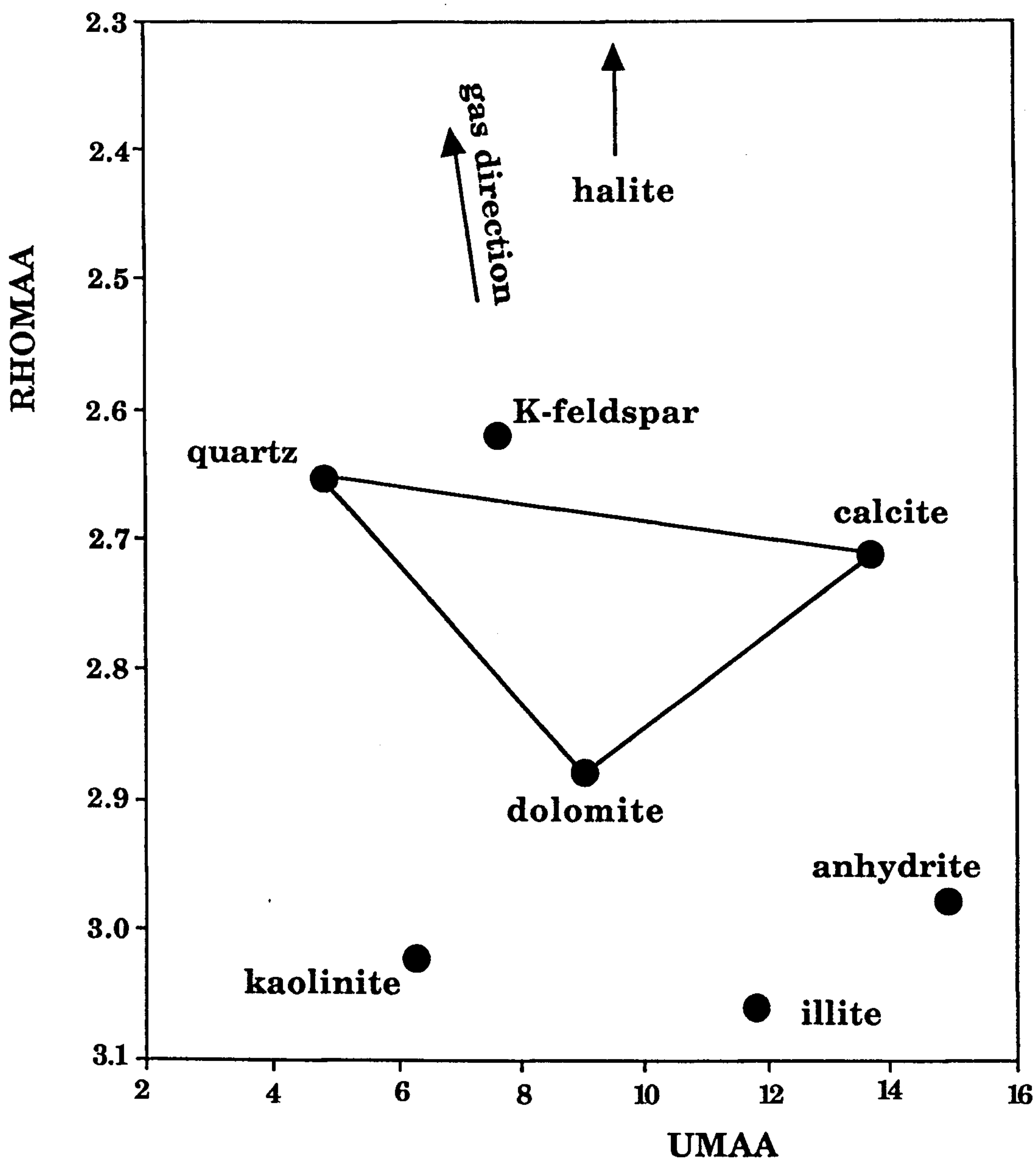
A NON-LINEAR PROJECTION IN THREE POROSITY LOG SPACE AS INTERCEPTS ON A 2-DIMENSIONAL PLOT



Clavier, C., and Rust,D.H., 1976, MID plot: A new lithology technique : The Log Analyst, v. XVII, no.6, p. 16 -24.

THE RHOMAA - UMAA PLOT :

A LITHODENSITY - NEUTRON SPECIES OF MID PLOT



Gardner, J.S. and Dumanoir, J.L., 1980, Lithodensity Log Interpretation : Trans. SPWLA, Paper N, 23 pp.

CALCULATION OF RHOMAA AND UMAA

INPUT:
- Fractional neutron porosity (PHIN)
- Fractional density porosity (PHID)
- Photoelectric factor (PE)

PHIH = (PHIN + PHID)/2

PHIH: 0.05 — 0.10 →

PHIH > 0.10:

PHIT = 1.03*PHIH

0.05 – 0.10:

D = 0.27 + 1.45*PHIN - 11.5*PHIH

PHIT = 1.03(SQRT[D^2+ 40(0.145PHIN + 0.125PHIH)] - D) /20

PHIH < 0.05:

If PHIN < PHID then PHIT = PHIH

If PHIN > PHID then : D = 0.27 + 2.21PHIN - 12.2PHIH

PHIT = 1.03(SQRT[D^2+ 40(0.221PHIN + 0.049PHIH)] - D) / 20

If PHIT < 0.01 then PHIT = 0.01

If PHIT > 0.45 then PHIT = 0.45

RHOB = 2.71 - 1.71PHID

RHOMAA = (RHOB - PHIT)/(1 - PHIT)

RHOMAA

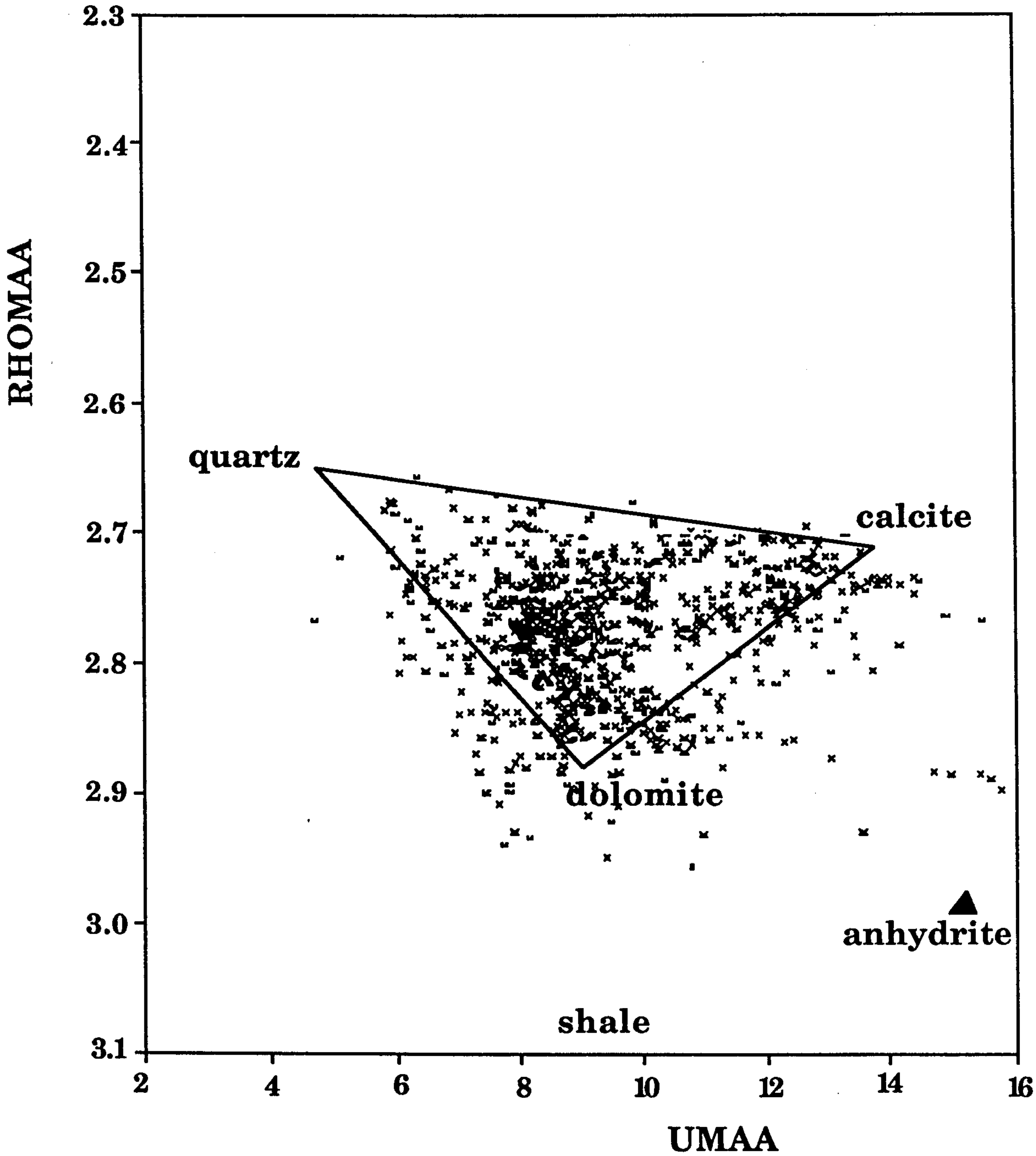
U = PE(RHOB + 0.188)/1.07

UMA = (U - 0.3PHIT)/(1 - PHIT)

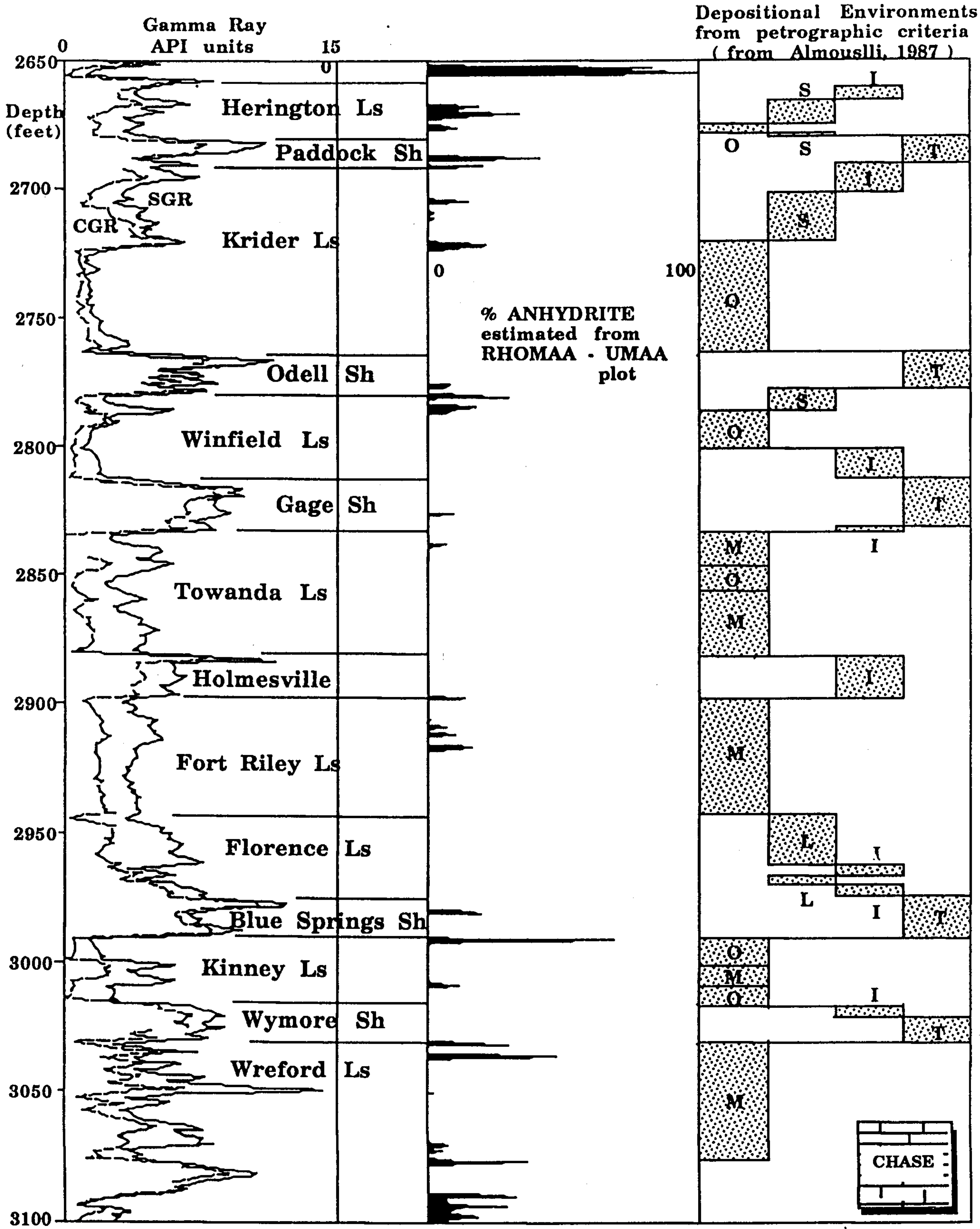
UMAA

Algorithm from BASIC program published by :
Elphick, R.Y., 1987, Nuclear log interpretation in hard rock formations : Geobyte, v. 2, no. 3, p. 44 - 47.

COMPLEX CARBONATE RHOMAA - UMAA PLOT

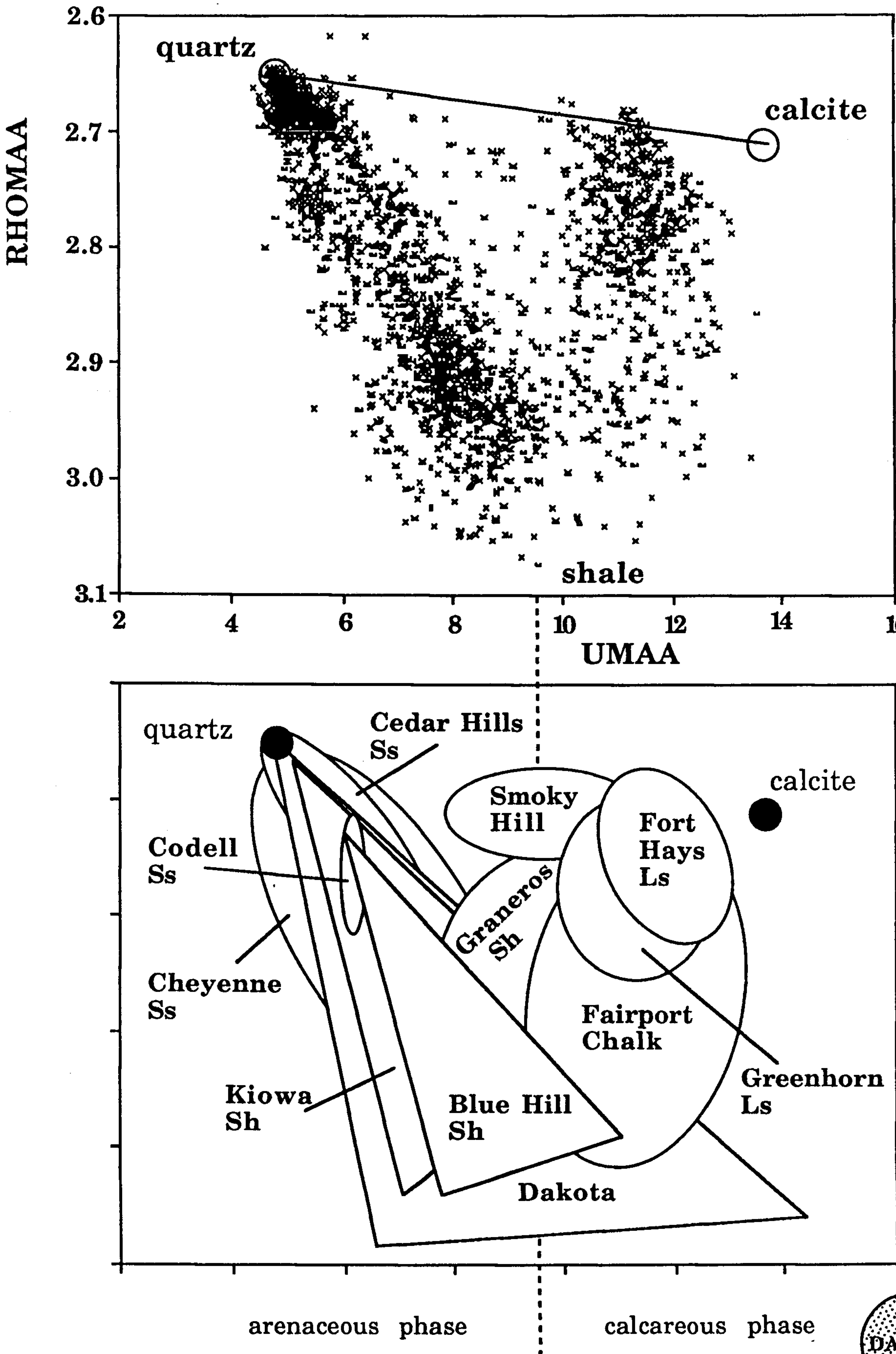


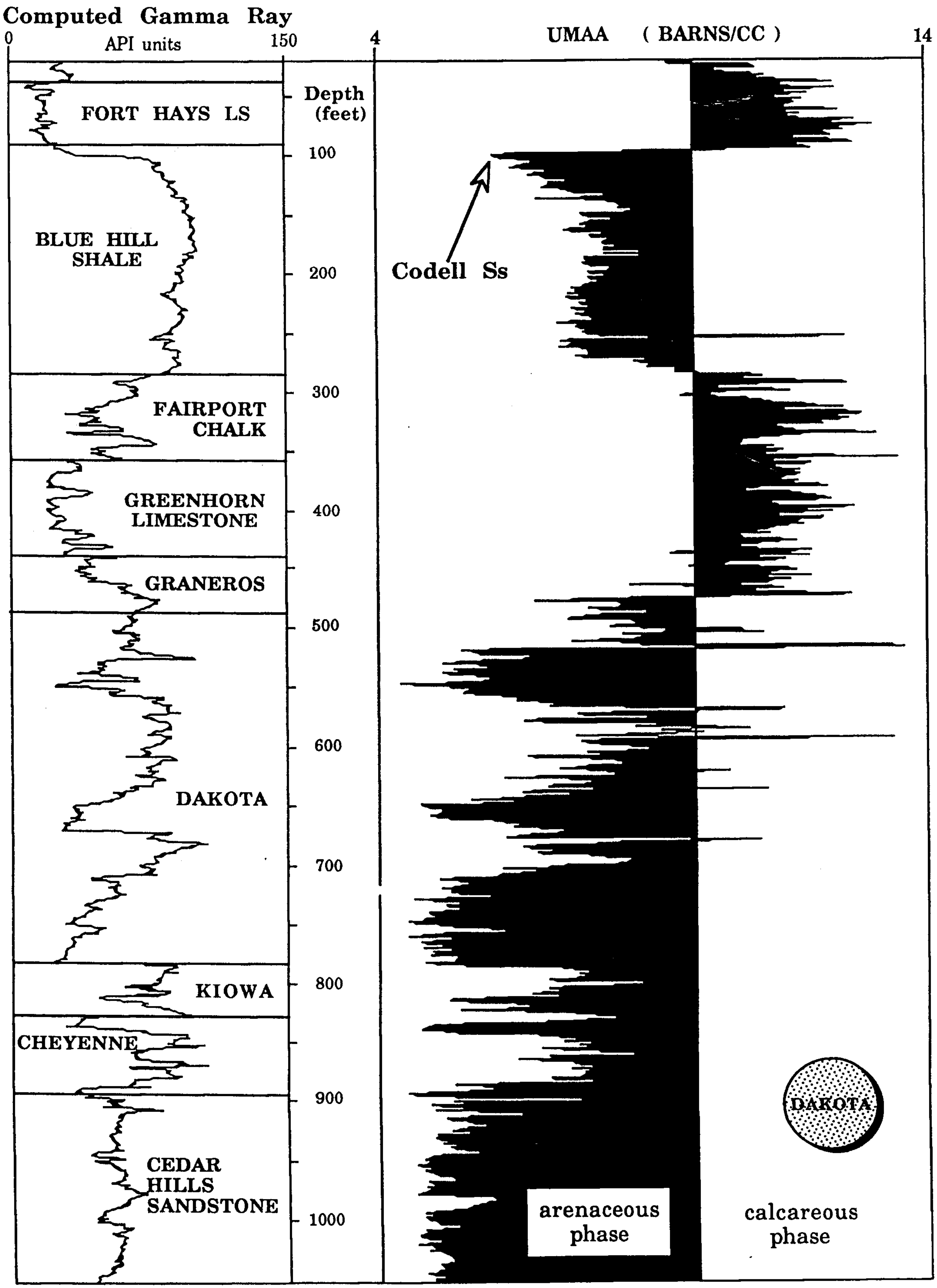
CHASE



M = marine; O = off-bank; S = subtidal; L = lagoonal; I = intertidal; T = supratidal

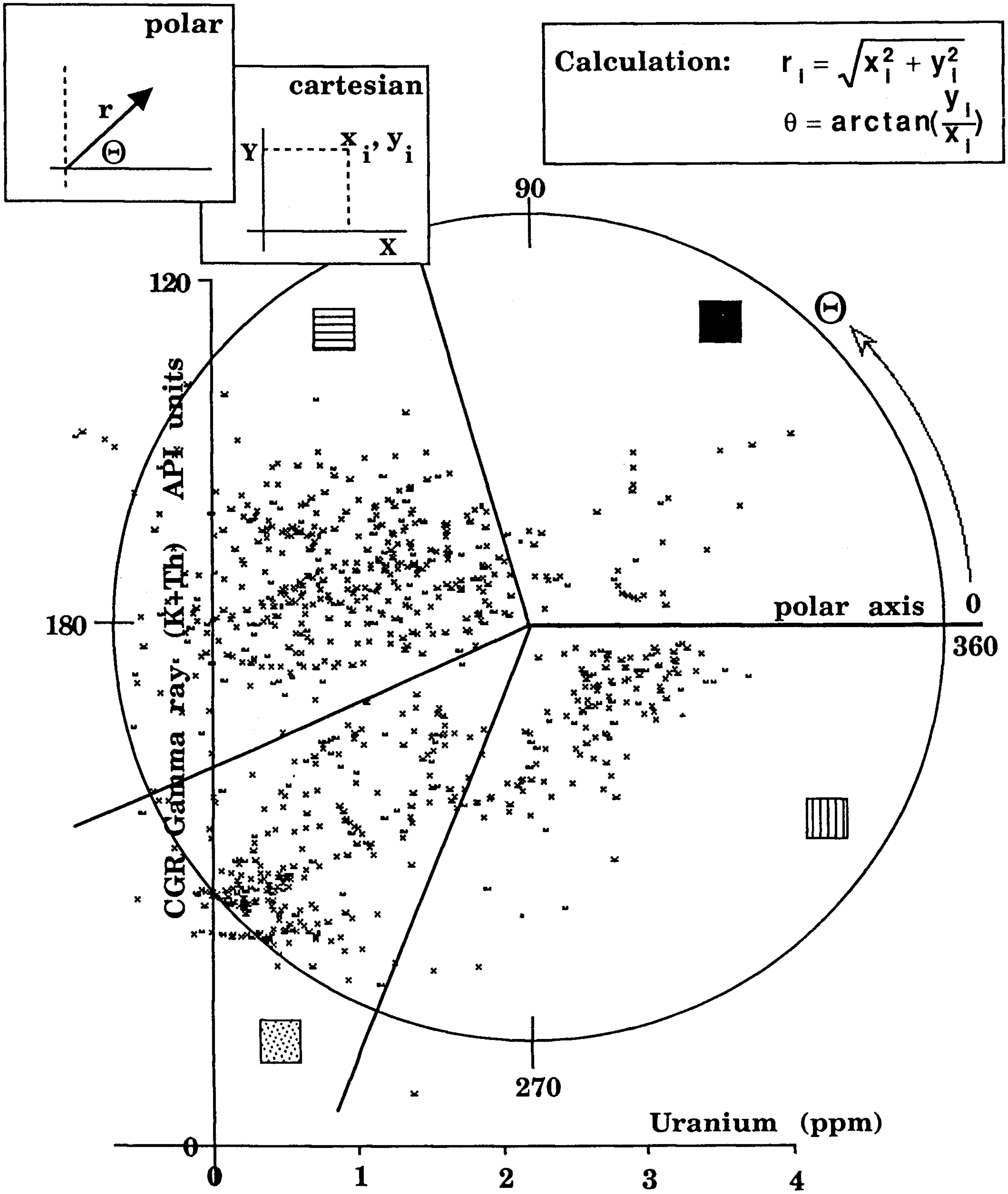
SANDSTONE - SHALE - LIMESTONE RHOMAA - UMAA PLOT

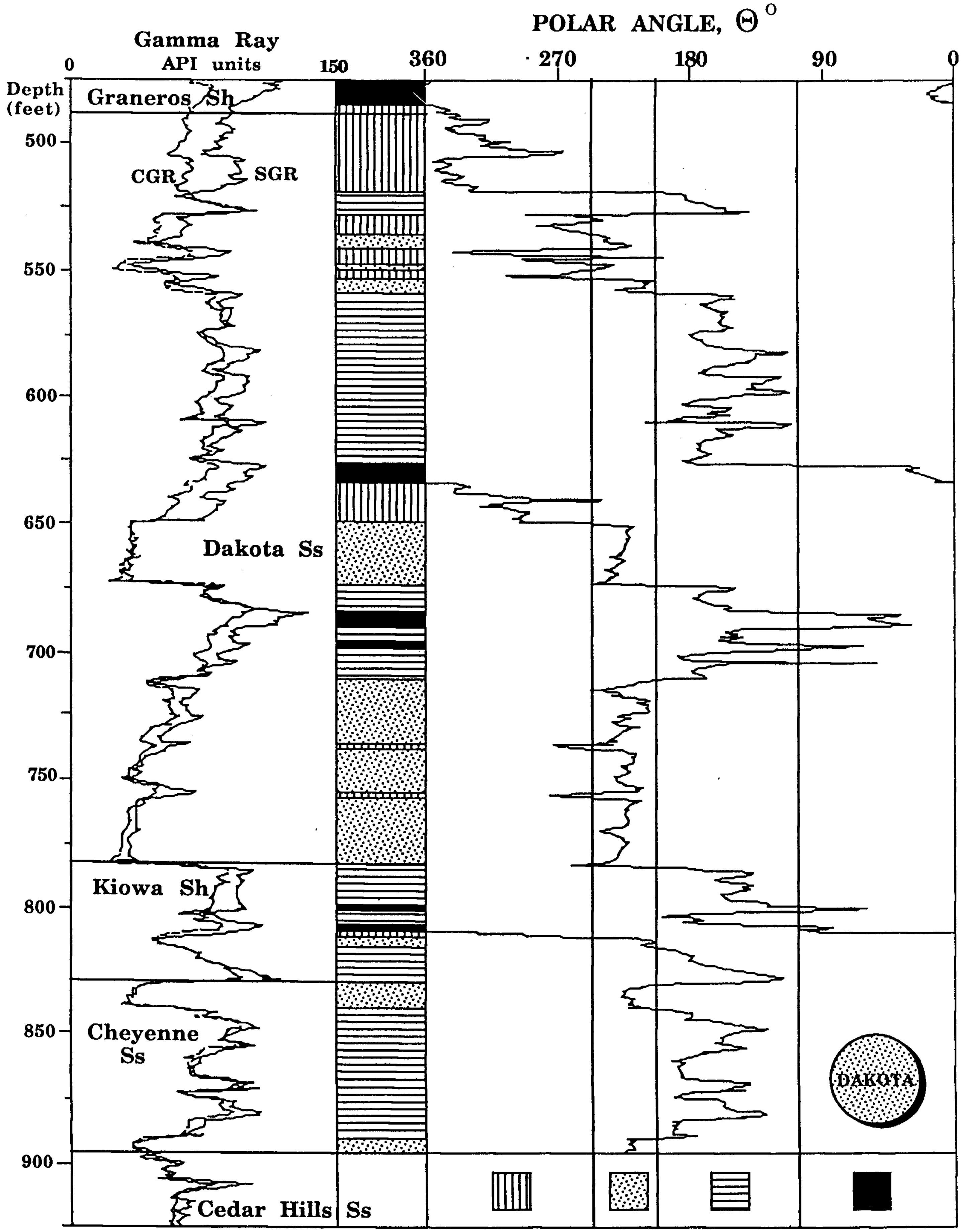



Computed Gamma Ray
0
API units
150
4
UMAA (BARNS/CC)
14
Depth
(feet)
100
200
300
400
500
600
700
800
900
1000
FORT HAYS LS
BLUE HILL
SHALE
FAIRPORT
CHALK
GREENHORN
LIMESTONE
GRANEROS
DAKOTA
KIOWA
CHEYENNE
CEDAR
HILLS
SANDSTONE
Codell Ss
arenaceous
phase
calcareous
phase
DAKOTA

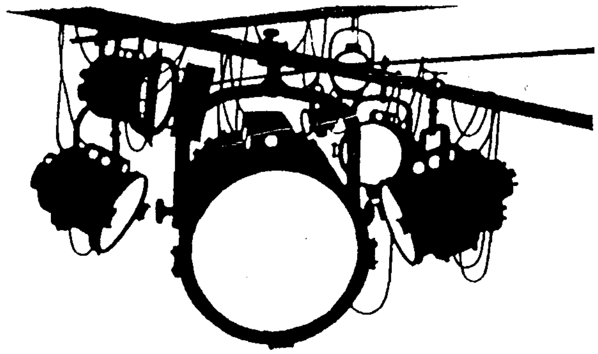
POLAR PROJECTION PLOT :

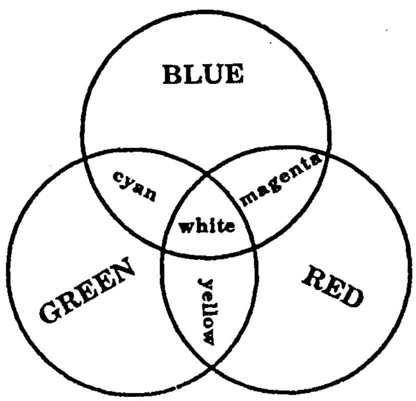
CONVERSION OF CARTESIAN TO POLAR COORDINATES AT USER - SELECTED FOCAL POINT. USE OF POLAR ANGLE, Θ

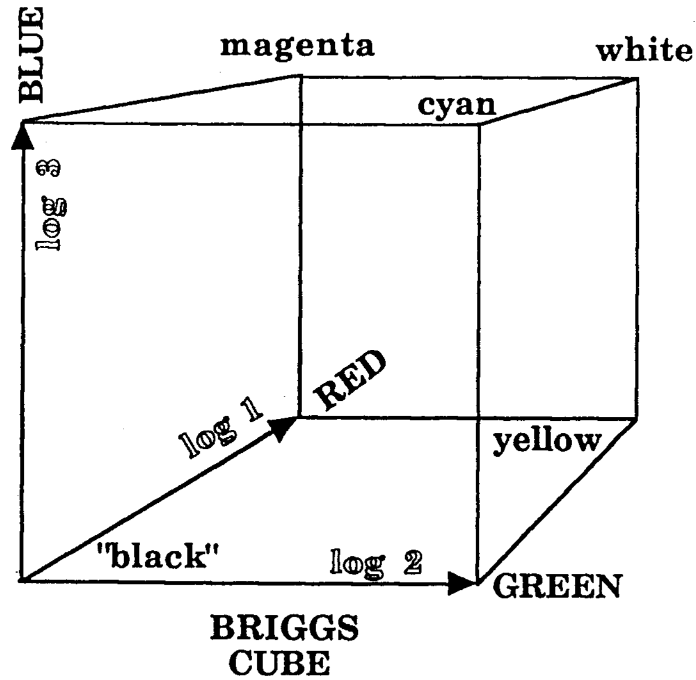


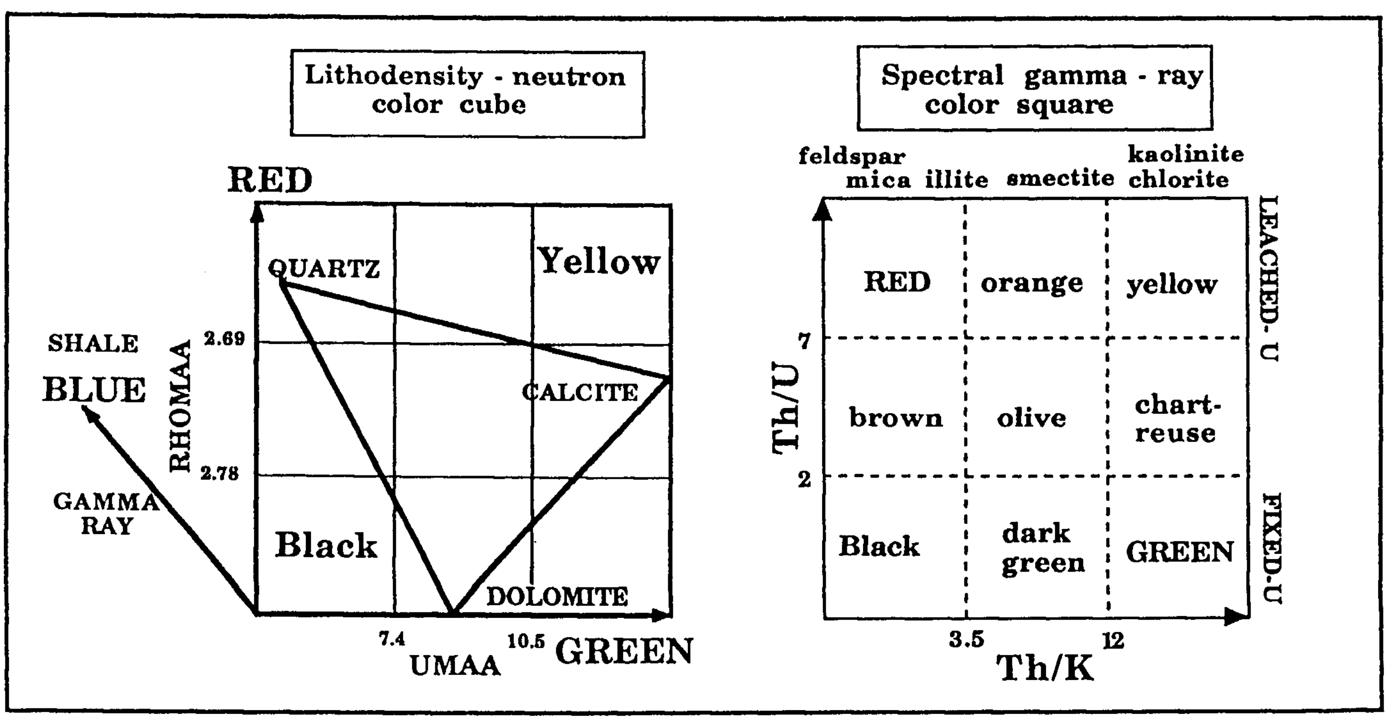

Gamma Ray
API units
0
150
POLAR ANGLE, Θ°
360
270
180
90
0
Depth
(feet)
500
550
600
650
700
750
800
850
900
Graneros Sh
CGR
SGR
Dakota Ss
Kiowa Sh
Cheyenne
Ss
Cedar Hills Ss
DAKOTA

COLOR AS INFORMATION: 3-D LOG CROSSPLOTS AND STRIPLOGS USING THE BRIGGS COLOR CUBE



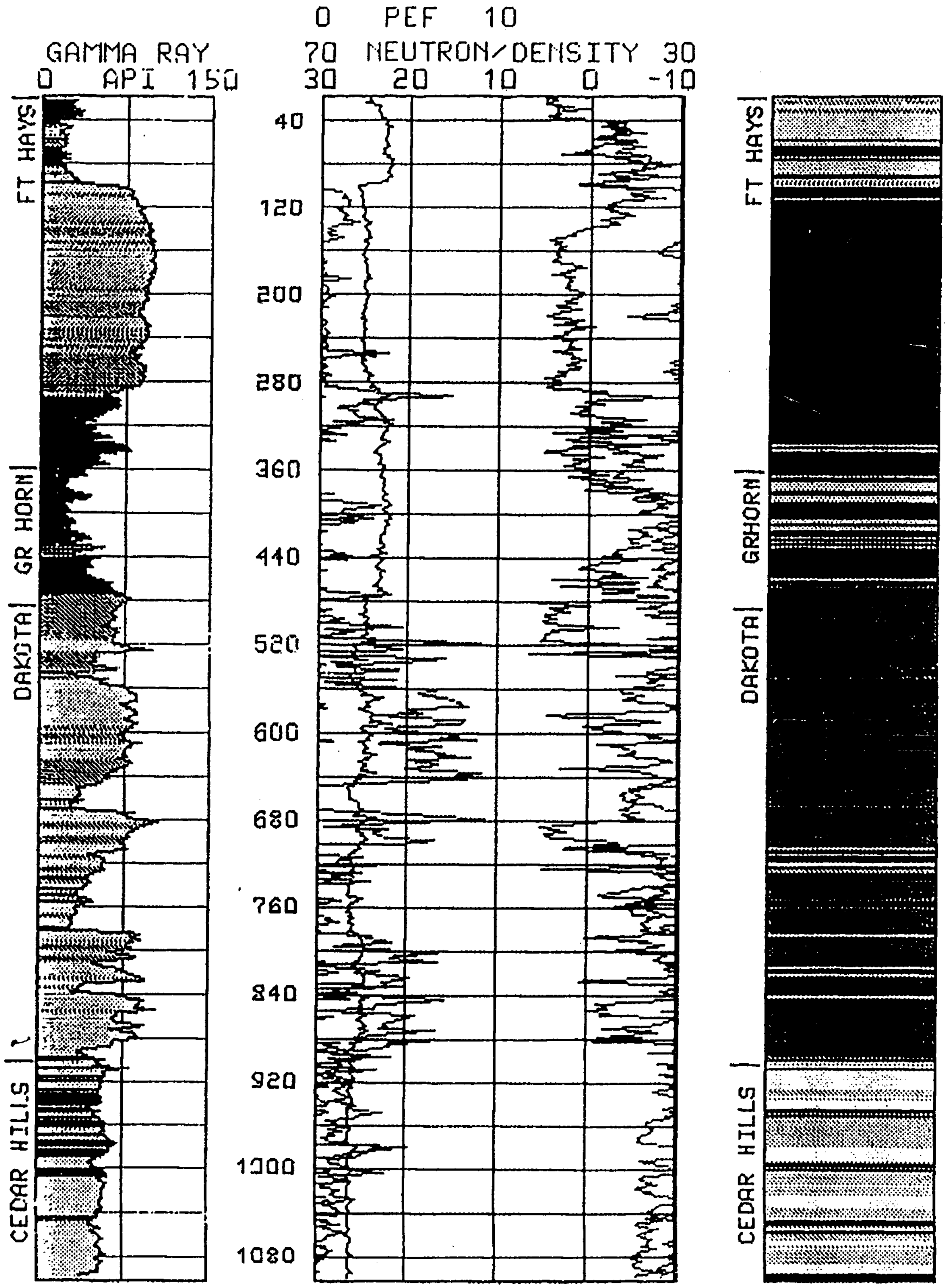






References: Briggs (1985) ; Collins and Doveton (1989)

KGS BRAUN #1 30-12S-18W ELLIS CO.
INITIAL DEPTH - 20.0 FINAL DEPTH - 1100.0
DATA INTERVAL = 2.0



Blank

TOP-DOWN METHODS :

Matrix algebra solutions of normative petrography

$$\boxed{CV = L}$$

BASICS OF MATRIX ALGEBRA

$$\begin{array}{cc} & \text{matrix A} \\ \text{r rows} & \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \\ & \text{c columns} \end{array} \qquad \begin{array}{c} \text{row vector} \\ \begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix} \\ \\ \begin{bmatrix} a_{11} \end{bmatrix} \\ \text{scalar} \end{array} \qquad \begin{array}{cc} \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} & \text{column vector} \end{array}$$

A matrix is denoted by a single capital letter (A), and is a rectangular array of elements which can be numbers or symbols (aii). If a matrix has r rows and c columns, it is said to be of order r X c (or known as an r X c matrix). Element a_{ij} is located in the ith row and the jth column of matrix A. When r = c, the matrix is square. A matrix with only one row is a row vector ; a matrix with only one column is a column vector. A matrix having a single row and a single column (an isolated number or element) is known as a scalar.

$$\text{Diagonal matrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{bmatrix} \qquad \text{Identity matrix (I)} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The leading diagonal of a square matrix corresponds to the a_{ii} elements (upper left to lower right). A symmetric matrix is a square matrix where all elements $a_{ij} = a_{ji}$ and so is symmetrical about the leading diagonal. The trace of a square matrix is the sum of the leading diagonal elements. A diagonal matrix has leading diagonal elements which are non-zero values and zero-value off-diagonal elements. An important diagonal matrix is the identity matrix whose diagonal elements are all one, and is denoted as I.

$$\text{if} \cdots D = \begin{bmatrix} 3 & 7 \\ 1 & 9 \\ 4 & 2 \end{bmatrix} \cdots\cdots \text{then} \cdots D^T = \begin{bmatrix} 3 & 1 & 4 \\ 7 & 9 & 2 \end{bmatrix}$$

The transpose of matrix D is written as D or D and is a matrix whose rows are the columns of D (or equivalently, whose columns are the rows of D.

ADDITION OF MATRICES

$$A + B = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13} \\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23} \\ a_{31}+b_{31} & a_{32}+b_{32} & a_{33}+b_{33} \end{bmatrix}$$

Only matrices of the same order may be added. **A+B = B+A**
Subtraction of matrices is analogous, but with subtraction of appropriate elements.

MULTIPLICATION OF MATRICES

The number of columns in the first matrix must be the same as the number of rows in the second matrix. If C=AB, then each element cij in the C matrix is the sum of the products obtained by multiplying the ith row of A by the jth column of B or :

$$c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj}$$

So, for example :

$$\begin{bmatrix} 3 & 1 & 2 \\ 4 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 3 & 0 \end{bmatrix} = \begin{bmatrix} 10 & 7 \\ 10 & 8 \end{bmatrix}$$

THE INVERSE MATRIX

Division of one matrix by another is not directly possible in the same manner as in scalar algebra. However, the equivalent operation may be made through multiplication by an inverse matrix. Now, if

$$AB=C$$

then the value of the matrix B is conceptually the division of C by A. However, if an inverse matrix A^{-1} can be found such that :

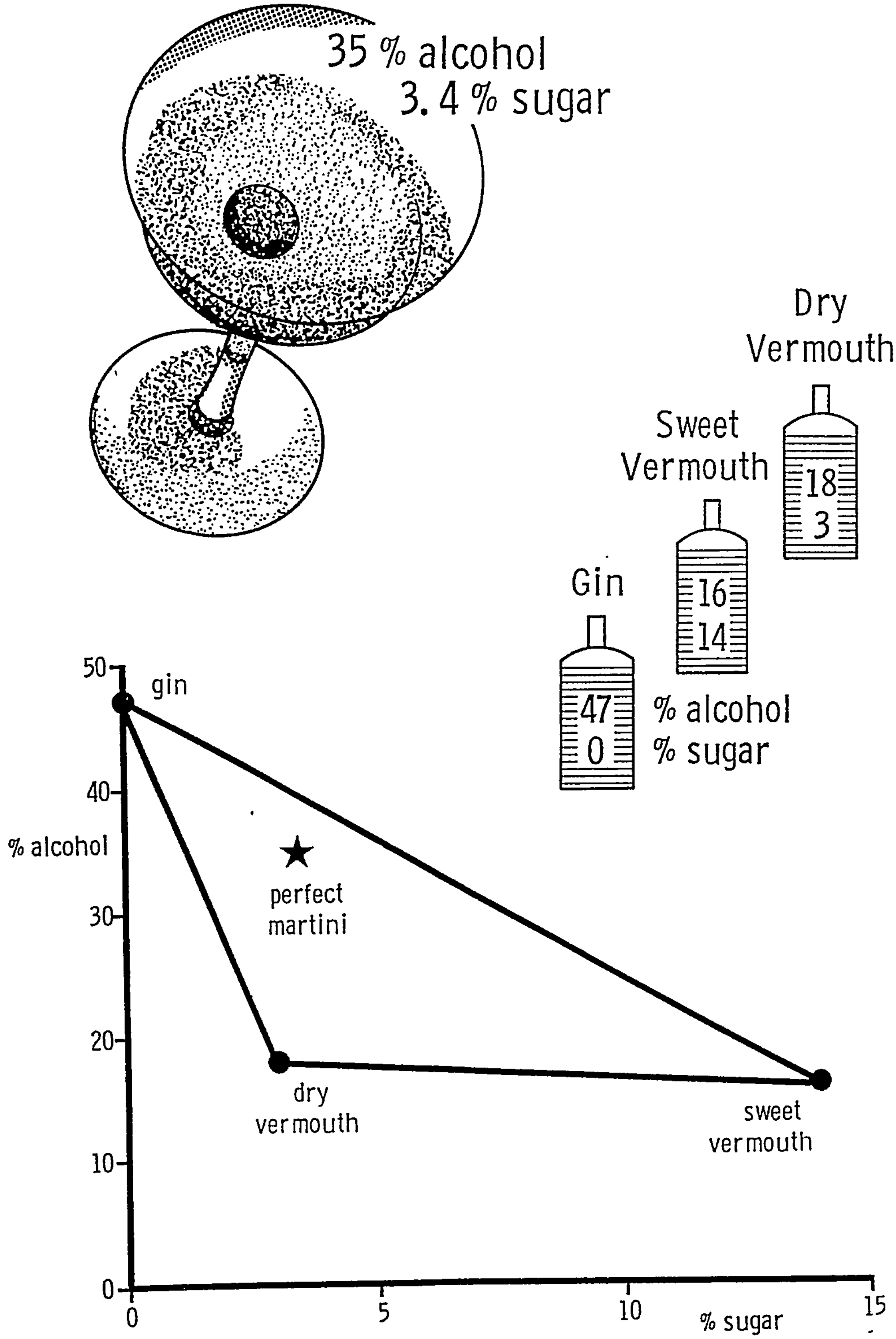
$$AA^{-1} = I \quad \text{and} \quad A^{-1}A = I,$$

where I is an identity matrix, then by multiplication :

$$A^{-1}AB = A^{-1}C$$
$$IB = A^{-1}C$$
$$B = A^{-1}C$$

This operation is most widely used to solve simultaneous equations.

THE PERFECT MARTINI



MATRIX ALGEBRA SOLUTION OF PERFECT MARTINI PROBLEM

G = GIN D = DRY VERMOUTH S = SWEET VERMOUTH

SIMULTANEOUS EQUATIONS

Alcohol:	47G	+	18D	+	16S	=	35
Sugar:	0G	+	3D	+	14S	=	3.4
Unity:	G	+	D	+	S	=	1

MATRIX ALGEBRA FORMAT

$$\begin{bmatrix} 47 & 18 & 16 \\ 0 & 3 & 14 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} G \\ D \\ S \end{bmatrix} = \begin{bmatrix} 35 \\ 3.4 \\ 1 \end{bmatrix}$$

$$C * V = L$$

MATRIX SOLUTION

$$V = C^{-1} * L$$

$$V = \begin{bmatrix} G \\ D \\ S \end{bmatrix} = \begin{bmatrix} 0.60 \\ 0.20 \\ 0.20 \end{bmatrix}$$

kiwi

Program to compute proportions of N components from (N-1) logs, using matrix algebra solution of simultaneous equation set

INPUT:

Well name, section, number of logs, names of logs and components, log responses of components, zone depths and log readings

OPERATION:

* Written in MICROSOFT BASIC (FORTRAN version in Doveton, 1986)
* Runs on IBM PC or compatible
* Interactive data entry version
* Will solve for up to 6 components, but can be expanded

```
100 REM*****************************************************************
110 REM  K I W I :
120 REM  COMPUTATION OF MINERAL AND FLUID COMPONENT PROPORTIONS
121 REM  BASED ON LOG RESPONSES.
122 REM
123 REM  MICROSOFT BASIC; IBM IMPLEMENTATION.
124 REM
125 REM  JOHN H. DOVETON; KANSAS GEOLOGICAL SURVEY; 1981
126 REM  (BASIC ADAPTATION: RICHARD BROWNRIGG; K.G.S.; 1985)
127 REM
128 REM *****************************************************************
130 OPTION BASE 1
140 DIM L$(5),C$(6)
150 DIM V(6,6),V2(6,6)
160 DIM D3(200),R1(6),R2(6),R4(200,6)
161 OPEN "O",#1,"KIWIWORK.TMP"
170 PRINT "*** K I W I PROGRAM ***"
180 PRINT "NAME OF WELL ";
190 LINE INPUT "? ";N$
200 PRINT "NAME OF SECTION ";
210 LINE INPUT "? ";S$
211 GOSUB 2000
220 PRINT
230 PRINT "NUMBER OF LOGS ";
240 INPUT N
250 IF N>5 THEN PRINT "***NO MORE THAN 5 CAN BE ACCEPTED!":LET N=5
260 FOR I=1 TO N
270   PRINT "NAME OF LOG ";I;
280   INPUT L$(I)
290 NEXT I
291 GOSUB 2080
300 LET M=N+1
310 PRINT
320 FOR I=1 TO M
330   PRINT "NAME OF COMPONENT ";I;
340   INPUT C$(I)
350 NEXT I
351 GOSUB 2190
360 PRINT
370 FOR I=1 TO N
380   FOR J=1 TO M
390     PRINT L$(I);" VALUE FOR ";C$(J);
400     INPUT V(I,J)
410   NEXT J
420   PRINT
430 NEXT I
440 FOR J=1 TO M
450   LET V(M,J)=1
460 NEXT J
461 GOSUB 2280
480 REM ****************************
490 REM INVERT MATRIX OF LOG COEFFICIENTS:
500 REM   V = MATRIX OF VALUES;  V2 = INVERTED MATRIX
510 REM   D = DETERMINATE      D2 = DIVISOR
520 REM   R = RATIO
530 REM ****************************
540 FOR I=1 TO M
550   LET V2(I,I)=1
560 NEXT I
570 LET D=1
580 FOR I=1 TO M
590   LET D2=V(I,I)
600   LET D=D*D2
610   FOR J=1 TO M
620     LET V(I,J)=V(I,J)/D2
630     LET V2(I,J)=V2(I,J)/D2
640   NEXT J
650   FOR J=1 TO M
660     IF (I-J)=0 THEN GOTO 720
670     LET R=V(J,I)
680     FOR K=1 TO M
690       LET V(J,K)=V(J,K)-R*V(I,K)
700       LET V2(J,K)=V2(J,K)-R*V2(I,K)
710     NEXT K
720   NEXT J
730 NEXT I
740 REM ****************************
750 REM NOW READ AND PROCESS LOG RESPONSES:
780 REM   D3(..) = DEPTHS          R1(..) = TMP.VR.,CALC. OF PROP.
790 REM   R2(..) = INPUT LOG VALUES   R3 = TMP.VR.
800 REM   R4(..) = COMPONENT'S PROPORTIONS, INDEXED BY DEPTH
801 REM ****************************
830 LET Z=0
840 LET R2(M)=1
850 PRINT
860 PRINT "ENTER LOG READINGS FOR EACH ZONE, AS:"
870 PRINT
880 PRINT "DEPTH";
881 GOSUB 2470
890 FOR I=1 TO N
900   PRINT ",";L$(I);
910 NEXT I
920 PRINT
930 PRINT
940 PRINT "...ONE ZONE PER LINE.   ENTER -1 FOR DEPTH TO QUIT."
950 FOR L=1 TO 200
960 LINE INPUT "? ";A$
970 LET A=0
980 GOSUB 1270
981 IF X<=0 GOTO 9000
```

```
990 LET D3(L)=X
1000 FOR I=1 TO N
1010   GOSUB 1270
1020   LET R2(I)=X
1030 NEXT I
1070   LET Z=Z+1
1080   FOR I=1 TO M
1090     LET R1(I) = 0
1100     FOR J=1 TO M
1110       LET R1(I)=R1(I)+V2(I,J)*R2(J)
1120     NEXT J
1130   NEXT I
1131   GOSUB 2610
1140   LET R3=0
1150   FOR I=1 TO M
1160     IF R1(I)<0 THEN LET R1(I)=0
1170     LET R3=R3+R1(I)
1180   NEXT I
1190   FOR I=1 TO M
1200     LET R1(I)=100*R1(I)/R3
1210     LET R4(L,I)=R1(I)
1220   NEXT I
1230 NEXT L
1260 REM ************************************************************
1270 REM  S U B R O U T I N E: ISOLATES THE NEXT NUMERIC STRING IN A$
1280 LET A=A+1
1290 IF A>LEN(A$) GOTO 1420
1300 LET X=ASC(MID$(A$,A,1))
1310 IF X<45 GOTO 1280
1320 IF X>57 GOTO 1280
1330 LET B=A
1340 LET A=A+1
1350 IF A>LEN(A$) GOTO 1400
1360 LET X=ASC(MID$(A$,A,1))
1370 IF X<45 GOTO 1400
1380 IF X>57 GOTO 1400
1390 GOTO 1340
1400 LET X=VAL(MID$(A$,B,(A-B)))
1410 RETURN
1420 LET X=0
1430 RETURN
1999 REM ************************************************************
2000 REM SUBROUTINE: PRINT HEADER, AND WELL NAME:
2001 PRINT #1,
2010 PRINT #1," * * * K I W I   P R O G R A M * * *"
2020 PRINT #1,
2030 PRINT #1,
2040 PRINT #1,
2050 PRINT #1," WELL NAME: ";N$
2060 PRINT #1,"    SECTION: ";S$
2070 RETURN
2079 REM ************************************************************
2080 REM SUBROUTINE: PRINT LOG KEY...
2090 PRINT #1,
2100 PRINT #1,
2110 PRINT #1,
2120 PRINT #1," KEY TO LOGS:"
2140 FOR I=1 TO N
2150   PRINT #1,USING "   LOG #=\           \";I;L$(I)
2170 NEXT I
2180 RETURN
2189 REM ************************************************************
2190 REM SUBROUTINE: PRINT THE KEY TO THE COMPONENTS...
2200 PRINT #1,
2210 PRINT #1,
2220 PRINT #1," KEY TO COMPONENTS:"
2230 FOR I=1 TO M
2240   LET A$=CHR$(64+I)
2250   PRINT #1,"    COMPONENT ";A$;"=";C$(I)
2260 NEXT I
2270 RETURN
2279 REM ************************************************************
2280 REM SUBROUTINE: PRINT THE LOG COEFFICIENTS...
2290 PRINT #1,
2300 PRINT #1,
2310 PRINT #1,
2320 PRINT #1," LOG COEFFICIENTS:"
2330 PRINT #1,"         ";
2340 FOR I=1 TO M
2350   LET A$=CHR$(64+I)
2360   PRINT #1,"       ";A$;
2370 NEXT I
2380 PRINT #1,
2390 FOR I=1 TO N
2400   PRINT #1,USING "   LOG # ";I;
2410   FOR J=1 TO M
2420     PRINT #1,USING "#######.##";V(I,J);
2430   NEXT J
2440   PRINT #1,
2450 NEXT I
2460 RETURN
2469 REM ************************************************************
2470 REM SUBROUTINE:  PRINT HEADING FOR RESPONSES AND PROPORTIONS OUTPUT...
2480 PRINT #1,
2490 PRINT #1,
2500 PRINT #1,
2510 PRINT #1," LOG RESPONSES AND COMPONENT PROPORTIONS:"
2511 PRINT #1,"   DEPTH";
2520 FOR I=1 TO N
2530   PRINT #1,USING "   # ";I;
2540 NEXT I
2550 PRINT #1,"    ";
2560 FOR I=1 TO M
2570   PRINT #1,"     ";CHR$(64+I);"  ";
2580 NEXT I
2590 PRINT #1,
2600 RETURN
2609 REM ************************************************************
2610 REM SUBROUTINE: PRINT USER'S LOG RESPONSES AND THEIR PROPORTIONS...
2620 PRINT #1,USING "########";D3(L);
2630 FOR A=1 TO N
2640   PRINT #1,USING "####.##";R2(A);
2650 NEXT A
2660 PRINT #1,"    ";
2670 FOR A=1 TO M
2680   PRINT #1,USING "####.##";R1(A);
2690 NEXT A
2700 PRINT #1,
2710 RETURN
9000 REM ************************************************************
9010 REM FINALLY,  GENERATE THE GRAPHIC...
9020 REM
9030 PRINT
9040 PRINT " SCALE PLOT FOR 133 CHARACTERS/LINE OUTPUT (Y/N)";
9050 INPUT Y$
9060 IF Y$="Y" GOTO 9120
9070 PRINT " ...PLOT WILL BE SCALED FOR 80 COLUMN OUTPUT."
9080 W=2
9091 LET P1$="*..10...20...30...40...50...60...70...80...90..100* %"
9100 LET P2$="*"+STRING$(50," ")+"*"
9110 GOTO 9150
9120 W=1
9131 LET P1$="*....+...10....+...20....+...30....+...40....+...50....+
        ...60....+...70....+...80....+...90....+..100* %"
9140 LET P2$="*"+STRING$(100," ")+"*"
9150 PRINT #1,
9160 PRINT #1,
9170 PRINT #1,
9180 PRINT #1," GRAPHIC COMPONENT LOG:"
9190 PRINT #1,
9200 PRINT #1," DEPTH";P1$
9210 LET D4=D3(1)
9220 FOR L=1 TO Z
9230   PRINT #1,USING "######";D4;
9240   IF D3(L)<=D4 GOTO 9280
9250   PRINT #1,P2$
9260   LET D4=D4+1
9270   GOTO 9230
9280   LET P3$=P2$
9290   MID$(P3$,LEN(P3$)-1,1)=CHR$(64+M)
9300   LET K=1
9310   FOR J=1 TO N
9320     IF R4(L,J)<0 GOTO 9360
9330     LET I=K+(R4(L,J)\W)
9340     MID$(P3$,I,1)=CHR$(64+J)
9350     LET K=I
9360   NEXT J
9370   PRINT #1,P3$
9371   LET D4=D4+1
9380 NEXT L
9390 PRINT #1," DEPTH";P1$
9400 PRINT #1,"     SCALE = ";W;" UNITS/SPACE."
9401 PRINT #1,
9410 PRINT #1,
9420 CLOSE #1
9421 PRINT:PRINT " ROUTE OUTPUT TO PRINTER (Y/N)";
9422 INPUT Y$
9423 IF Y$<>"Y" THEN PRINT:PRINT " AT DOS-LEVEL, ENTER 'TYPE
        KIWIWORK.TMP' TO VIEW THE OUTPUT":GOTO 9500
9430 OPEN "I",#1,"KIWIWORK.TMP"
9440 IF EOF(1) GOTO 9480
9450 LINE INPUT #1,A$
9460 LPRINT A$
9470 GOTO 9440
9480 CLOSE
9490 KILL "KIWIWORK.TMP"
9500 END
```

EXAMPLE OF KIWI PROGRAM RUN IN VIOLA LIMESTONE SECTION OF CHERTY DOLOMITES AND LIMESTONES

KIWI PROGRAM

****

CITIES SERVICE BECK A-1 SW-SE-SE 14-5S-12E
LOWER VIOLA (MIDDLE ORDOVICIAN)

****

KEY TO LOGS
- LOG 1 = NEUTRON
- LOG 2 = DENSITY
- LOG 3 = SONIC

KEY TO COMPONENTS
- COMPONENT A=DOLOMITE
- COMPONENT B=CHERT
- COMPONENT C=CALCITE
- COMPONENT D=POROSITY

****

LOG COEFFICIENTS	A	B	C	D
LOG 1	5.00	-5.00	.00	100.00
LOG 2	2.87	2.65	2.71	1.00
LOG 3	43.50	55.10	47.50	189.00

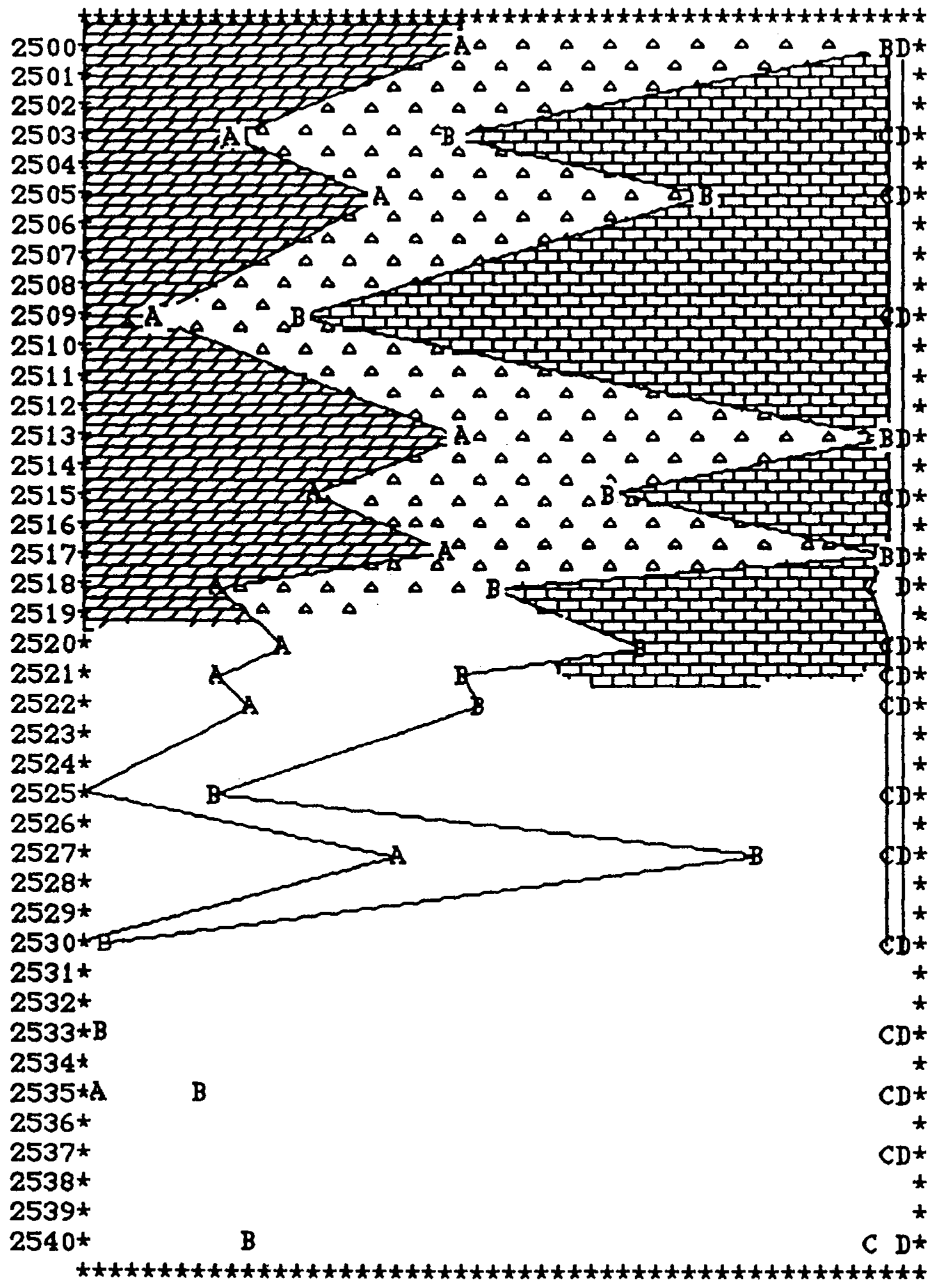
****

LOG RESPONSES AND COMPONENT PROPORTIONS

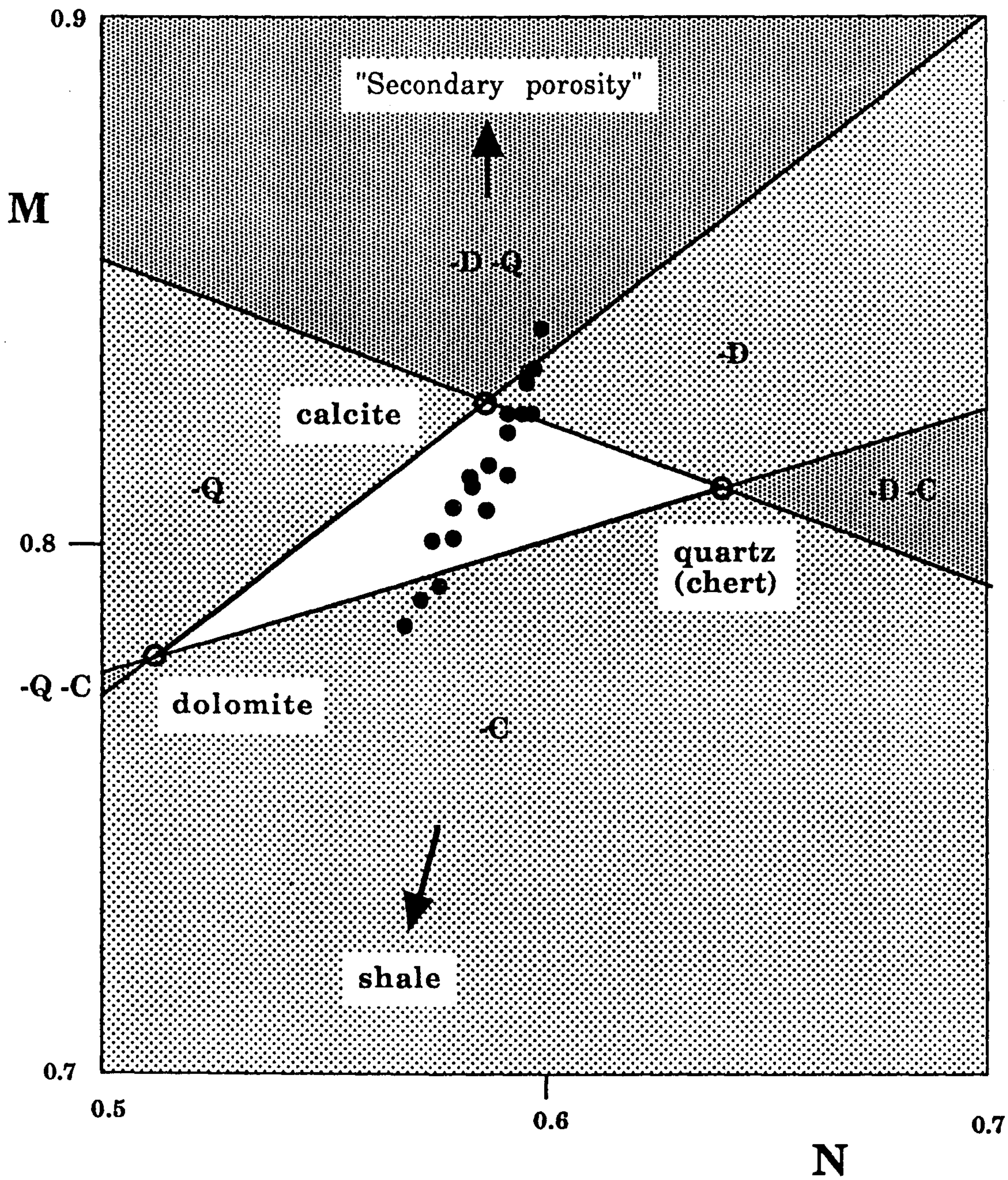
DEPTH	1	2	3	A	B	C	D
2500	2.40	2.72	53.80	.59	.64	-.26	.03
2503	2.00	2.68	52.30	.17	.27	.53	.02
2505	1.70	2.71	51.90	.36	.41	.21	.02
2509	1.80	2.67	51.90	.07	.18	.73	.02
2513	1.80	2.72	52.90	.51	.58	-.11	.02
2515	2.00	2.69	52.60	.27	.36	.35	.02
2517	1.70	2.71	53.40	.47	.59	-.09	.02
2518	1.50	2.67	53.10	.15	.35	.48	.03
2520	.90	2.69	52.50	.24	.43	.31	.02
2521	1.00	2.69	51.40	.16	.29	.53	.02
2522	1.20	2.70	50.90	.20	.27	.51	.02
2525	1.00	2.67	51.20	.00	.15	.83	.02
2527	.90	2.72	51.30	.37	.45	.16	.01
2530	1.20	2.66	50.80	-.10	.03	1.05	.02
2533	.90	2.67	50.00	-.09	.01	1.06	.01
2535	.80	2.68	50.40	.01	.13	.85	.01
2537	1.20	2.65	50.10	-.23	-.11	1.32	.02
2540	1.70	2.65	52.90	-.01	.20	.78	.03

****

GRAPHIC COMPONENT LOG



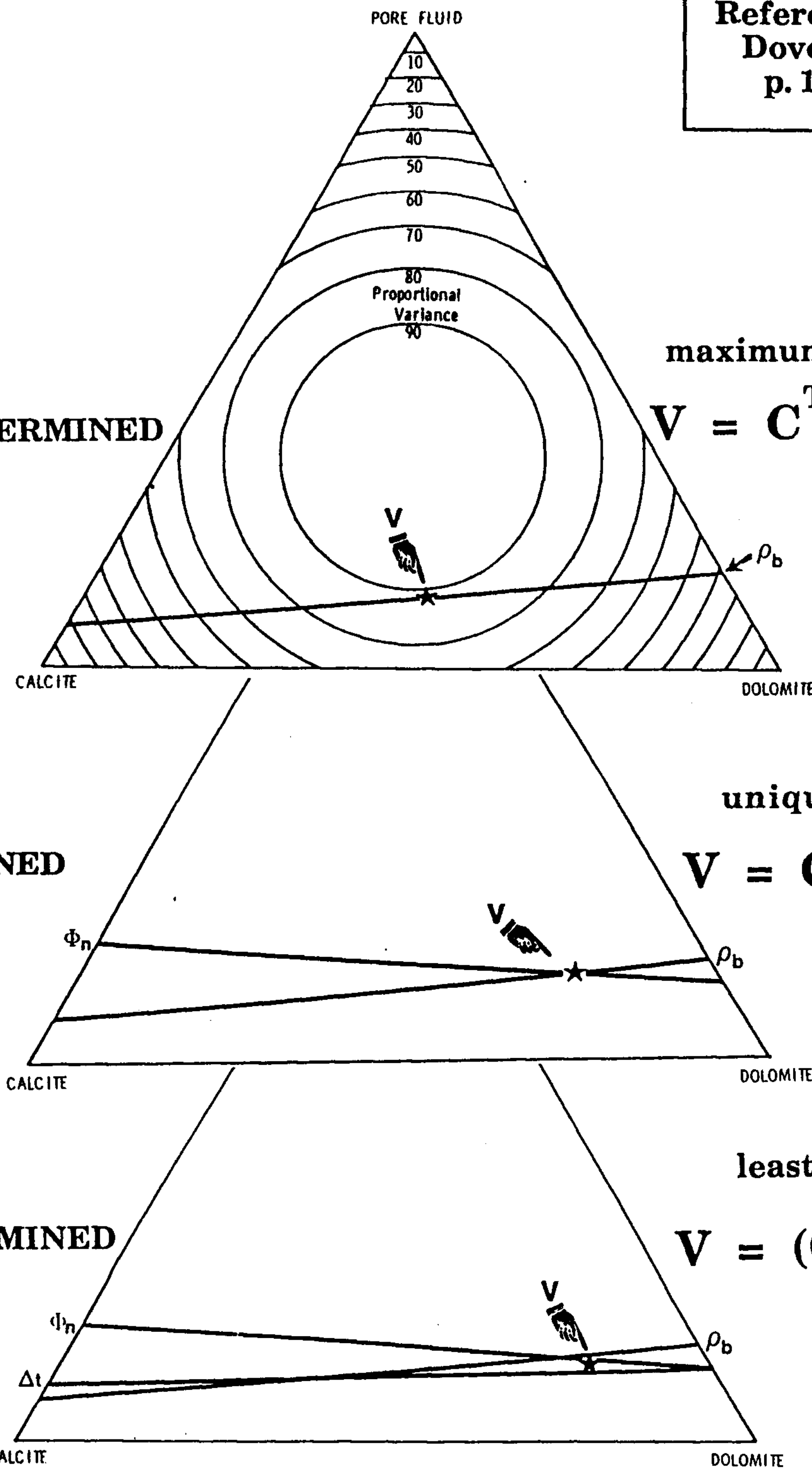
EXPLANATION OF " NEGATIVE COMPONENTS" IN CONTEXT OF KIWI OUTPUT EXAMPLE



MATRIX ALGEBRA SOLUTIONS OF COMPONENT PROPORTIONS (V) FROM LOG READINGS (L), BASED ON COMPONENT LOG RESPONSES (C), FOR ALL SYSTEMS OF DETERMINACY

$$CV = L$$

Reference :
Doveton (1986)
p. 170 - 177



UNDERDETERMINED

maximum variance

$$V = C^T(CC^T)^{-1} L$$

DETERMINED

unique

$$V = C^{-1}L$$

OVERDETERMINED

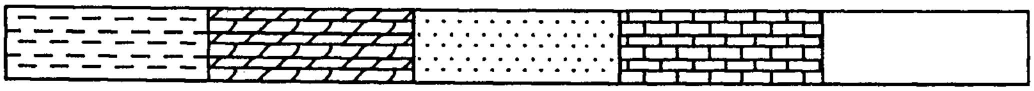
least - squares

$$V = (C^TC)^{-1} C^TL$$

EXAMPLE OF A MATRIX ALGEBRA SOLUTION OF COMPONENTS IN A COMPLEX CARBONATE

```
WELL NAME:    CHASE
LOCATION:                                  DATE:
DEPTH:        2650.00 TO  3100.00 BY   .50 FEET
LOGS:         UTOT      CGR       DENS       CNL%
COMPONENTS:    SH      DOL    SS      LS     POR
```

KEY



SH DOL SS LS POR

INPUT COEFFICIENTS:

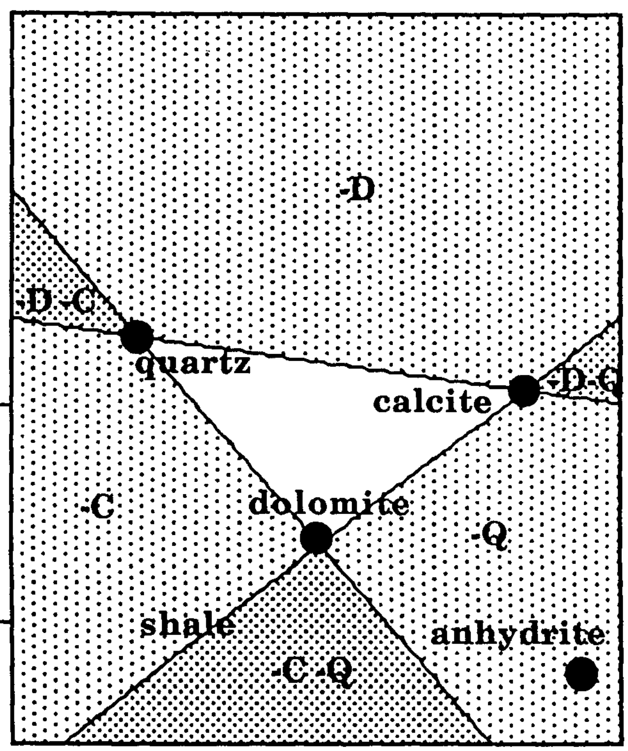
	SH	DOL	SS	LS	POR
UTOT	6.20	9.00	4.78	13.80	.50
CGR	100.00	5.00	5.00	5.00	5.00
DENS	2.40	2.87	2.65	2.71	1.00
CNL%	28.00	3.00	-4.00	.00	100.00

EQUATIONS EQUAL UNKNOWNS   N+1=M

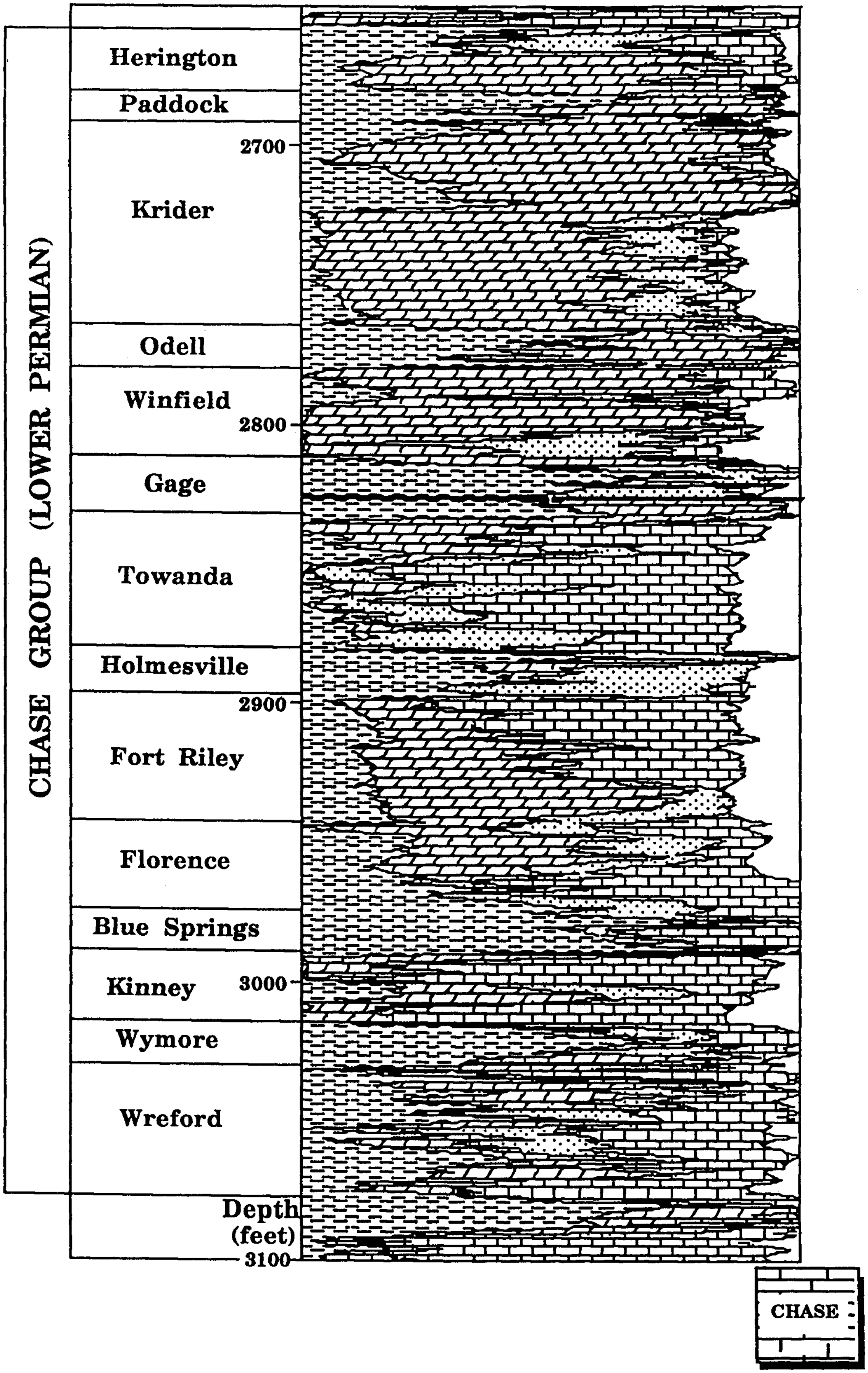
SOLUTION CALCULATED BY MATRIX INVERSION ....

Key to interpretation of "negative components"

RHOMAA



UMAA


CHASE GROUP (LOWER PERMIAN)
Herington
Paddock
2700
Krider
Odell
Winfield
2800
Gage
Towanda
Holmesville
2900
Fort Riley
Florence
Blue Springs
Kinney
3000
Wymore
Wreford
Depth (feet)
3100
CHASE

EXAMPLE OF A MATRIX ALGEBRA SOLUTION OF COMPONENTS IN A SANDSTONE - SHALE SEQUENCE

WELL NAME: BRAUN
LOCATION: ELLIS DATE: 1/25/88
DEPTH: 488.00 TO 895.00 BY .50 FEET
LOGS: UMAA RHOMAA
COMPONENTS: HI-Z LO-Z SS

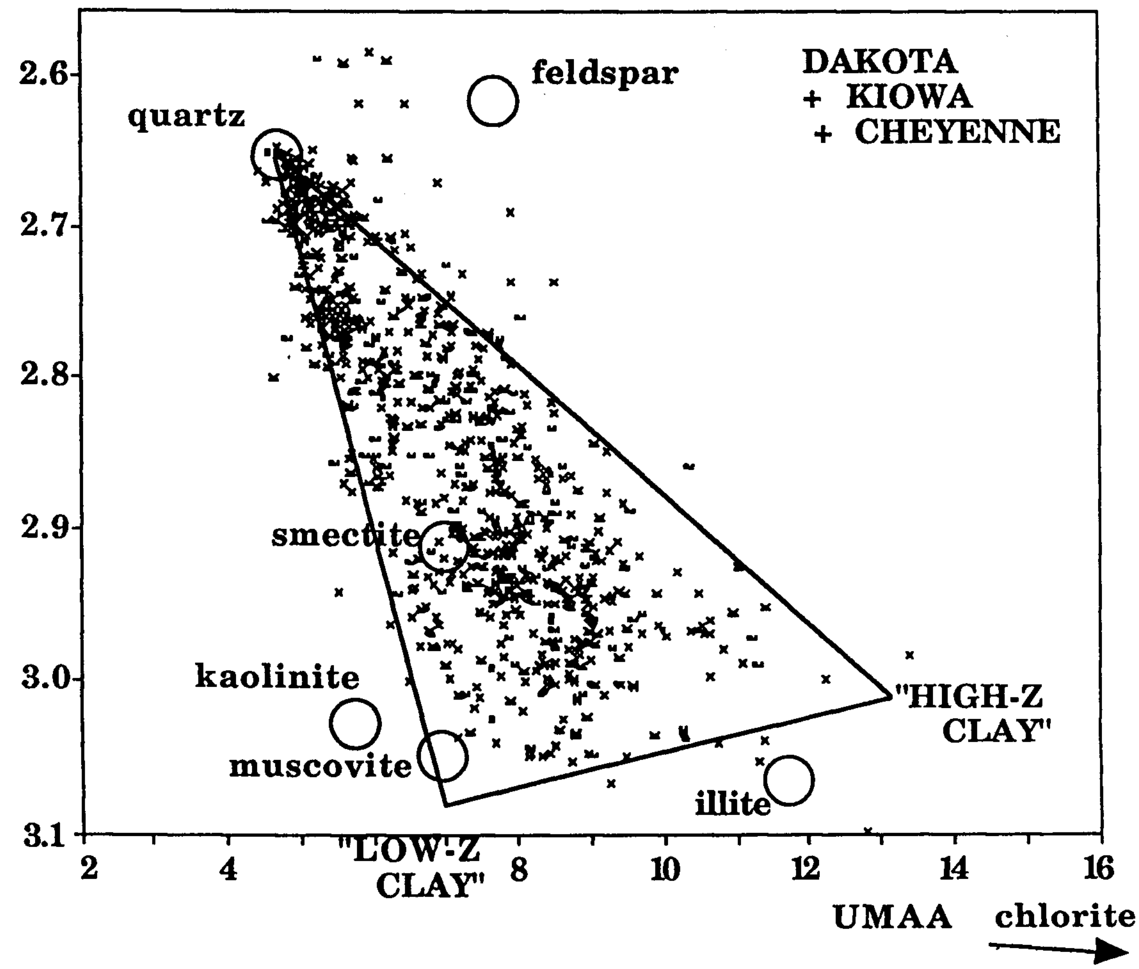
KEY

INPUT COEFFICIENTS:

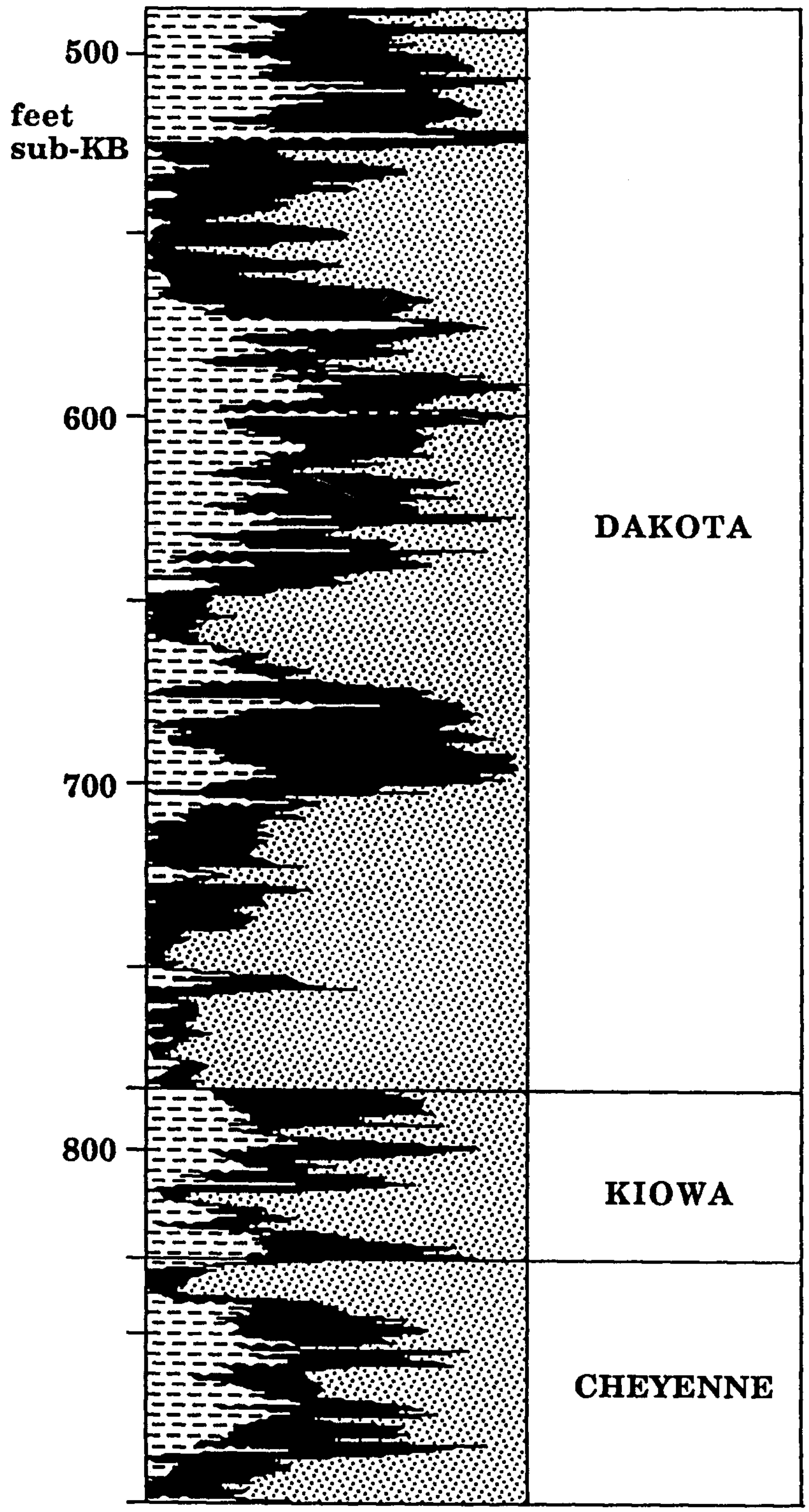
	HI-Z	LO-Z	SS
UMAA	13.20	7.00	4.79
RHOMAA	3.02	3.08	2.64

EQUATIONS EQUAL UNKNOWNS N+1=M

SOLUTION CALCULATED BY MATRIX INVERSION ....

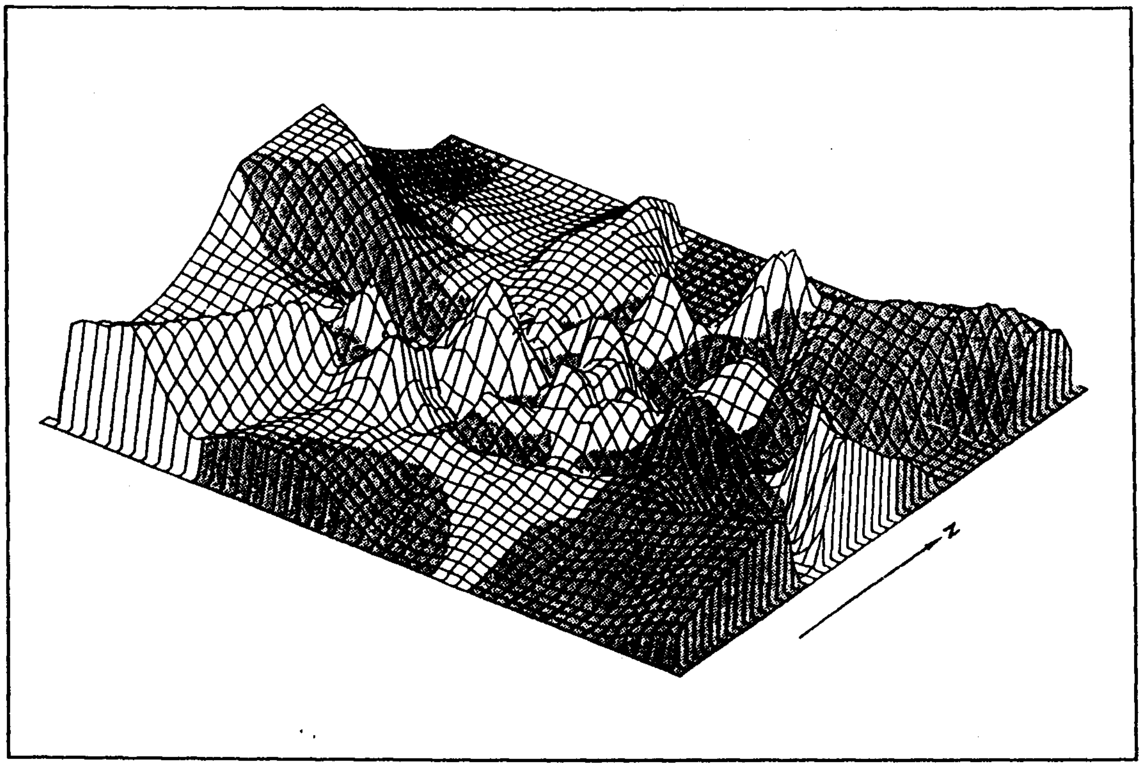


COMPOSITION PROFILE

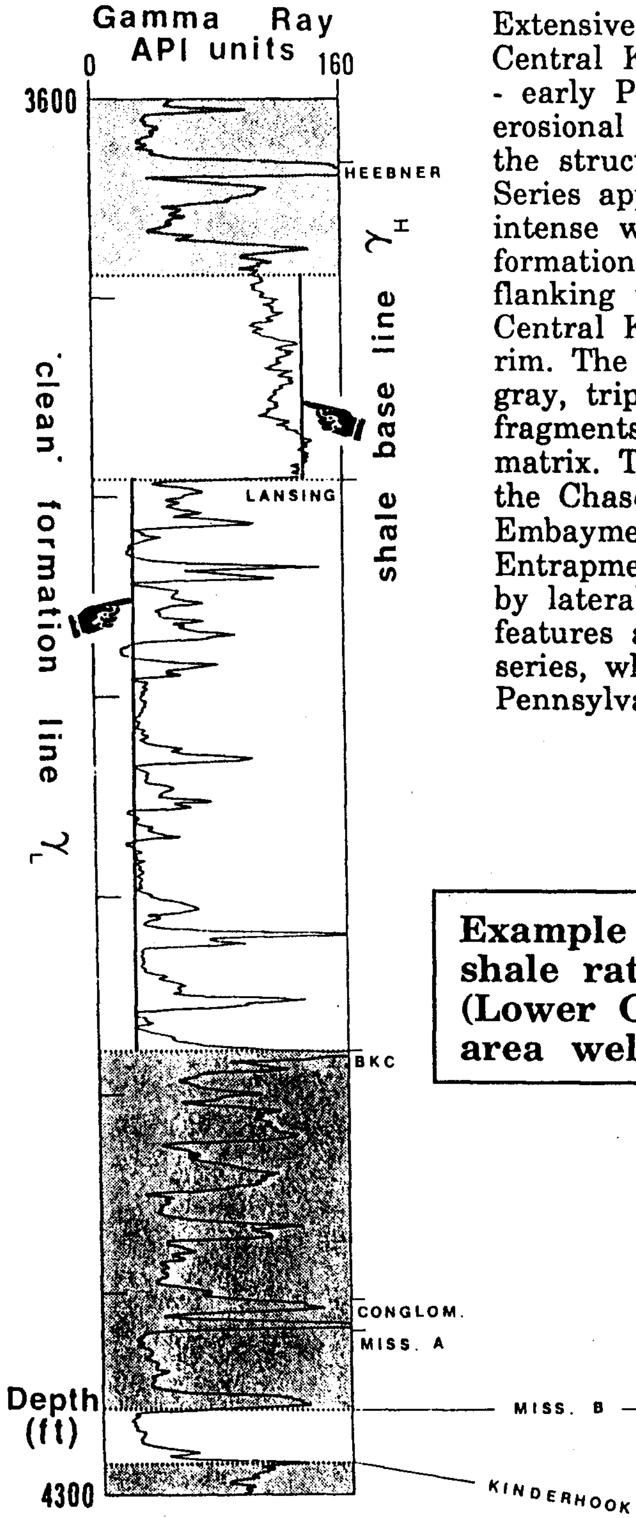


Blank

MAPPING APPLICATIONS

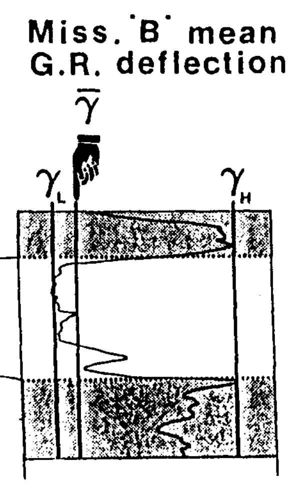


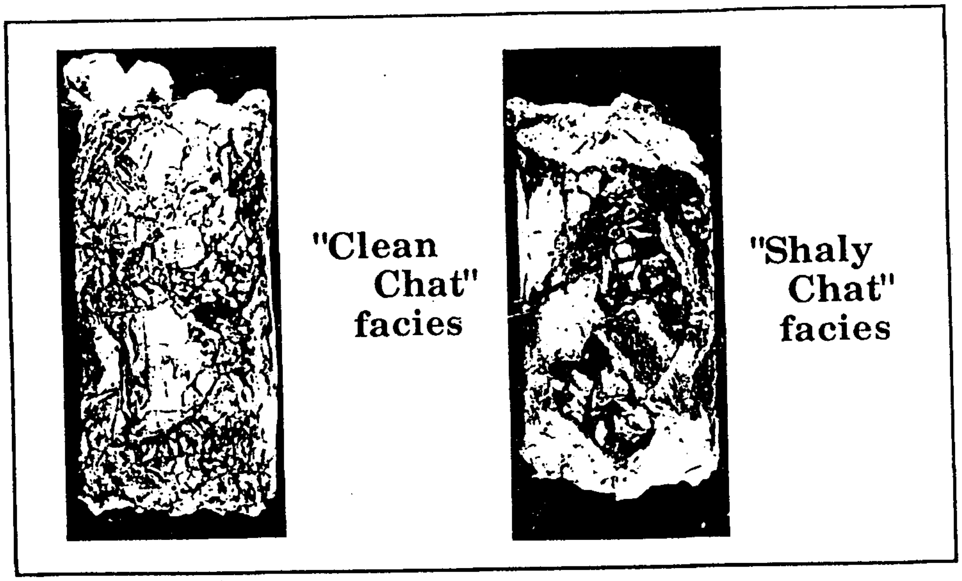
SHALE RATIO MAPPING IN THE MISSISSIPPI CHAT

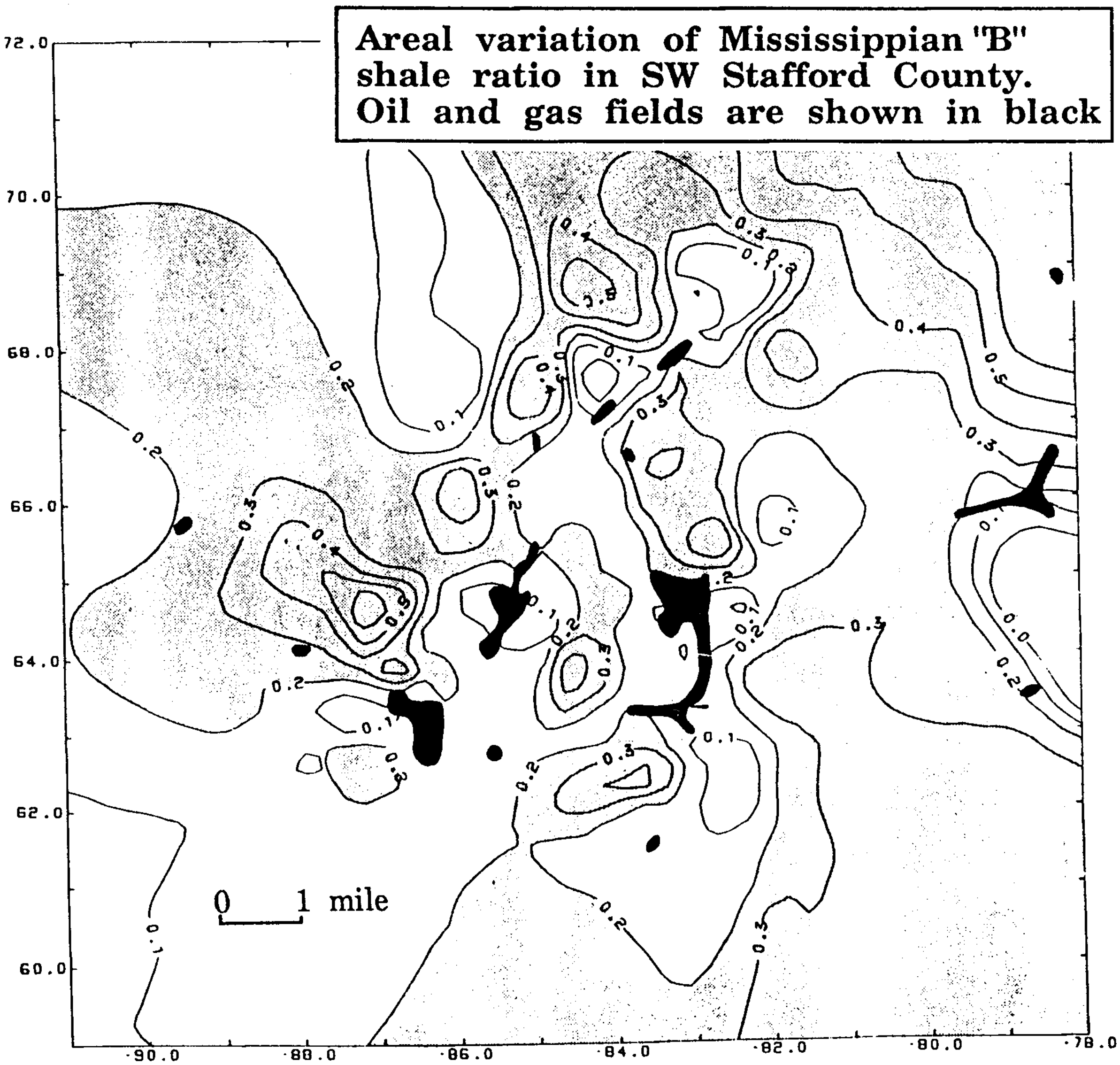


Extensive relative upward movement of the Central Kansas Uplift in late Mississippian - early Pennsylvanian times caused erosional stripping from the higher part of the structure. The Mississippian Osage Series appears to have been subjected to intense weathering, with a resulting formation of a thick residuum of chert flanking the southern margin of the Central Kansas Uplift as a broad arcuate rim. The Osage consists mostly of white to gray, tripolitic to porcelaneous chert fragments in a reddish or gray-green clay matrix. The Osage Series is second only to the Chase - Council Grove of the Hugoton Embayment as a source of gas in Kansas. Entrapment of gas appears to be controlled by lateral permeability changes, structural features and pinchouts at the edge of the series, which is overlapped by the basal Pennsylvanian unconformity.

Example of calculation of mean shale ratio of the "Mississippian B" (Lower Osage Series) in a study area well





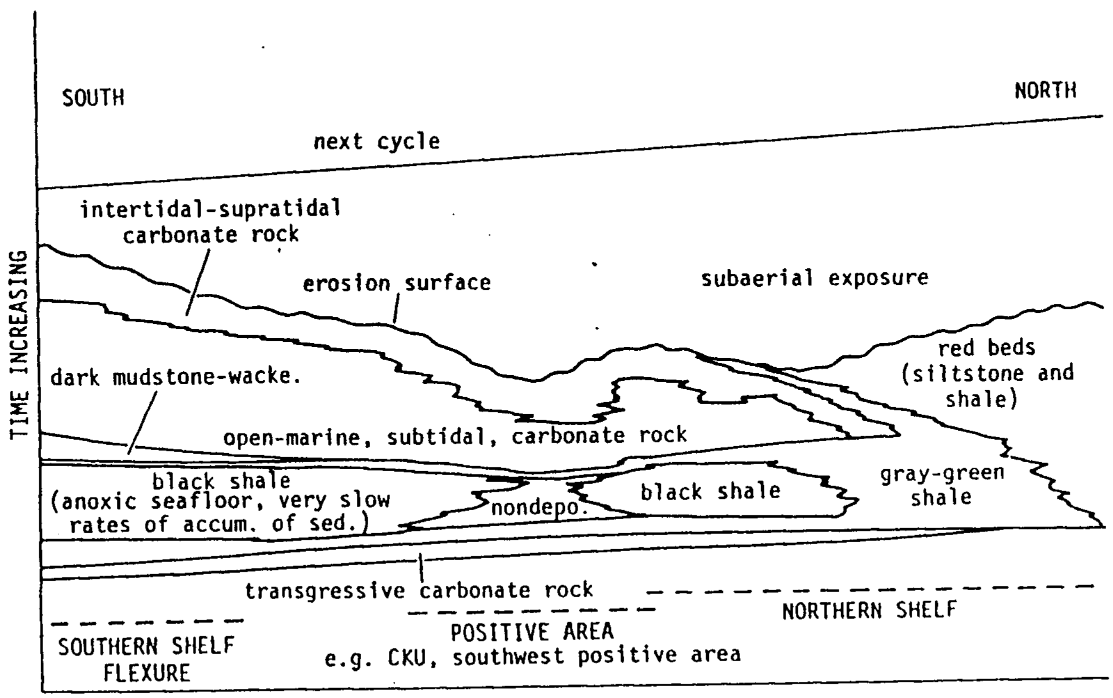


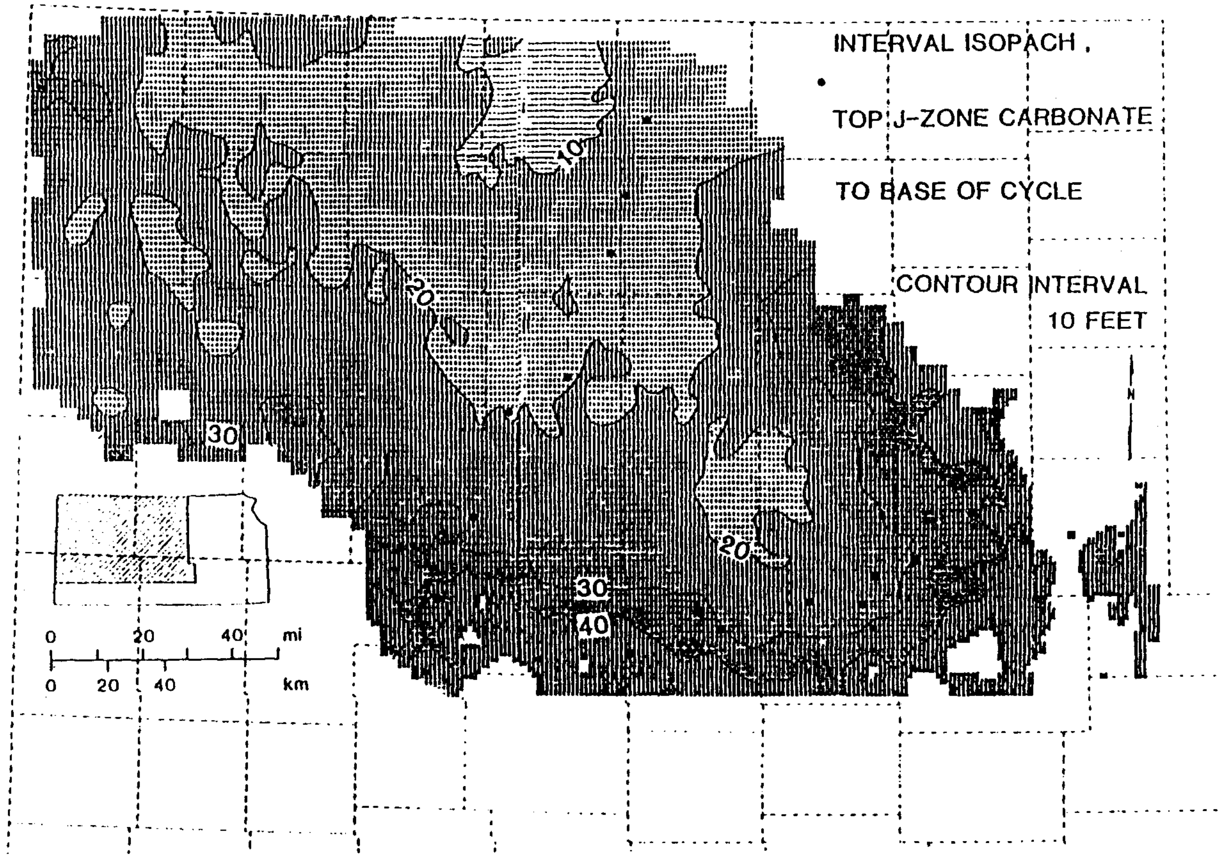
Areal variation of Mississippian "B" shale ratio in SW Stafford County. Oil and gas fields are shown in black

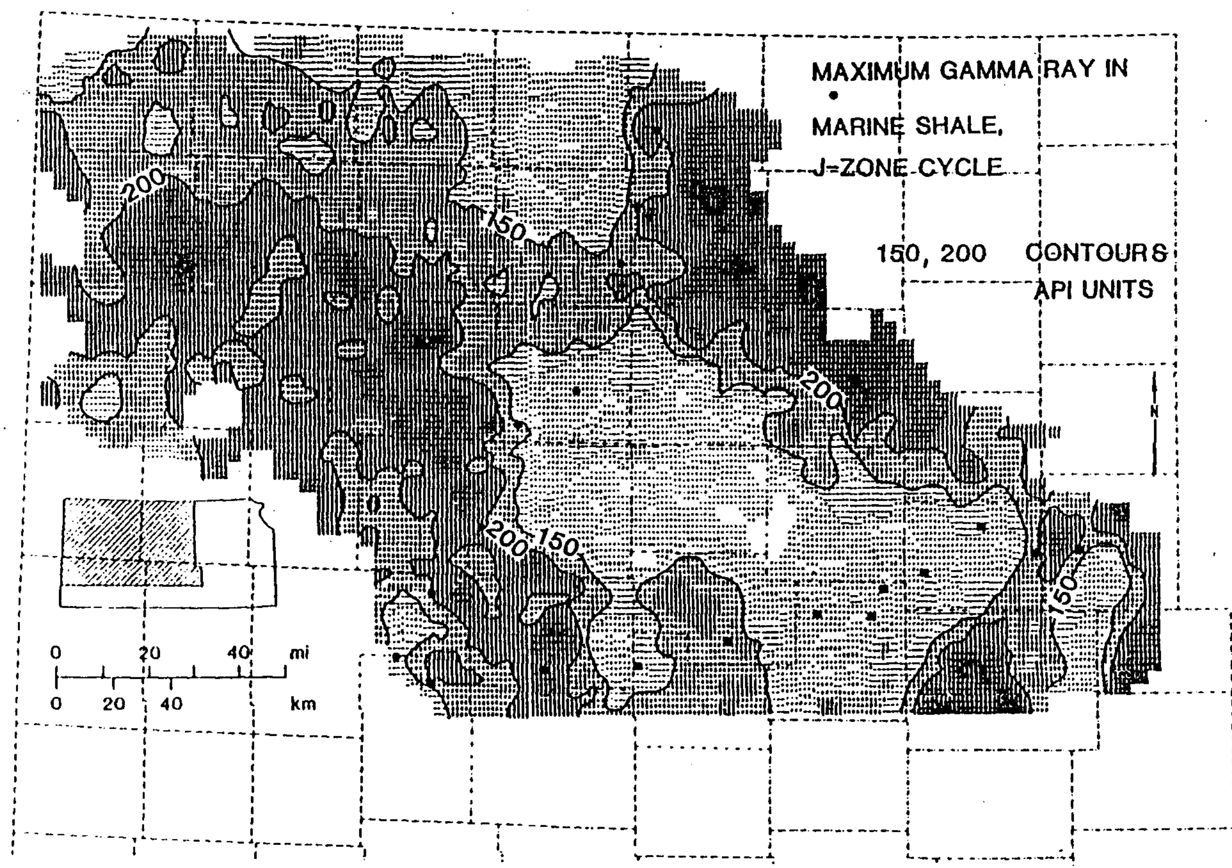
Reference : Harbaugh, Doveton, and Davis (1977)

REGIONAL MAPPING OF MAXIMUM GAMMA - RAY READING IN A MARINE SHALE AS A PALEOGEOGRAPHIC INDICATOR

Watney (1985) mapped maximum gamma - ray values of the transgressive marine shale of the Pennsylvanian J-zone cycle, based on several thousand wells from western Kansas. Comparison of this map with an isopach of part of the cycle illustrates a simple pattern which is readily interpreted in terms of regional facies variation and the paleogeography of the western Kansas shelf in Pennsylvanian times. The Central Kansas Uplift is demonstrated to be a major depositional "high", flanked by deeper marine facies with probable increased organic contents and reducing environments, which were responsible for higher uranium fixation and increased maximum gamma ray readings.

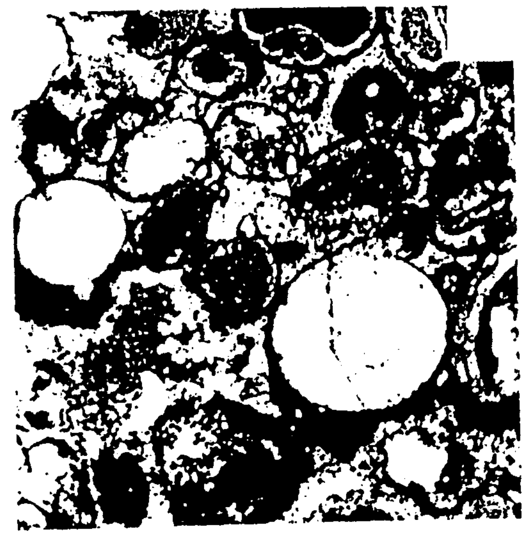



INTERVAL ISOPACH ,
TOP J-ZONE CARBONATE
TO BASE OF CYCLE
CONTOUR INTERVAL
10 FEET
10
20
30
20
30
40
0
20
40
mi
0
20
40
km
N

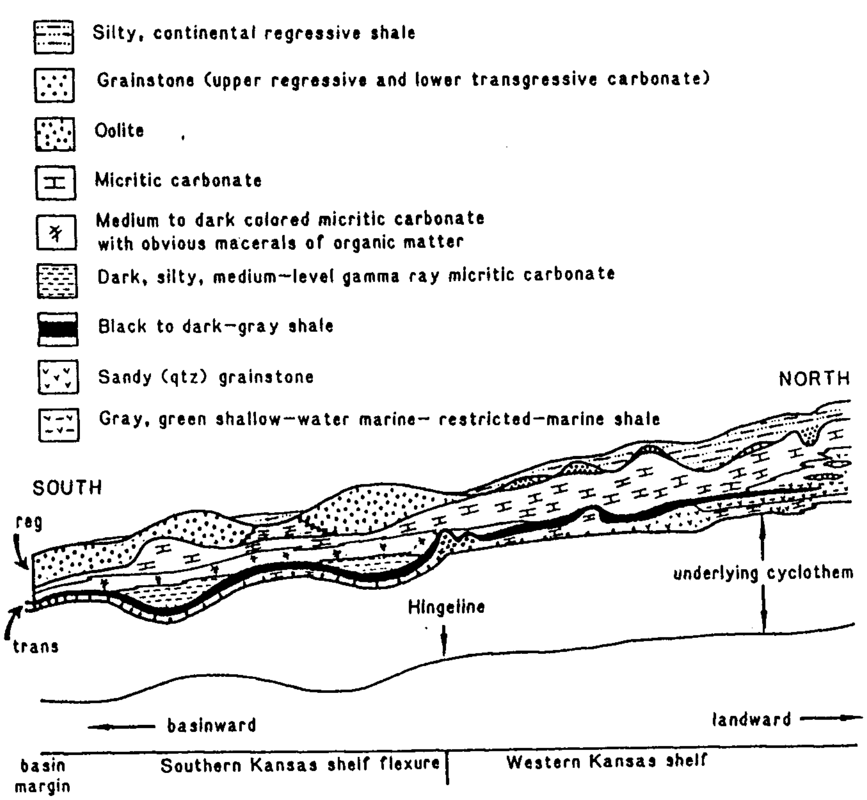

MAXIMUM GAMMA RAY IN
MARINE SHALE,
J-ZONE CYCLE
150, 200 CONTOURS
API UNITS
200
150
200
200
150
150
0
20
40
mi
0
20
40
km
N

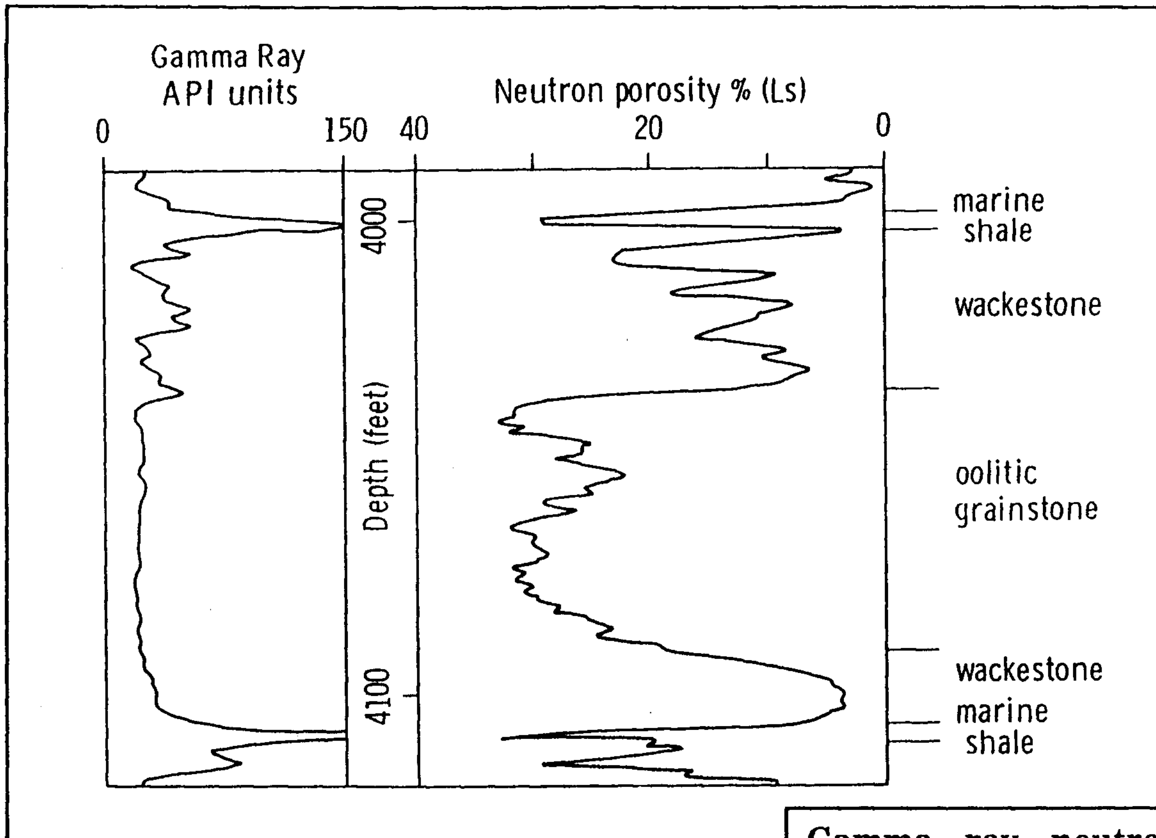
MAPPING PENNSYLVANIAN OOLITE SHOALS FROM ENHANCED POROSITY ASSOCIATED WITH THE DEVELOPMENT OF OOMOLDS

Oomoldic porosity zones occur locally in Kansas Pennsylvanian limestones, and the dissolution of the original ooids can result in substantial increases in porosity, unless filled with later cement. The oolitic zones are found typically at the top of regressive limestones. Watney (1985) concluded that the lower part was deposited as open- marine micrite sediments on a broad open shelf. The upper part represents a later stage of regression with localized development of oolite shoals. The recognition of the enhanced porosity motif allowed him to map the thickness of the oolite grainstone facies. The map shows northward -pointing fingers in a broad facies band which resemble recent spillover lobes of oolite in the Bahamas. Watney (ibid) suggested that the cause of the oolite distribution was a break in the Pennsylvanian paleoslope (striking approximately west to northwest), which acted as a focus for waves and currents during shoaling conditions operating late in the deposition of the limestone.

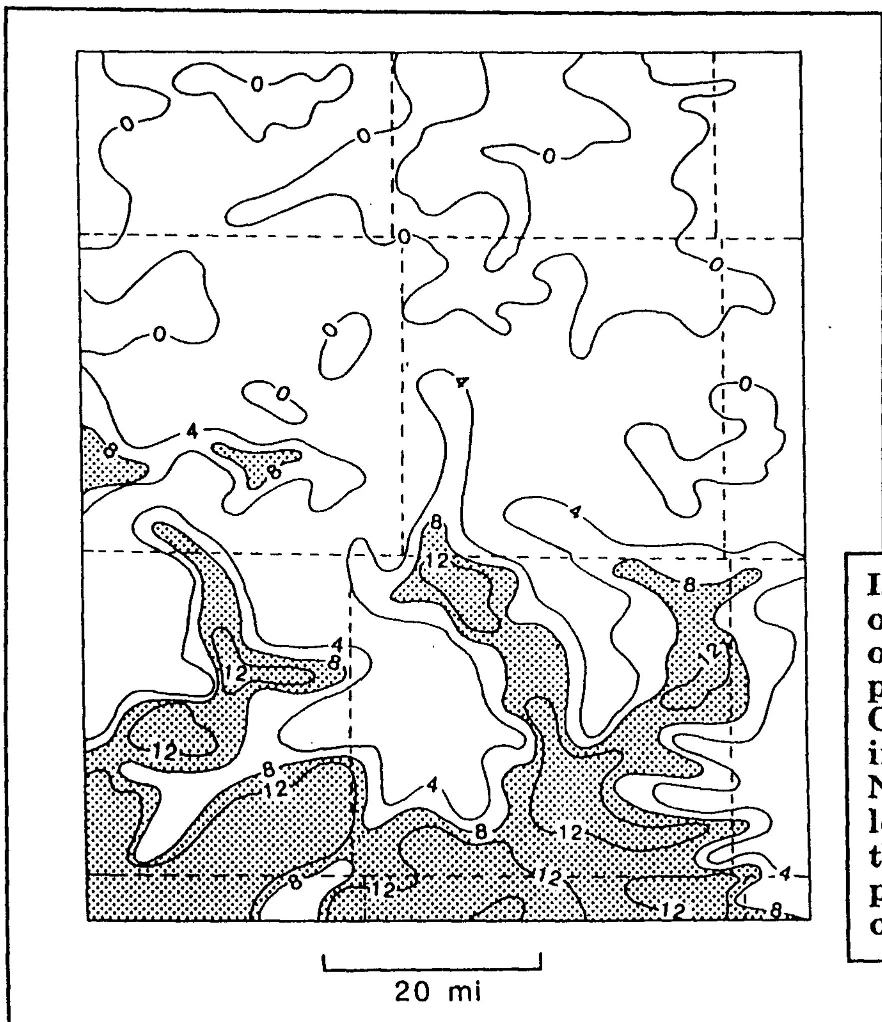


Spar-filled oomolds in Pennsylvanian oolitic grainstone





Gamma - ray, neutron logs of the J zone, Kansas City Group (Pennsylvanian) from a well in central Kansas, with typical high-porosity oolitic grainstone.

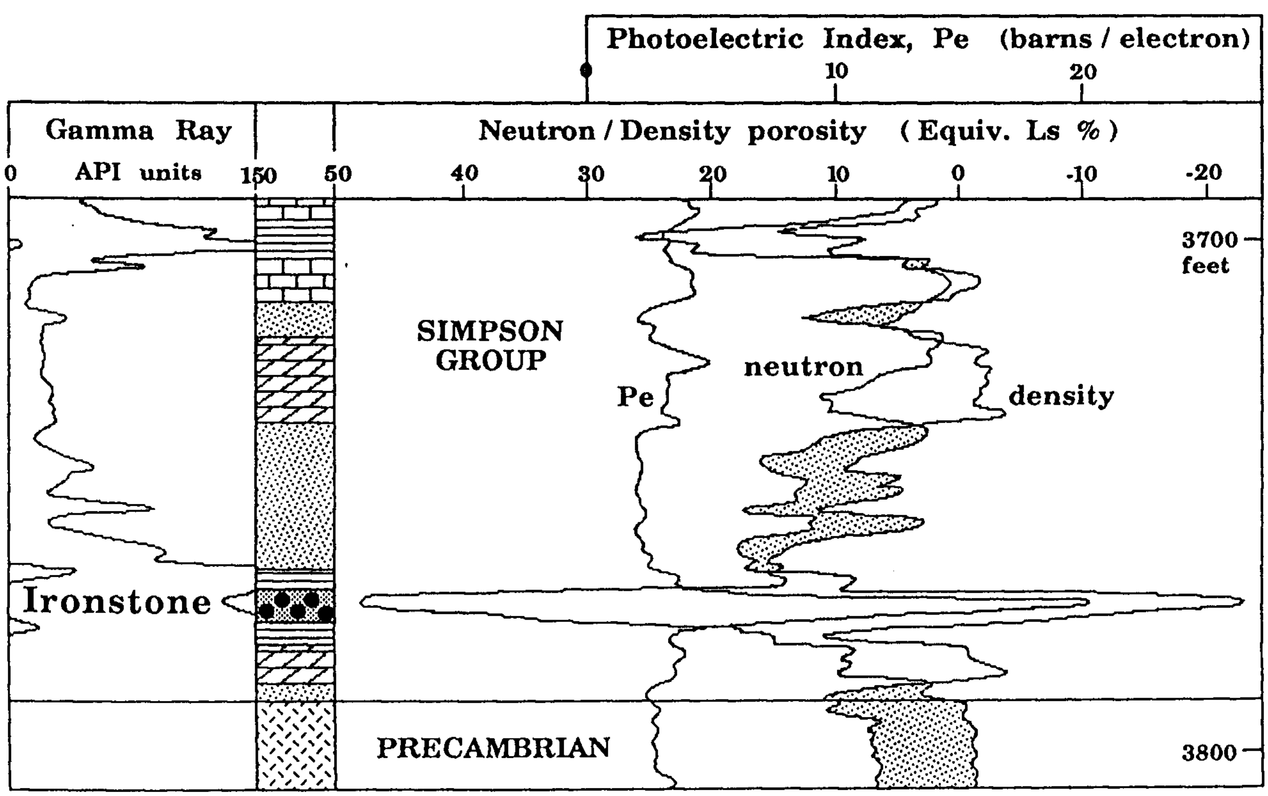


Isopach map (feet) of the high-porosity oolitic grainstone phase of the Kansas City Group J zone in central Kansas. Note the thick lobate features in the south which probably represent oolite shoals

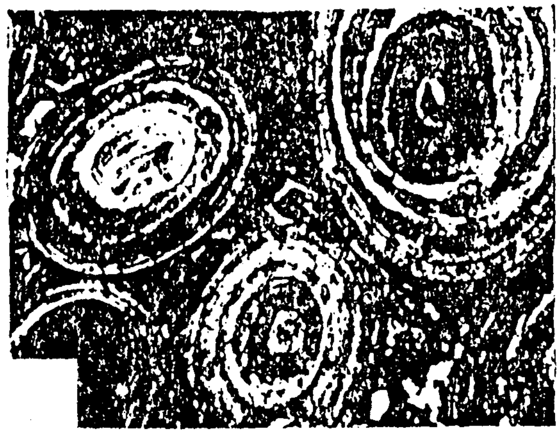
MAPPING THE DISTRIBUTION OF AN ORDOVICIAN OOLITIC IRONSTONE, USING THE DENSITY LOG

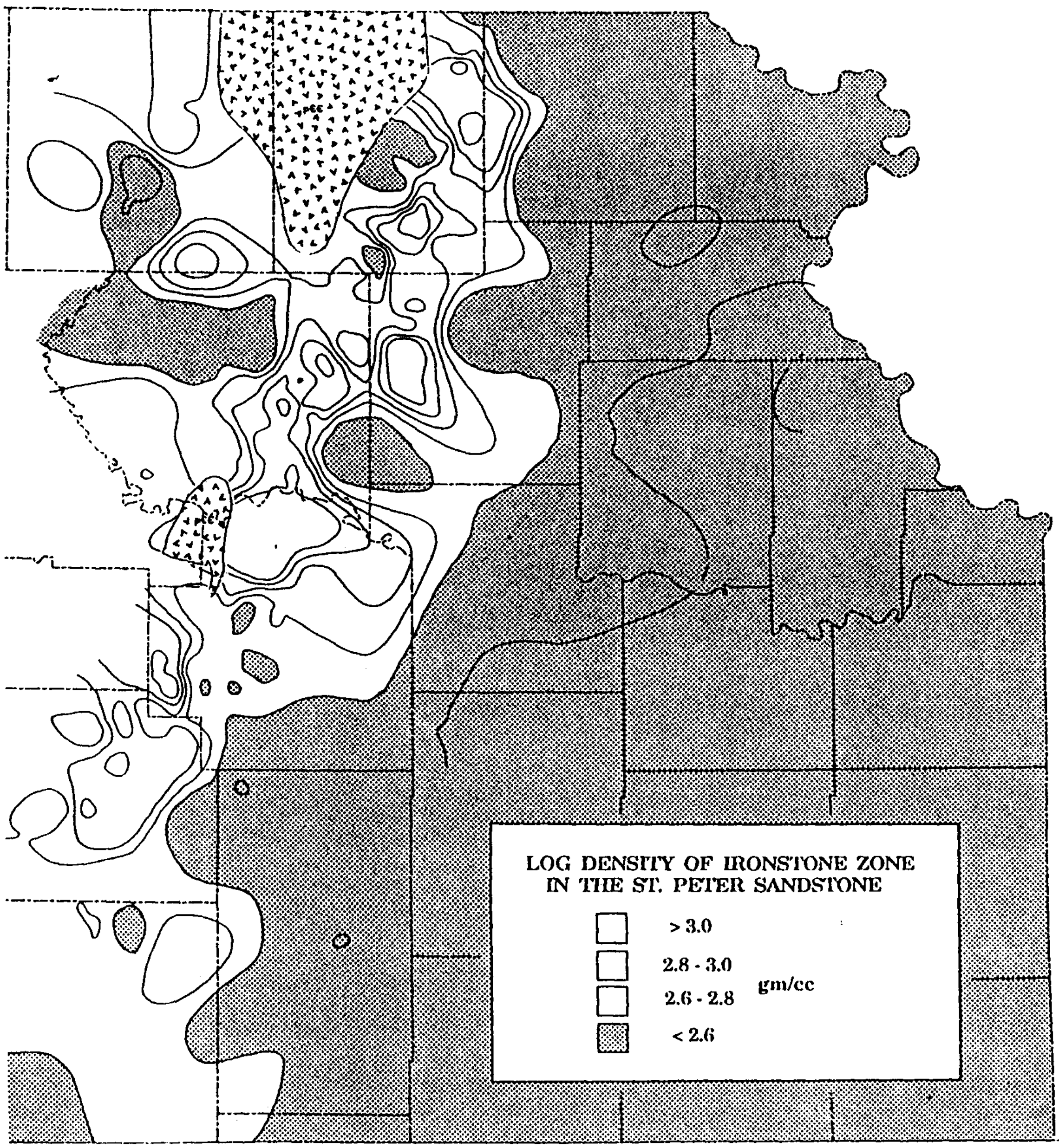
In northeast Kansas, an iron oolite zone is found locally in the shale from the lower part of the St. Peter Sandstone (Middle Ordovician) and consists primarily of goethite. The zone ranges from 5 to 10 feet in thickness and is often missed in the drill - cuttings, although the associated hematite turns the mud -pit red. Goethite has an unusual log signature of a very high neutron response (over 60 porosity units) caused by its hydroxyl content, and a density of 4.34 gm/cc. Consequently, neutron and density logs are particularly useful to both locate occurrences of the ironstone zone and map lateral gradations in its composition. Mapping revealed anomalies in close proximity to structurally high features on the Nemaha Anticline where the St. Peter is absent. These absences can therefore be attributed to non-deposition rather than erosional removal. The highs probably represent an archipelago of Middle Ordovician islands of Precambrian granite which is locally rich in magnetite. Intense weathering processes on the islands concentrated the iron for goethite formation in the peripheral nearshore environments.

(Berendsen and Doveton, 1988)



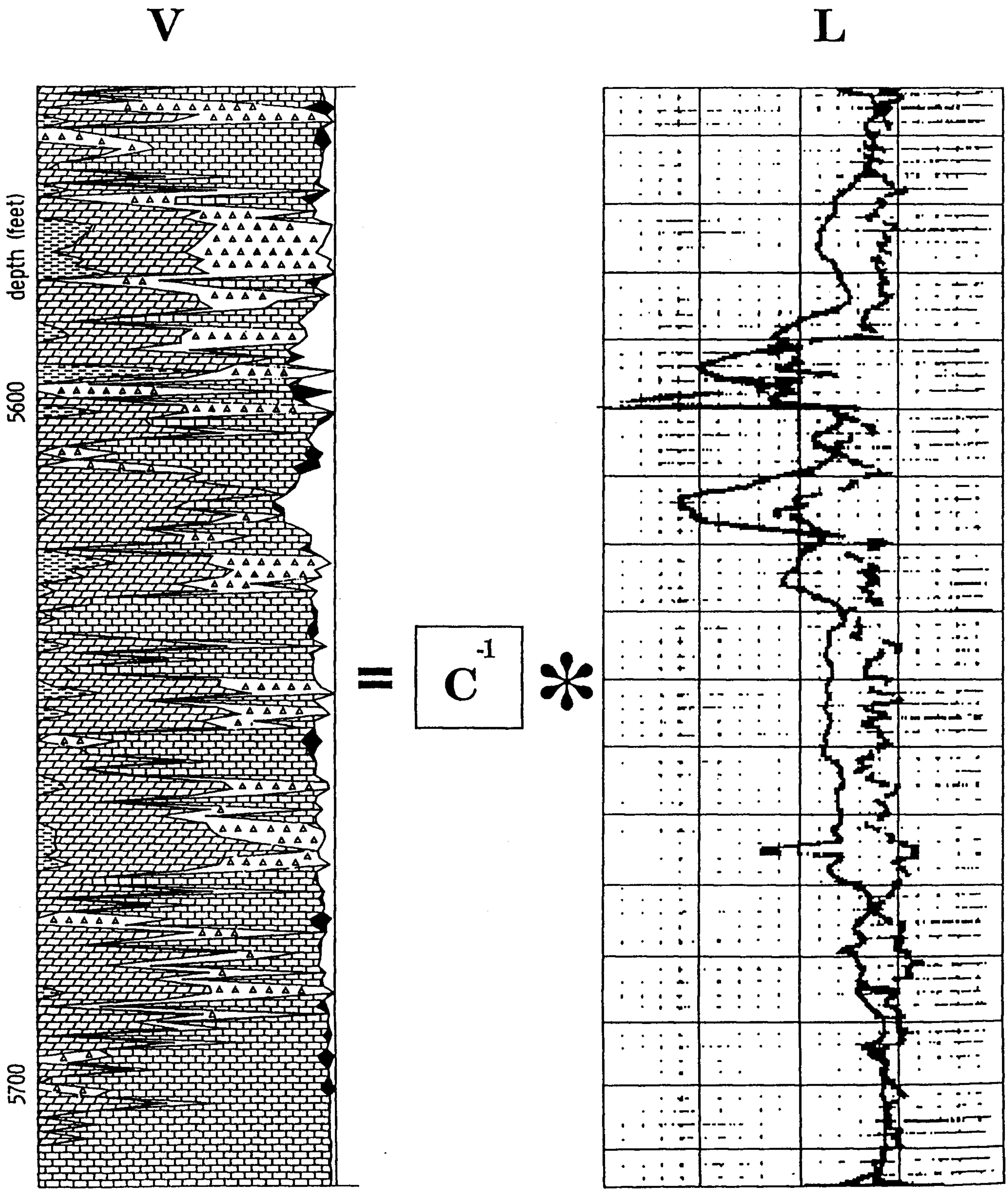
Thin - section of goethite oolites in the ironstone zone



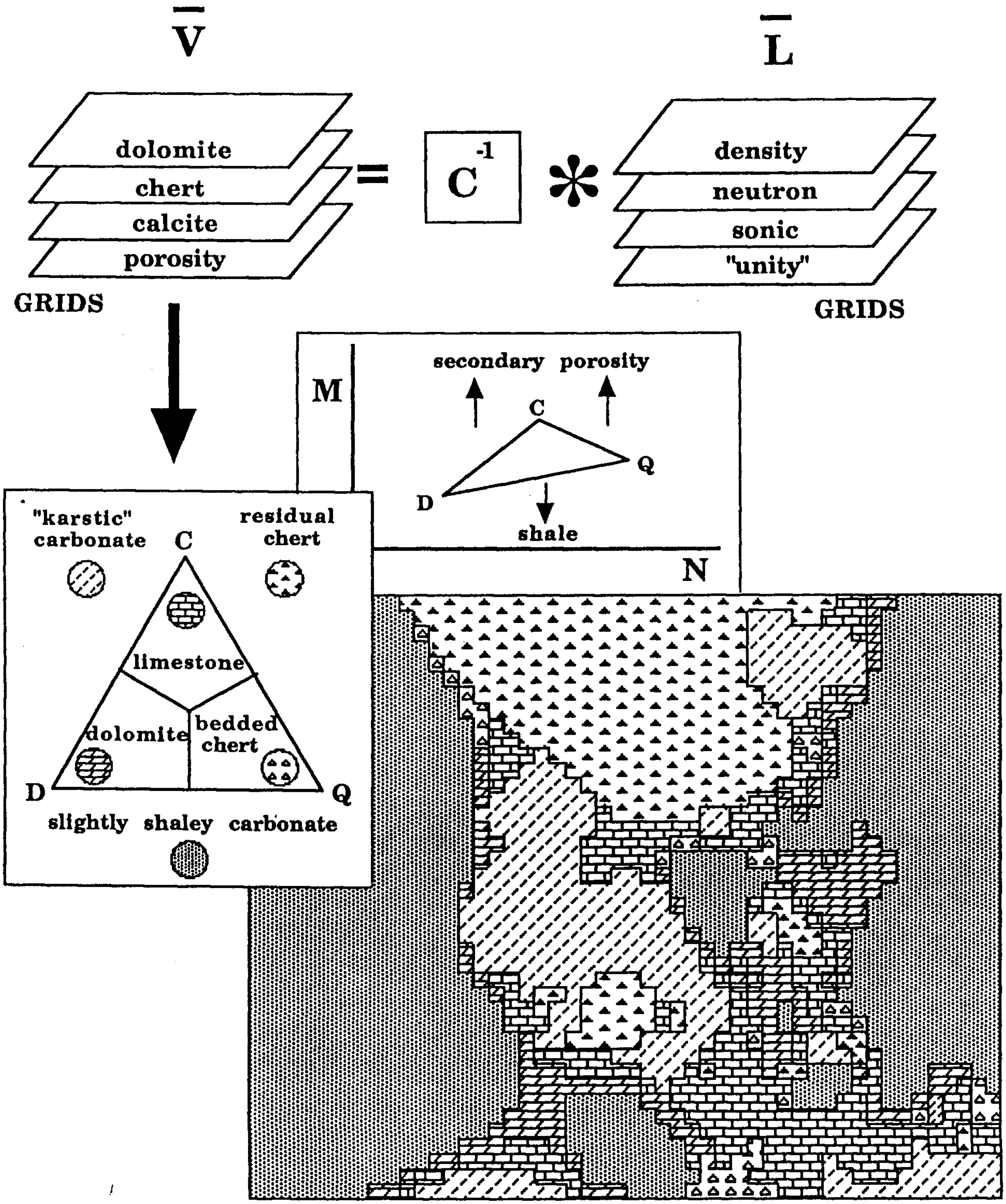


MATRIX ALGEBRA SOLUTION APPLIED TO VIOLA LIMESTONE SECTION MINERAL AND FLUID COMPOSITIONS

$$\boxed{CV = L \quad \therefore \quad V = C^{-1}L}$$



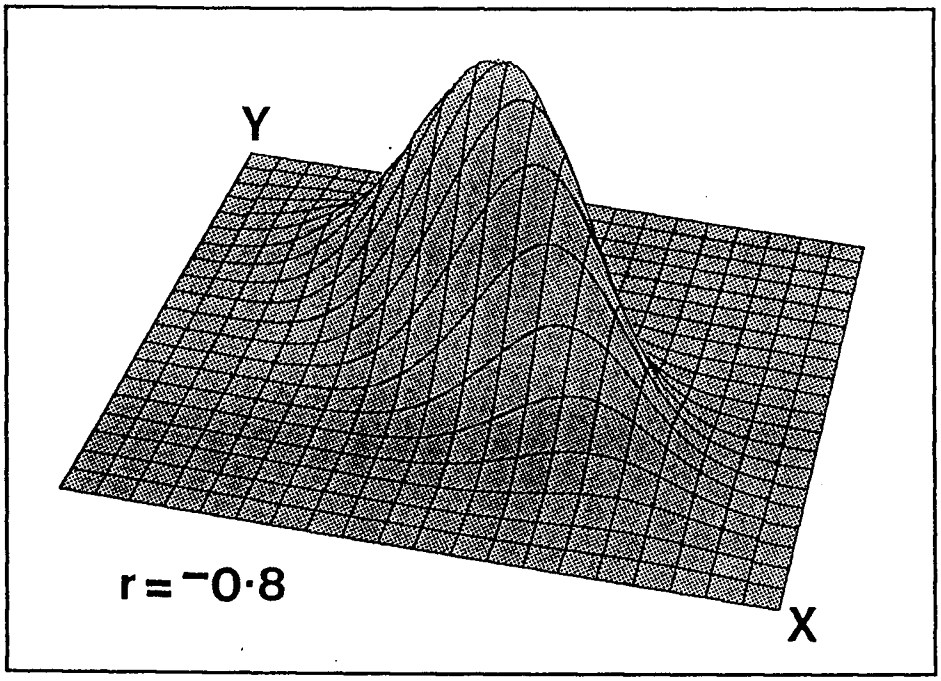
MATRIX ALGEBRA SOLUTION APPLIED TO VIOLA LIMESTONE LITHOFACIES MAPPING BY GRID-TO-GRID OPERATIONS



Reference: Bornemann and Doveton (1983)

Blank

STATISTICS



FUNDAMENTALS OF PROBABILITY

0.0 ← probability range → 1.0

absolute impossibilty | absolute certainty

RULES OF PROBABILITY

<u>Mutually Exclusive Events</u> cannot occur simultaneously.

<u>Additive Rule of Probability</u> :
If events A and B are mutually exclusive,

$$P(A \text{ or } B) = P(A) + P(B)$$

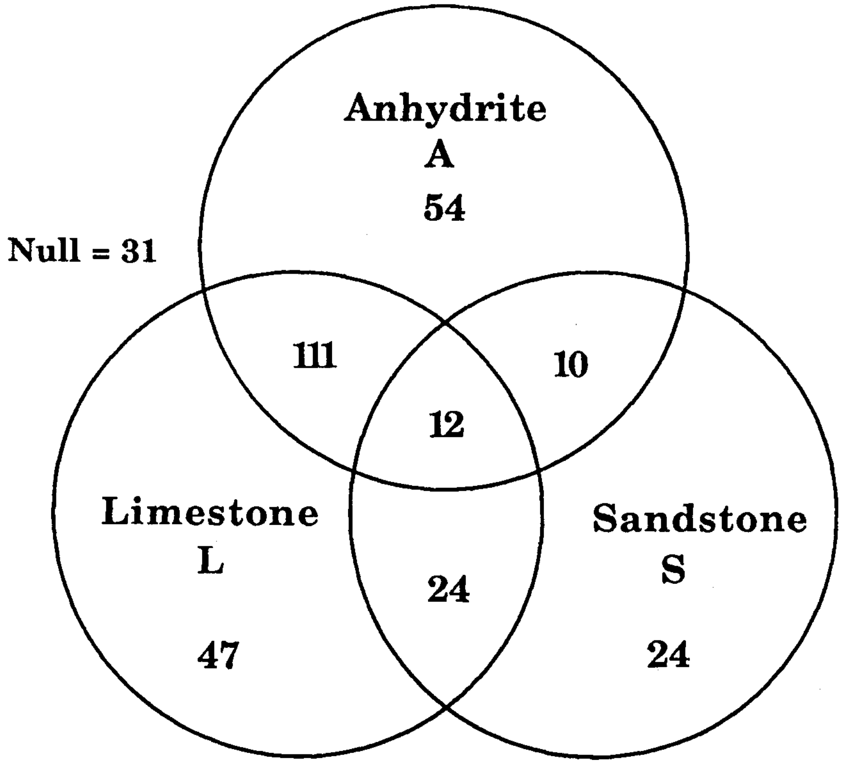
<u>Multiplicative Rule of Probability</u> :
If events A and B are independent, their <u>Joint Probabilty</u> of occurrence is :

$$P(A,B) = P(A) \times P(B)$$

If the joint probability is not the product of the individual probabilities, the events are <u>Conditional.</u> The probability of one event depends upon the occurrence of the other :

$$P(A,B) \neq P(A) \times P(B)$$

OCCURRENCE OF SANDSTONE, LIMESTONE AND ANHYDRITE IN BOREHOLE SECTIONS OF THE MORRISON FORMATION (JURASSIC) OF KANSAS



S L A	O	E
0 0 0	31	37.2
0 0 1	54	55.2
0 1 0	47	60.6
0 1 1	111	90.0
1 0 0	24	10.7
1 0 1	10	15.9
1 1 0	24	17.5
1 1 1	12	25.9

O = observed frequencies

Total number of "trials" = 313

$p_S = 70/313 = 0.224$ $\quad p_{\bar{S}} = 0.776$

$p_L = 194/313 = 0.620$ $\quad p_{\bar{L}} = 0.380$

$p_A = 187/313 = 0.597$ $\quad p_{\bar{A}} = 0.403$

The expected frequencies (E) of the joint occurrences (and non- occurrences) of sandstone, limestone and anhydrite, if they are independent, may be calculated by applying the multiplicative rule of probability to the marginal probabilities above.
e.g. probability of sandstone and limestone, but no anhydrite, assuming these are independent, is:

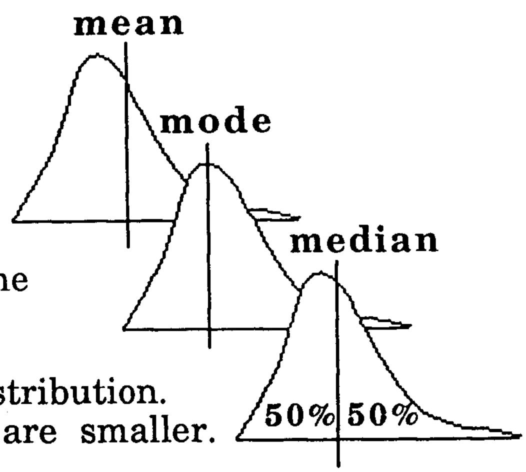
$$p_{SL\bar{A}} = p_S \times p_L \times p_{\bar{A}} = 0.056$$

The expected number (E) is then:

$$p_{SL\bar{A}} \times 313 = 17.5$$

MOMENTS OF A DISTRIBUTION

DESCRIPTORS OF LOCATION



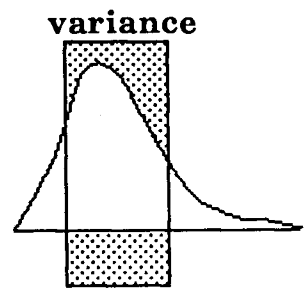
MEAN: The arithmetic average of the observations in a sample, $\bar{X}$. The "expected value" E(X) is the population parameter, μ

$$\bar{X} = \sum X_i / n$$

MODE : The value that occurs with the most frequency.

MEDIAN: The 50th percentile of a distribution. Half the observations are larger; half are smaller.

DESCRIPTORS OF DISPERSION



The parameter of VARIANCE is the expected squared difference of the observations from the population mean :

$$\sigma^2 = \frac{\sum (X_i - \mu)^2}{n}$$

The SAMPLE VARIANCE is the average squared difference of the observations in a sample from the sample mean :

$$s^2 = \frac{\sum X_i^2 - n\bar{X}^2}{(n-1)}$$

This equation can be rearranged as the computationally simpler :

$$s^2 = \frac{n \sum X_i^2 - (\sum X_i)^2}{n(n-1)}$$

The STANDARD DEVIATION, s, is equal to the square root of the variance.

The UNCORRECTED SUM OF SQUARES is $\sum X_i^2$ and the CORRECTED SUM OF SQUARES, SS $= \sum (X_i - \bar{X})^2$

HIGHER ORDER MOMENTS

The kth moment of a distribution is given by :

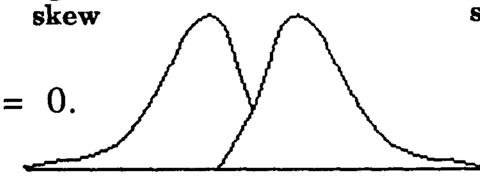
$$m_k = \frac{1}{n}\sum(X_i - \bar{X})^k$$

The first moment about the origin is the MEAN and represents the center of gravity of the distribution.
The second moment about the mean is the VARIANCE. The standard deviation corresponds to the radius of gyration, if the distribution was rotated about the center of gravity.

SKEWNESS (Third moment)

A measure of symmetry :

$$m_3 = \frac{1}{n}\sum(X_i - \bar{X})^3$$



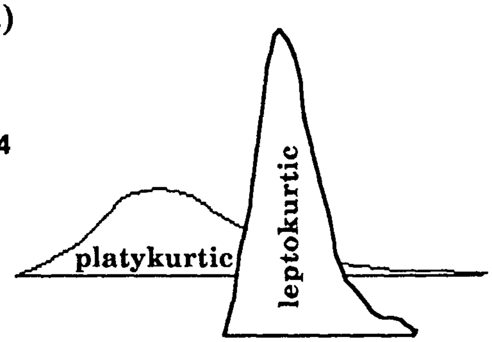
For a symmetrical distribution, $m_3 = 0$.
If $m_3 > 0$, the distribution has a POSITIVE SKEW ; if $m_3 < 0$, the distribution has a NEGATIVE SKEW.
The size of m_3 is influenced by the units of measurement. For comparison between variables, a dimensionless measure of skewness is computed by the ratio :

$$Sk = m_3 / \sqrt{m_2^3}$$

KURTOSIS (Fourth moment)

A measure of "peakedness" :

$$m_4 = \frac{1}{n}\sum(X_i - \bar{X})^4$$



The effect of measurement unit is eliminated by the dimensionless ratio :

$$Kt = m_4 / m_2^2$$

Distributions with higher kurtosis are more "peaked" and are termed LEPTOKURTIC ; with lower kurtosis, are "flatter" and termed PLATYKURTIC. The normal distribution is often used as a reference standard and has a kurtosis value of 3.

BINOMIAL DISTRIBUTION

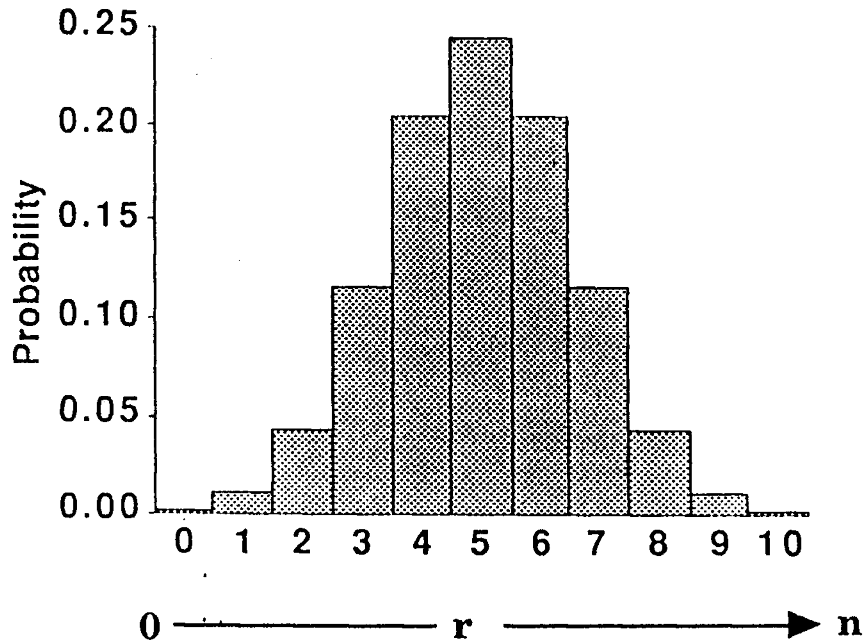
Discrete probability distribution with two possible outcomes ("success" or "failure") in a series of independent trials.

Probability of success = p

Probability of failure = q = (1-p)

The binomial distribution enumerates all the possible outcomes of r successes (ranging from 0 to n) in n trials:

$$P(r) = \frac{n!}{(n-r)!\,r!} q^{n-r} p^{r}$$



Sample mean, $\overline{X}$ = np
Sample variance, s^2 = npq

The proportion of "successes", p, in a sample of n trials is an estimate of the population parameter, π . The standard deviation of estimates, p, about π is the STANDARD ERROR and may be estimated as :

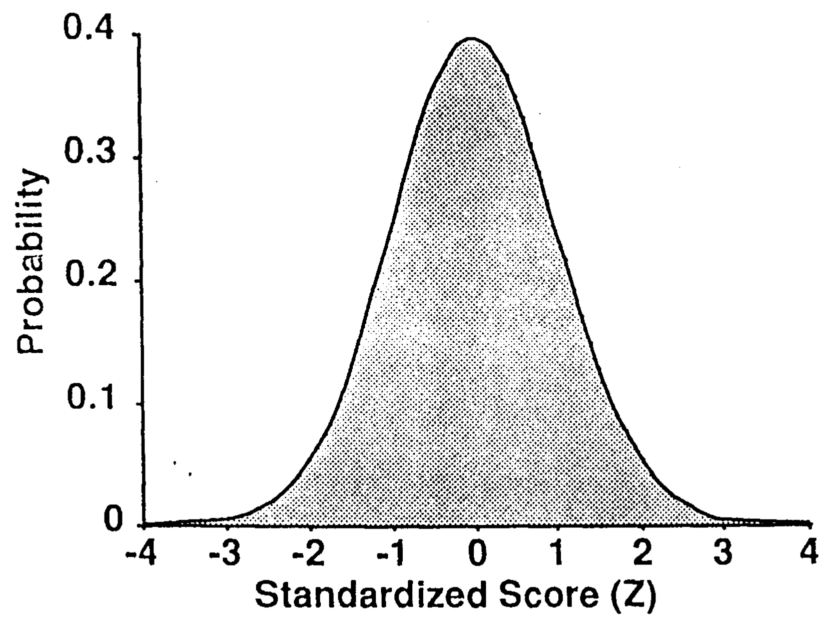
$$s_e = \sqrt{\frac{pq}{n}}$$

NORMAL DISTRIBUTION

A continuous distribution, which is the limit of the binomial, when the number of "trials" is extended to infinity. Also known as the "Gaussian curve" from the work of Gauss, who realized that it described the distribution of measurement error.

$$P(X) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$$

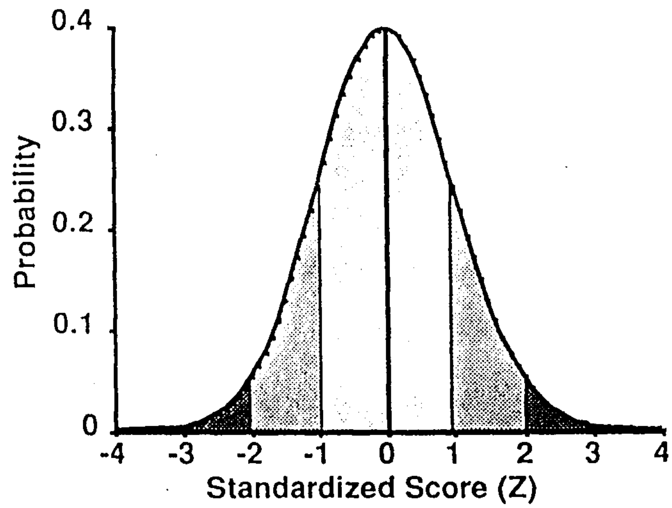
Notice that the normal distribution is described completely by the MEAN and VARIANCE.



In statistical work, the STANDARD NORMAL DISTRIBUTION is used, which has a mean of zero and a standard deviation of one. Observational data may be STANDARDIZED for comparison between variables with differing measurement units by the Z- transformation:

$$Z = (X - \overline{X}) / s$$

USE OF THE NORMAL DISTRIBUTION IN STATISTICAL INFERENCE



The probabilities of occurrence of different values from a normal distribution are given by integrated areas under the curve. 68% will be within +/- one standard deviation of the mean ; 95% within two.

THE CENTRAL LIMITS THEOREM

If sets of random samples are drawn from ANY population, their means will tend to be normally distributed.

The mean of the sample means is equal to the population mean :

$$\overline{(X)} = \mu$$

The variance of the sample means is equal to the variance of the population divided by the size of the samples :

$$s_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

The STANDARD ERROR (of the estimate of the mean) is the standard deviation of the sample means from the population mean, and is :

$$s_e = \sqrt{\frac{\sigma^2}{n}}$$

The Central Limits Theorem allows us to formulate statistical tests based on the normal distribution, even when the sampled population is not itself normally distributed.

The Z- test can be used to check the probability that a sample belongs to a given population. First a Z- value is calculated :

$$Z = \frac{\bar{X} - \mu}{s_e}$$

Z is a measure of the distance of the sample mean from the population mean in standard error units. If, say, Z = 2 and we decide that the sample does not belong to the population, then we run the risk of 5% that we are wrong.

STATISTICAL HYPOTHESIS TESTS

A process of INDUCTIVE LOGIC -- making generalizations from a limited sample to an entire population -- the validity of the conclusions will be probabilistic.

PROCEDURE :
State a NULL HYPOTHESIS of NO DIFFERENCE :

e.g. $H_0: \mu_1 = \mu_2$

State an ALTERNATIVE HYPOTHESIS that covers all other contingencies,

$$H_1: \mu_1 \neq \mu_2$$

Set a test criterion for the rejection of the null hypothesis. This is based on probability and chosen to reflect the cost of being wrong.
Calculate a test statistic and relate this to the expected value at the criterion probability.
On the basis of the comparison, either accept or reject the null hypothesis.

SIGNIFICANCE LEVELS
The acceptable risk of committing a Type I error -- the probability that the null hypothesis is indeed true even when "rejected" by the statistical test. The significance level is denoted by α and is selected BEFORE the test. In most geological applications a value of 0.05 is customarily used, in common with most scientific applications. This can be argued as rather strict for geology, since we record the results of ancient, uncontrolled "experiments", unlike most of the other sciences. Type II error -- the probability of accepting an hypothesis when it is false -- is generally unknown, so that the null hypothesis is usually set as the one we are attempting to reject. By using a conservatively low level of α the chance of a Type II error is minimized.

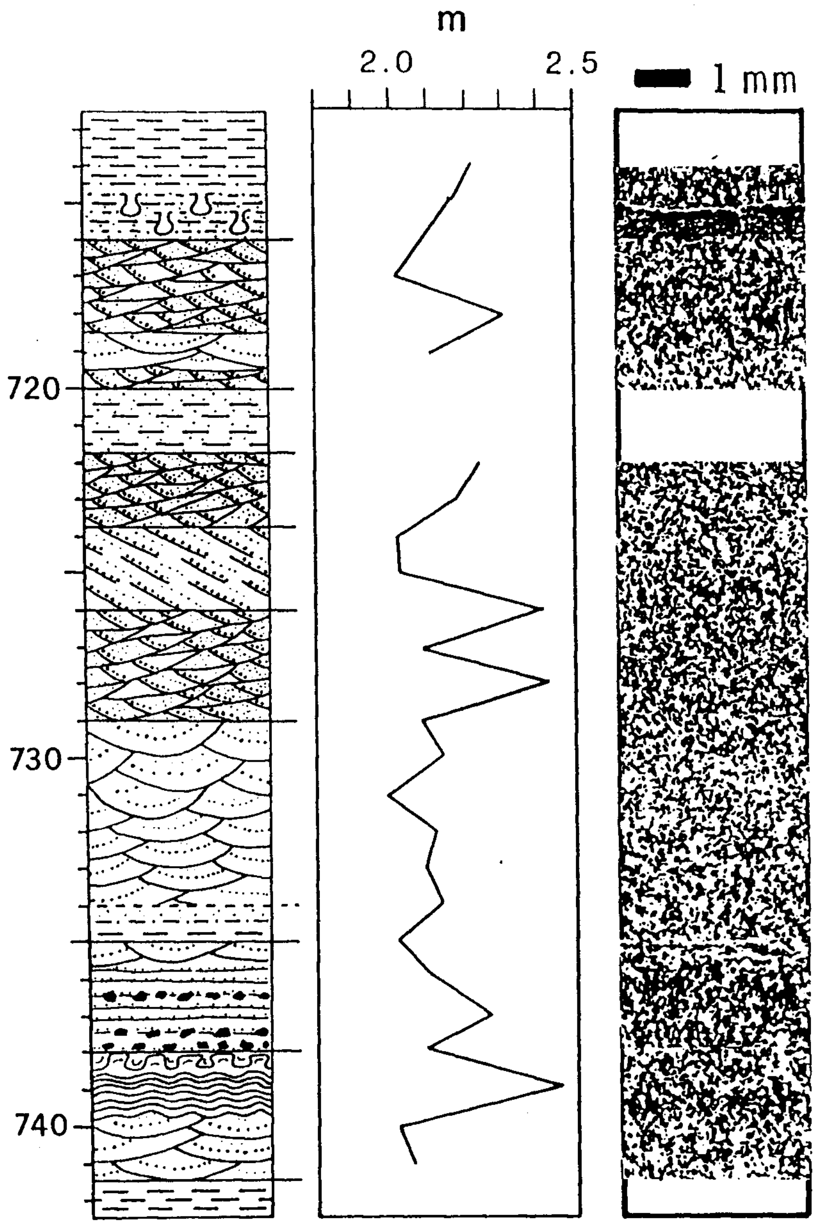
DEGREES OF FREEDOM
The number of independent items of information in a sample. Usually, the number of observations minus the number of parameters estimated.

COMMON TESTS

t- test : EQUALITY OF MEANS : $H_0: \mu_1 = \mu_2$ $\quad t = \dfrac{X_1 - X_2}{s_e}$

F- test : EQUALITY OF VARIANCES : $H_0: \sigma_1^2 = \sigma_2^2$ $\quad F = \dfrac{s_1^2}{s_2^2}$

EXAMPLE : IS THERE A SIGNIFICANT DIFFERENCE BETWEEN CEMENTATION FACTORS (M) CALCULATED FROM CORE MEASUREMENTS OF RIPPLED ZONES AND CROSS-BEDDED UNITS IN A SECTION OF THE SKINNER SANDSTONE (PENNSYLVANIAN) ?



RIPPLED -ZONES (m1)	CROSS-BEDS (m2)
2.09	2.01
2.02	2.02
2.30	2.08
2.10	2.13
2.23	1.99
2.17	2.11
2.42	2.09
2.09	2.01
2.42	2.01
	2.05
$n_1 = 9$	$n_2 = 10$
$\overline{m}_1 = 2.20$	$\overline{m}_2 = 2.05$
$s_1 = 0.148$	$s_2 = 0.049$

F-test

$$H_0: \sigma^2_{m1} = \sigma^2_{m2}$$

$$F = s^2_{m1} / s^2_{m2} = 0.148^2 / 0.049^2 = 9.12$$

$$\nu_1 = n_1 - 1 = 8 \qquad \nu_2 = n_2 - 1 = 9$$

Critical F value @ α 0.05 and 8 & 9 df = 3.23

The critical value is exceeded by the calculated F-value, and so the null hypothesis is rejected. The alternative hypothesis is accepted : $H_1: \sigma^2_{m1} \neq \sigma^2_{m2}$

This means that the cementation factor is considered to be significantly more variable in the rippled zones than in the cross-bedded units.

t -test

Because there is a significant difference between the variances of the two groups, a MODIFIED t -test is calculated. The best estimate of the standard error of the difference of the means is then :

$$s_e = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

The approximate number of degrees of freedom is given by the formula :

$$\nu = \frac{(s_1^2 / n_1 + s_2^2 / n_2)^2}{(s_1^2 / n_1)^2 / (n_1 - 1) + (s_2^2 / n_2)^2 / (n_2 - 1)}$$

$$\nu \approx 9.6$$

Then, $$t = \frac{m_2 - m_1}{s_e} = 2.90$$

Critical t value @ α 0.05 and 9 df = 1.833

The critical value is exceeded and so the null hypothesis rejected. The alternative hypothesis is accepted : $H_1: \mu(m_1) \neq \mu(m_2)$

The cementation factor is considered to be significantly higher in the rippled zones than in the cross-bedded units.

CHI-SQUARE DISTRIBUTION AND TEST

For a large number of "trials" with a binomial result, the discrete binomial distribution is often approximated by the continuous normal distribution. In a similar fashion, the CHI-SQUARE DISTRIBUTION is used as a continuous approximation of the MULTINOMIAL DISTRIBUTION, where an observed event may take one of several outcomes.

The distribution is used for the CHI-SQUARE TEST, where if measurements fall into m different and mutually exclusive classes, then :

$$\chi^2 = \sum_1^m \frac{(O - E)^2}{E}$$

where O = the observed frequency in each class, and E = the expected frequency (as predicted from a null hypothesis distribution).

It is most commonly used as a TEST OF ASSOCIATION between properties which are categorized. The number of degrees of freedom are the number of classes minus the number of independent values used in estimating the expected frequencies. The chi-square test is the best known example of a NON-PARAMETRIC TEST. Most hypothesis tests (such as the t- and F tests) are PARAMETRIC , because they work with the estimation of parameters such as means and variances. However, non-parametric tests are necessary for analysis of nominal and ordinal data (categorical scales of measurement).

EXAMPLE : TEST OF ASSOCIATION BETWEEN SANDSTONE (S), LIMESTONE (L), AND ANHYDRITE (A) IN THE MORRISON FORMATION (JURASSIC)

S L A	O	E
0 0 0	31	37.2
0 0 1	54	55.2
0 1 0	47	60.6
0 1 1	111	90.0
1 0 0	24	10.7
1 0 1	10	15.9
1 1 0	24	17.5
1 1 1	12	25.9

The expected frequencies are those calculated as the product of marginal probabilities i.e. no association. The null hypothesis is that there is no significant difference between observed and expected, and therefore no reason to believe an association between the components.

The calculated $\chi^2 = 37.6$

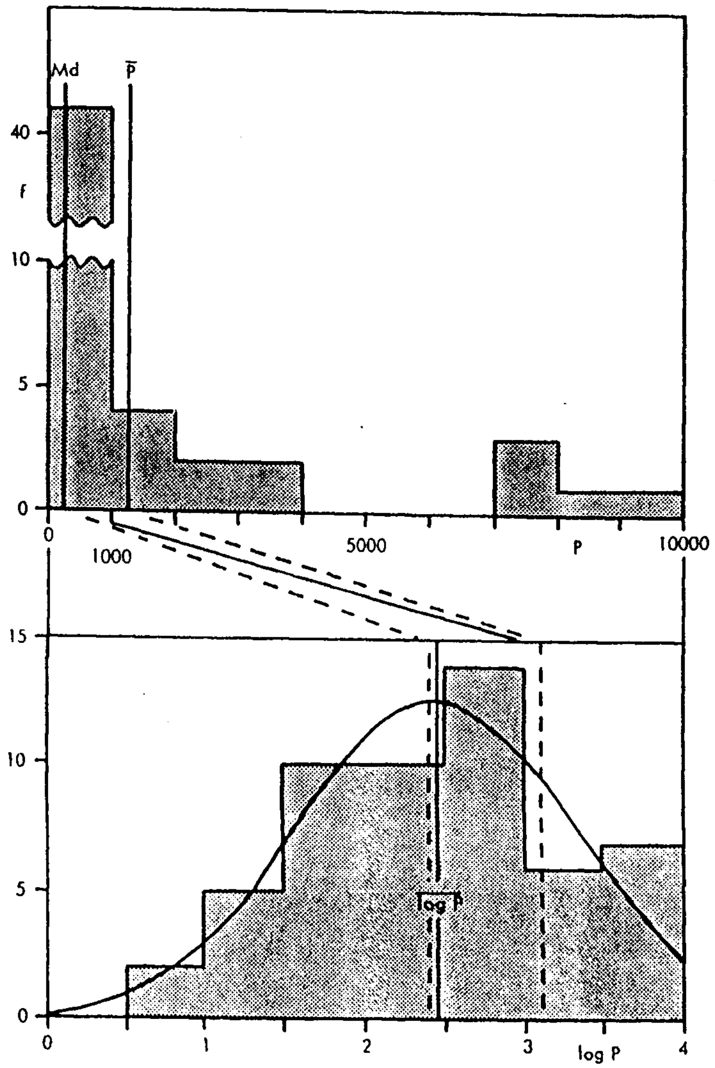
$\chi^2 @ \alpha 0.05 \cdot \text{and} \cdot 1 \text{df} = 3.84$

So the null hypothesis is rejected and it is accepted that there is significant association between the components.

LOGNORMAL DISTRIBUTION

The normal distribution is the appropriate model for the description of error, which can be thought of as the summation of small, random arithmetic displacements from a systematic value. At one time the normal distribution was also considered to be a "universal law" of variation of measurements in natural populations. Sample distributions which were markedly skewed were explained as mixtures of different normal populations. More recently, it was realized that some samples may the product of processes that are MULTIPLICATIVE in character. So, for example, the formation of particles is partly the result of crushing and breakage. The proportioning process is a division of the size, which generates a LOGNORMAL DISTRIBUTION. The values of the logarithms of the raw variable are normally distributed. The mean of a lognormal distribution is then equivalent to the geometric mean of the data.

EXAMPLE : DISTRIBUTION OF VOLUMES OF CHEROKEE OILFIELDS IN SOUTHEAST KANSAS



Arithmetic mean,
$\bar{P}$ = 1258 M bbls

Median,
Md(P) = 257 M bbls

Mean logarithm,
$\overline{\log P}$ = 2.45
which is equivalent to a Geometric mean of 283 M bbls

POISSON DISTRIBUTION

The Poisson distribution is a limiting case of the binomial distribution, when p, the probability of "success", becomes very small, and (1-p), the probability of "failure" approaches unity. Then :

$$P(X) = e^{-np}(np)^X / X!$$

where e = 2.718 and X is the number of "events" in an interval. The distribution describes the occurrence of rare, random "events", whose probability is p, in a sequence of n trials. There are many cases where the complementary occurrence of "non-events" is not observable, so that neither p nor n are known. However, the distribution only requires the product of n and p (np) to be known, and this is the mean number of events per interval. For the Poisson model :

$$\text{mean}(\bar{X}) = \text{variance}(s^2) = np$$

$s^2 < \bar{X}$ ... more uniform		$s^2 > \bar{X}$ ... more clustered

EXAMPLE : ANHYDRITE OCCURRENCE IN THE CHASE GROUP

Section : 2650 - 3100 feet depth = 450 feet

Subdivided (arbitrarily) into 18 sequential intervals

Number of anhydrite zones = 34

Mean number of anhydrites per interval ($\bar{X}$) = 34/18 = 1.89

Variance (s^2) = 2.81

Initial indication : more clustered than random

X	n_o	P(X)	n_e
0	6	0.151	2.72
1	1	0.286	5.14
2	5	0.270	4.86
3	2	0.170	3.06
4	3	0.080	1.44
5	1	0.030	0.55

X = number of anhydrite "events"

n_o = number of intervals observed with X events

P(X) = Poisson probability, using mean of 1.89 as estimate of np

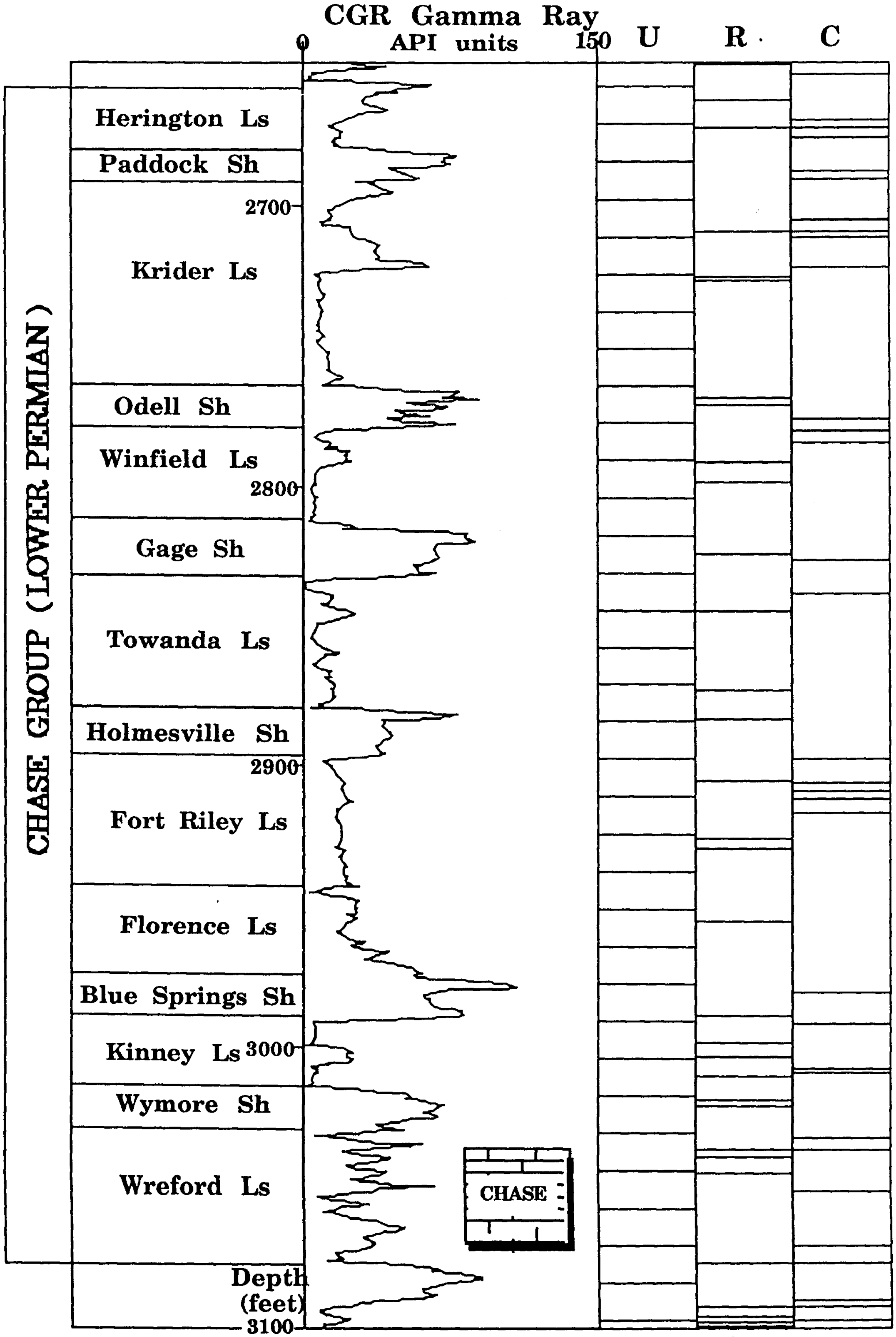
n_e = number of intervals expected with X events for Poisson model

Number of degrees of freedom (df) = 18 -2 = 16

Critical χ^2 value @ $\alpha = 0.05$ and 16 df = 26.3

Calculated χ^2 value = $\Sigma\,(O\text{-}E)^2 / E$ = 35.4

This exceeds the critical value, and so the null hypothesis of Poisson (random) event distribution is rejected.



U = uniform ; R = random (Poisson) ; C = clustered

COVARIANCE AND CORRELATION

COVARIANCE is the joint variation of two variables around their common mean :

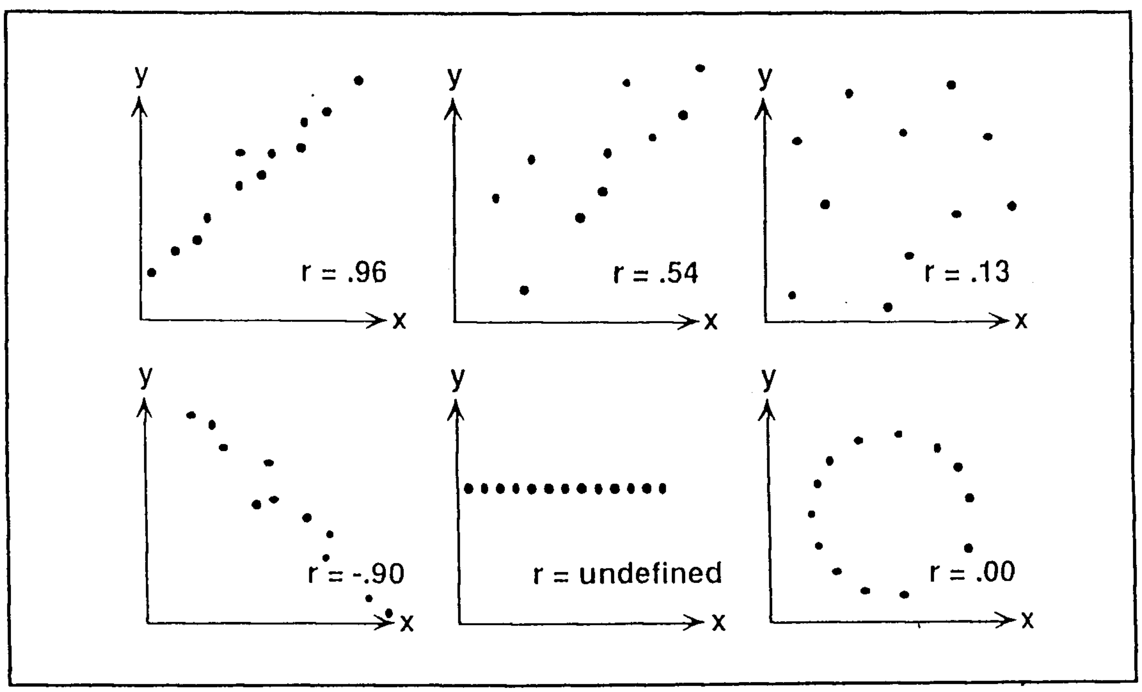
$$cov(X, Y) = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)}$$

$$= \frac{\sum X_i Y_i - (\sum X_i \sum Y_i / n)}{(n-1)}$$

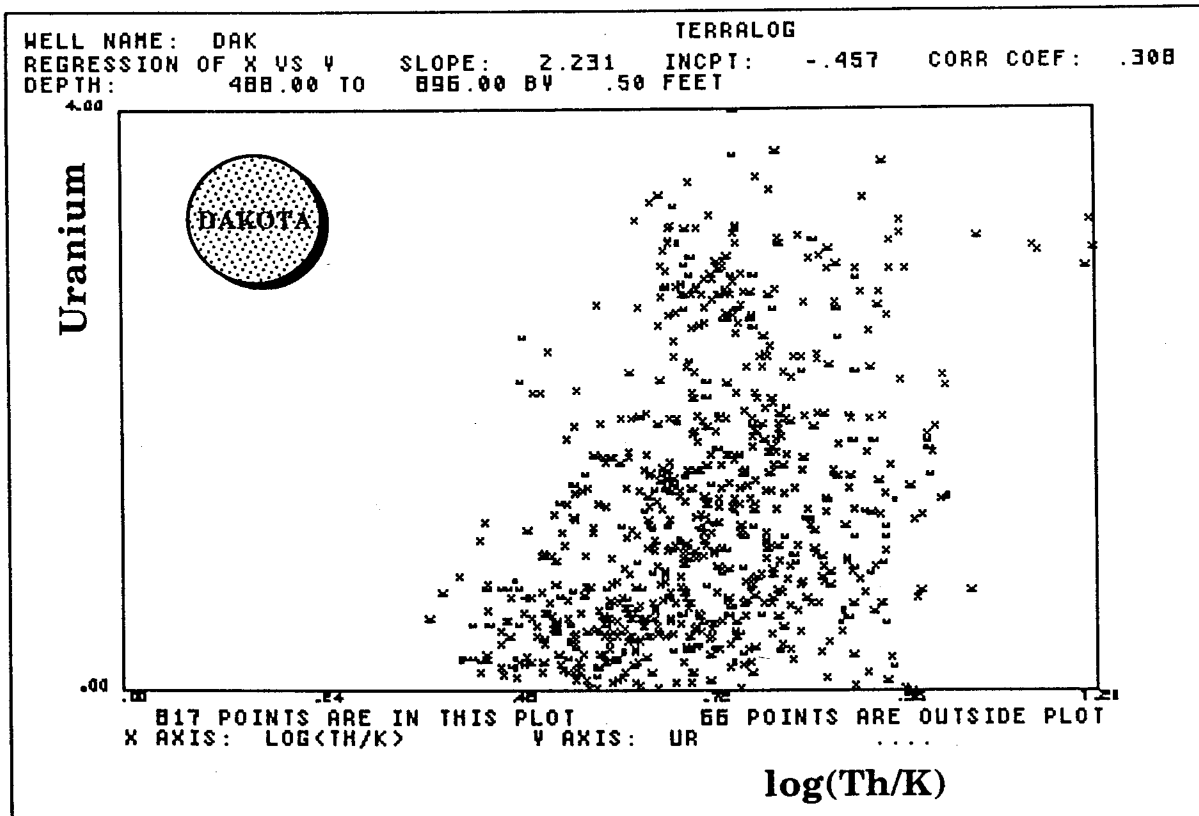
CORRELATION is a standardized measure of the linear relation between two variables. It is the covariance of the two variables when they have been standardized, and so is independent of measurement units :

$$r_{XY} = \frac{cov(X, Y)}{s_X s_Y}$$

This measure is the PEARSONIAN PRODUCT-MOMENT CORRELATION COEFFICIENT. Since the product of the standard deviations cannot exceed the covariance, the correlation coefficient is constrained between +1 and -1. A value of +1 is given by a perfect linear relationship : -1 corresponds to a perfect inverse relationship ; 0 means no linear relationship. The correlation parameter is denoted by ρ , the sample estimate is indicated by r.



EXAMPLE : IS THERE A CORRELATION BETWEEN URANIUM AND THE THORIUM/POTASSIUM RATIO IN THE DAKOTA/KIOWA/CHEYENNE ?



Because the thorium/potassium measure is a ratio, we attempt to linearize its variation and compute its correlation with uranium :

$$r = 0.308$$

There is an indication of a weak positive correlation, but is this significant ? Or is this value simply a sample estimate of a population correlation parameter of zero ? This question may be resolved by use of the fact that the sampling distribution of r about 0 when ρ is zero is approximately normal. Then a t- test can be used with the null hypothesis that the correlation parameter is zero. Then :

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

with (n-2) degrees of freedom. The calculated t- value = 17.9. Since the sample size is 817, there are 815 degrees of freedom and so :

$$t = 1.96 \cdot @0.05\alpha \cdot \text{and} \cdot 815\text{df}$$

The critical value is exceeded and the null hypothesis is rejected. It is accepted that there is some form of significant correlation between uranium and the Th/K ratio. The correlation sign suggests a relative enrichment in uranium with decreasing levels of potassium.

RANK CORRELATION

Since it is impossible to estimate parameters measured on an ordinal scale (ordered categories), some form of non-parametric measure of association must be used for this type of data. If measurements of two ordinal variables are ranked in order for each of the two variables, a SPEARMAN'S RANK CORRELATION COEFFICIENT, r' , may be calculated :

$$r' = 1 - \frac{6 \sum D_i^2}{n(n^2 - 1)}$$

where n = the number of individuals in the sample, and D_i is the difference between the ranks of the ith individual measured on the two variables. The statistic is based on the consequence that there are n! possible different combinations of the ranks of the two variables. The possible values are constrained between +1 and -1, with the same implications as the Pearson correlation coefficient.

The potential significance of the computed r' may be checked by a t- test of the hypothesis that the parameter ' is zero :

$$t = \frac{r' \sqrt{n-2}}{\sqrt{1 - r'^2}}$$

The rank correlation coefficient is also useful in cases where the variables are drastically different properties with radically different units. Since the Pearsonian correlation is a measure of dgree of LINEARITY between variables, a strong non-linear relationship may give a disappointingly weak coefficient. However, by using the Spearman measure for ranked data, the coefficient is sensitive to a MONOTONIC pattern in the data, regardless of whether the trend is non-linear.

A good example is provided by data tabulated in Winsauer and others (1952) classic paper in which they derived the Humble equation. By using their figures for formation factor and porosity the cementation factor, m, can be calculated for each of the sandstone samples. It is interesting to examine the possible associations with the other measured variables of tortuosity (by ion transit time), permeability, mean grain diameter, standard deviation grain size, packing, roundness, and skewness. The wide range of units and great potential for non-linear relationships is a sound recommendation to restrict initial analysis to the examination of trends of ordering. An example calculation of Spearman's rank correlation coefficient is given on the following page, together with a tabulation of the coefficients and their significance .

EXAMPLE : CALCULATION OF SPEARMAN'S RANK CORRELATION COEFFICIENT BETWEEN CEMENTATION FACTOR AND PERMEABILITY

m	R_m	k	R_k
1.78	15	90	11
2.05	26	7	2
1.62	2	220	16
1.61	1	1920	22
1.64	3	4400	26
1.77	13	145	14
1.86	18	25	6
1.75	10	410	19
2.05	25	3	1
1.96	24	9	3
1.70	6	200	15
1.77	16	36	7.5
1.80	17	70	9
1.69	5	330	17
1.77	12	98	12
1.71	7	1560	21
1.72	8	36	7.5
1.76	11	1180	20
1.88	20	3200	25
1.89	21	2100	23
1.86	19	18	4
1.67	4	2200	24
1.94	23	19	5
1.94	22	88	10
1.77	14	370	18
1.74	9	130	13

m = cementation factor
R_m = rank of m

k = permeability
R_k = rank of k

$\sum D_i^2 = 4793.5$

$r'_{mk} = -0.64$

Calculated t = 4.08

n = 26
df = n-2 = 24

t = 1.71 @0.05α and 24df

Null hypothesis is rejected, and it is accepted that there is a significant negative association between m and permeability

SPEARMAN CORRELATION OF CEMENTATION FACTOR WITH OTHER VARIABLES :

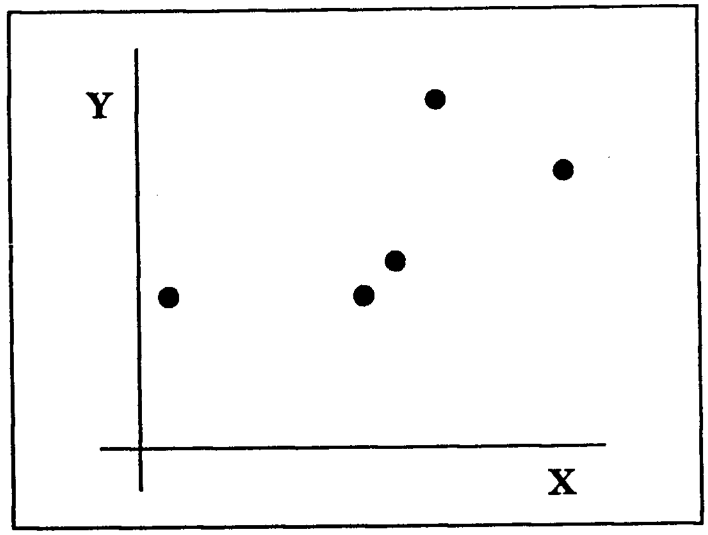
SIGNIFICANT		NOT SIGNIFICANT	
Tortuosity	+0.72 *	Skewness	-0.22
Permeability	-0.64 *	Mean diameter	+0.20
Grainsize range	+0.34 *	Packing	+0.19
Roundness	-0.34 *		

LINEAR REGRESSION OF Y ON X

DATA SET

n samples

X_1	Y_1
X_2	Y_2
X_3	Y_3
....	
X_n	Y_n



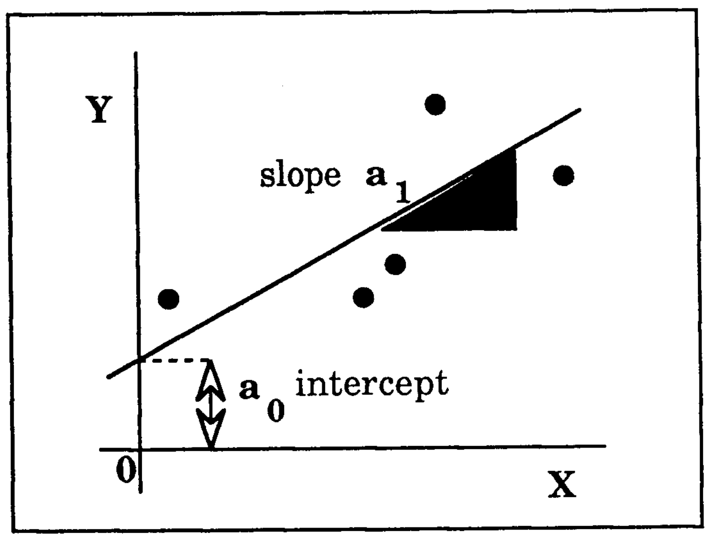
dependent (predicted) variable

independent (predictor) variable

$$\hat{Y} = a_0 + a_1 X$$

intercept

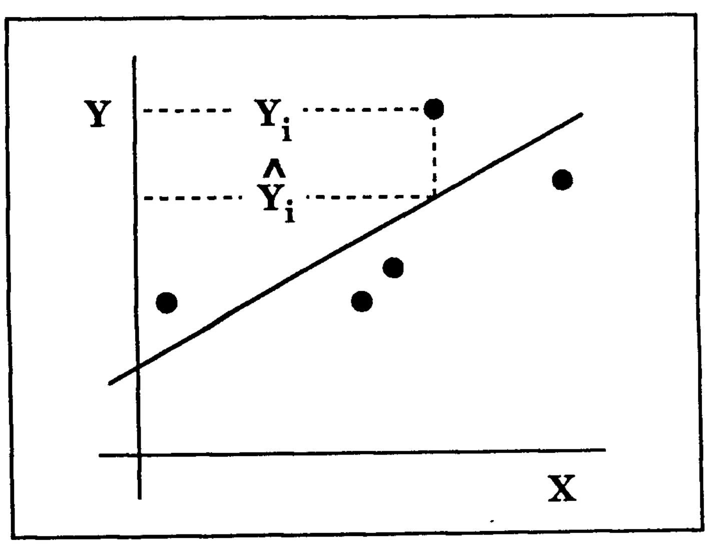
slope



The regression line of Y on X is fitted using the "principle of least squares", which minimises the sum of the squared deviations of Y from its predicted value, $\hat{Y}$

$$\sum (Y_i - \hat{Y}_i)^2 = G$$

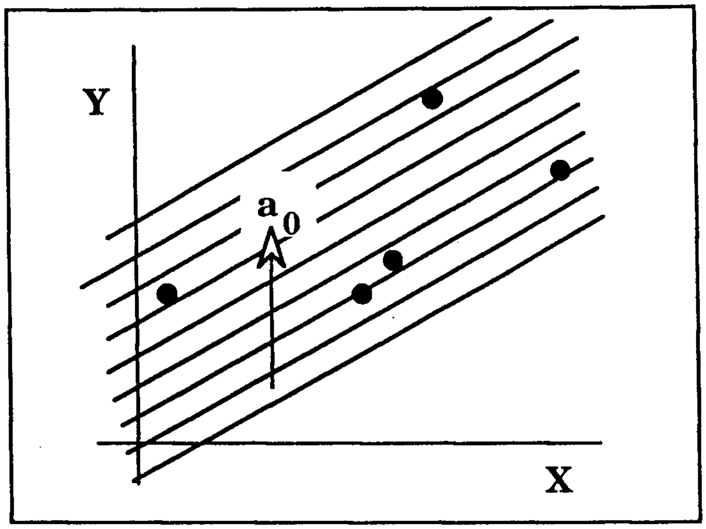
where G is the minimum possible value.

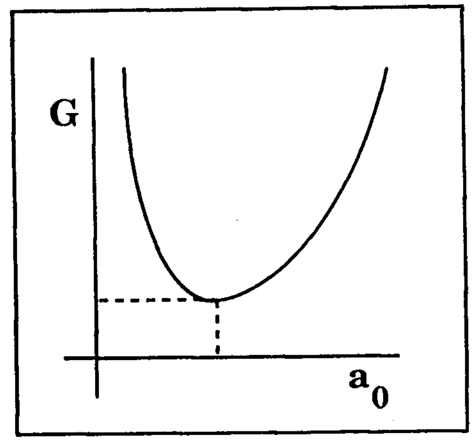


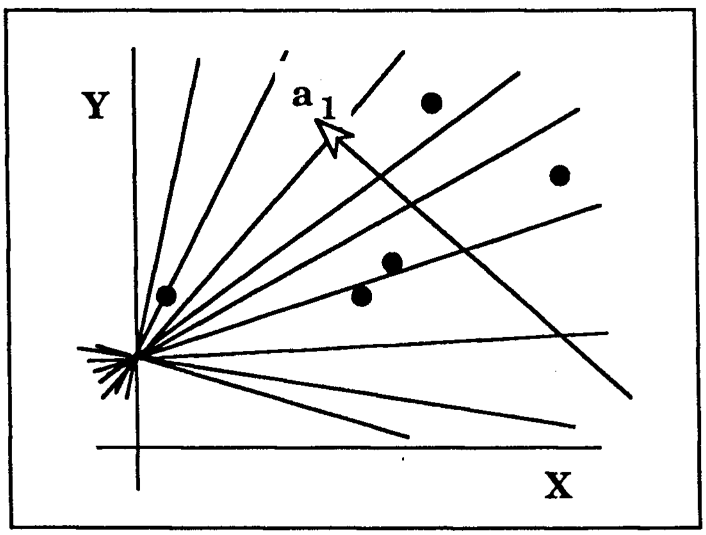
$$\sum(Y_i - \hat{Y}_i)^2 = G = \text{minimum}$$

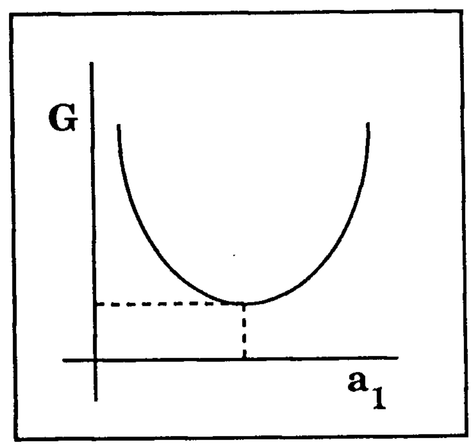
$$\text{But} \ldots\ \hat{Y}_i = a_0 + a_1 X_i$$

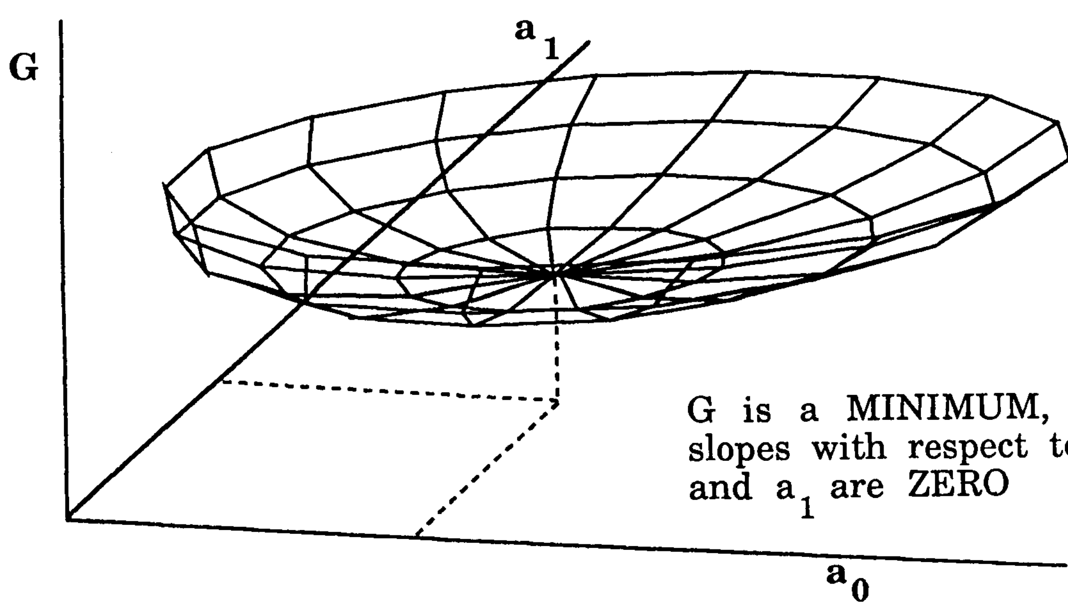
$$\text{So} \ldots \sum(Y_i - a_0 - a_1 X_i)^2 = G = \text{minimum}$$











G is a MINIMUM, when slopes with respect to a_0 and a_1 are ZERO

The slope of an equation is given by the first differential i.e. for equation $y = f(x)$, the slope is dy / dx

If $\sum(Y_I - a_0 - a_1 X)^2 = G$, G is a minimum when the partial differentials with respect to both a_0 and a_1 are zero i.e.

$$\frac{\partial G}{\partial a_0} = 0 \quad \text{and} \quad \frac{\partial G}{\partial a_1} = 0$$

Differentiating :

$$\partial G / \partial a_0 = \sum - (Y_I - a_0 - a_1 X_I) = 0$$
$$\partial G / \partial a_1 = \sum - X_I (Y_I - a_0 - a_1 X_I) = 0$$

Rearranging :

$$n a_0 + a_1 \sum X_I = \sum Y_I$$
$$a_0 \sum X_I + a_1 \sum X_I^2 = \sum X_I Y_I$$

Rewriting in matrix form :

$$\begin{bmatrix} n & \sum X \\ \sum X & \sum X^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum Y \\ \sum XY \end{bmatrix}$$

X **A** **Y**

$$\mathbf{XA} = \mathbf{Y}$$

$$\mathbf{A} = \mathbf{X}^{-1}\mathbf{Y}$$

The vector A is the solution for the intercept a_0 and the slope a_1 of the regression line of Y on X

SOURCES OF VARIATION -- IS THE REGRESSION TREND OF Y ON X SIGNIFICANT?

Sum of squares, regression:	$SS_R = \sum (\hat{Y} - \bar{Y})^2$
Sum of squares, deviation:	$SS_D = \sum (Y - \hat{Y})^2$
Sum of squares, total:	$SS_T = \sum (Y - \bar{Y})^2$

$$SS_T = SS_R + SS_D$$

GOODNESS - OF - FIT is the proportion of the total variation accounted for by the regression:

$$R^2 = SS_R / SS_T$$

R is equal to the correlation coefficient between X and Y.

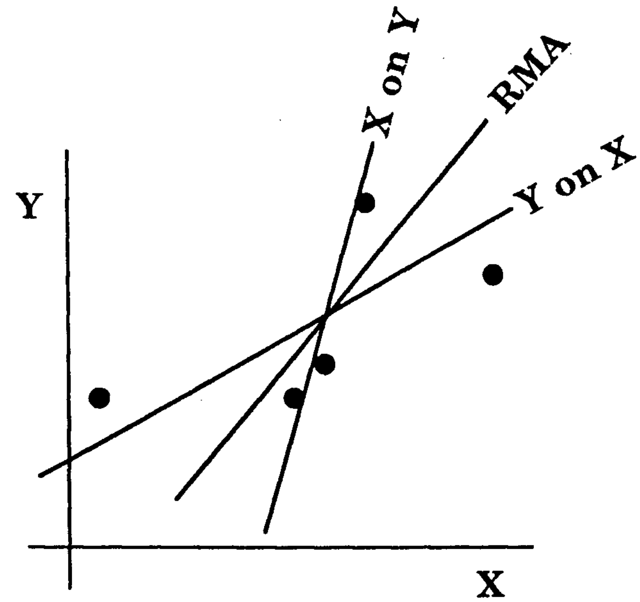
ANALYSIS OF VARIANCE

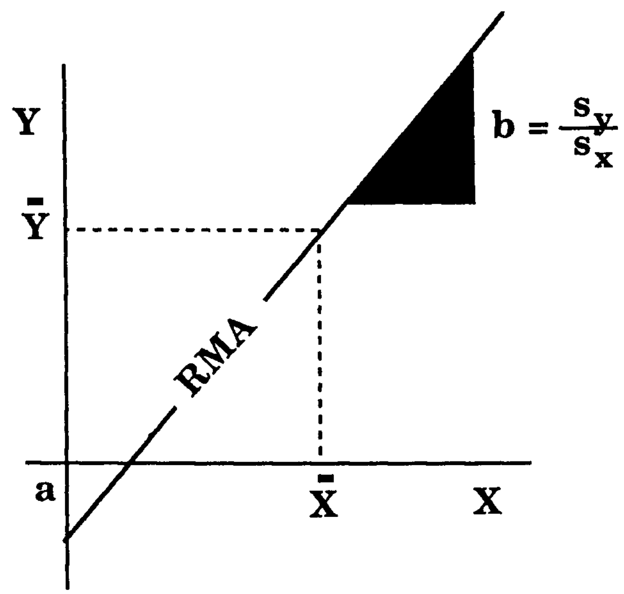
Source of variation	Sum of squares	Degrees of freedom	Mean squares	F - test
Linear regression	SS_R	1	MS_R	MS_R / MS_D
Deviation	SS_D	n - 2	MS_D	
Total variation	SS_T	n - 1		

If this value exceeds the critical F-test value at 1 and (n-2) degrees of freedom at a preselected level of significance, then the null hypothesis that the variance about the trend is no different than the variance about the mean is rejected. In this case, the alternative hypothesis is accepted and the trend considered to be significant.

Blank

AN ALTERNATIVE LINE - FIT : THE REDUCED MAJOR AXIS





Slope = Y standard deviation divided by X standard deviation (S_Y / S_X)
The sign of the slope is the same as that of the correlation coefficient.
The intercept can then be calculated, because the RMA line passes through the mean:

$$\bar{Y} = a + b\bar{X}$$

EXAMPLE : LINEAR REGRESSION OF URANIUM ON LOG(Th/K) IN THE DAKOTA/KIOWA/CHEYENNE FORMATIONS

WELL NAME: DAK TERRALOG
LOCATION:
DATE:
DEPTH: 488.00 TO 896.00 BY .50 FEET
DEPENDENT VARIABLE: UR
INDEPENDENT VARIABLE: LOG(TH/K)

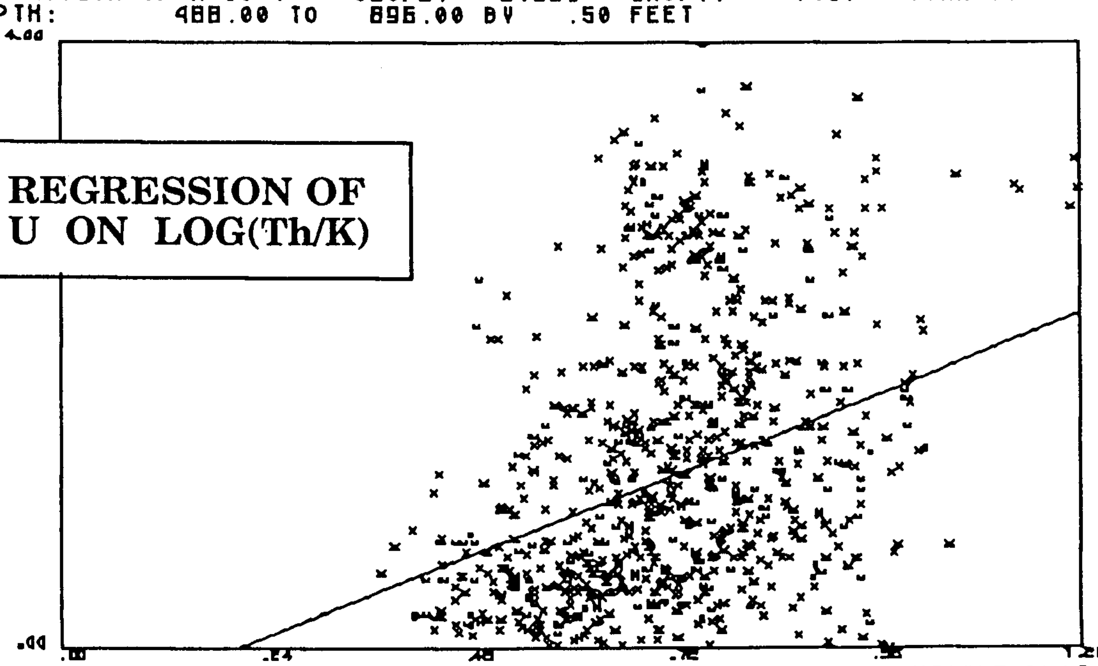
DAKOTA

SIMPLE REGRESSION

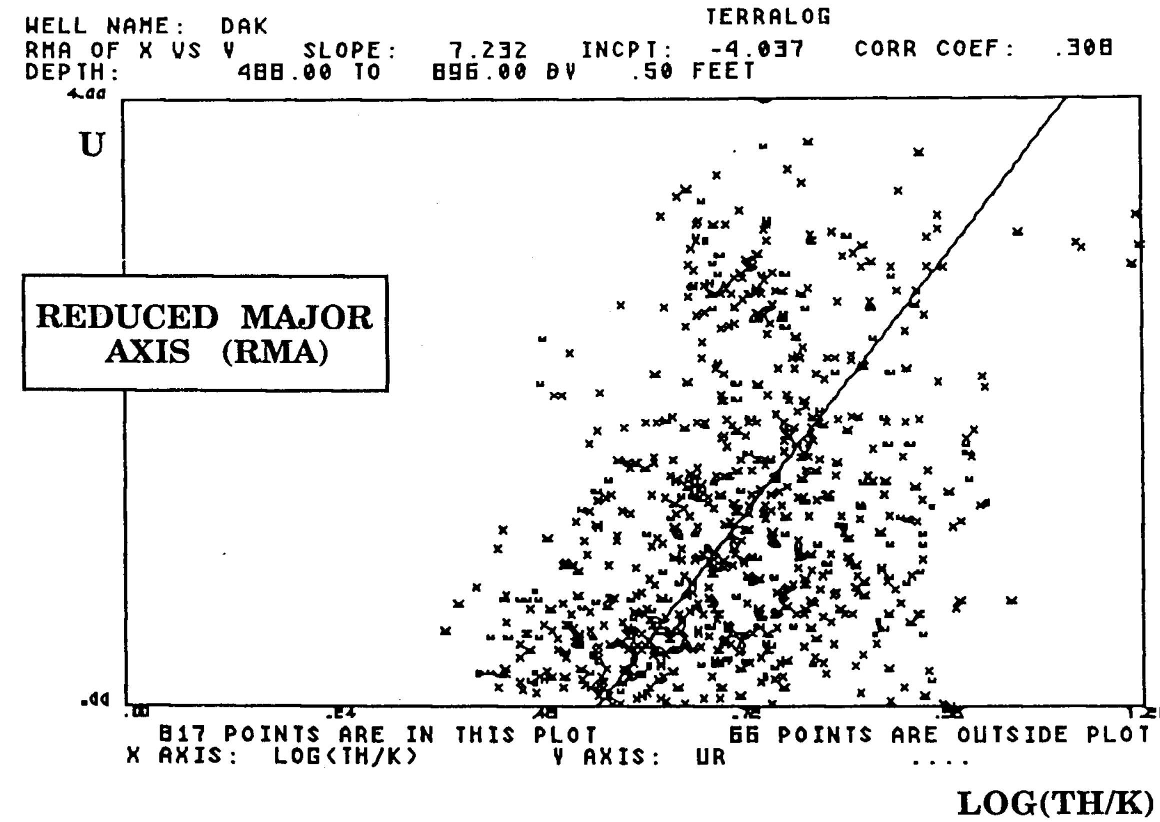
ORDER OF REGRESSION	0	1	2	3	4
REGRESSION COEF A:	1.140	-.457	-2.322	-5.894	-7.964
REGRESSION COEF X:		2.231	7.496	22.648	33.371
REGRESSION COEF X**2:			-3.589	-24.136	-44.375
REGRESSION COEF X**3:				8.980	25.000
REGRESSION COEF X**4:					-4.640
NUMBER OF SAMPLES	817	817	817	817	817
SUM OF SQUARES TOTAL	781.0	781.0	781.0	781.0	781.0
SUM OF SQUARES REGRESSION	.0	74.3	81.3	84.4	73.2
SUM OF SQUARES DEVIATION	781.0	706.7	699.7	696.6	707.8
GOODNESS OF FIT	.000	.095	.104	.108	.093
COEF. OF MULTIPLE CORR.	.000	.308	.322	.328	.306 ....

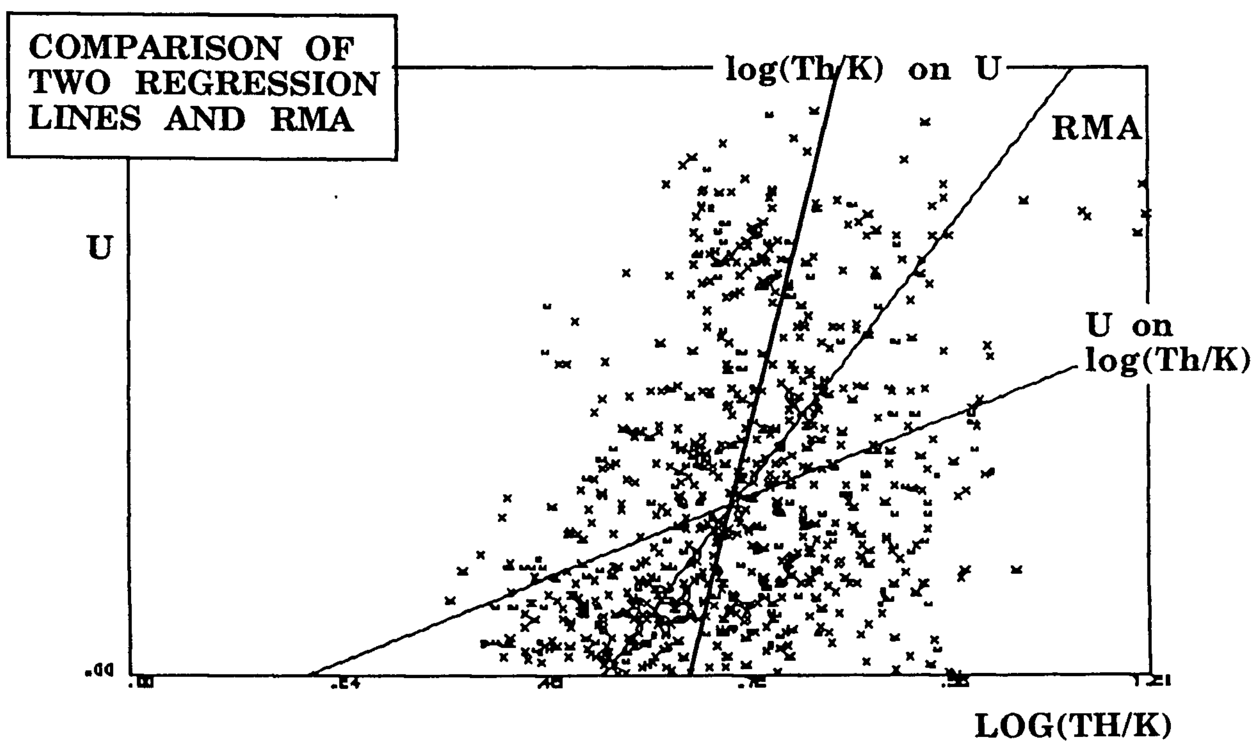
F ratio = 85.8 , and is significant at the 5% level

WELL NAME: DAK TERRALOG
REGRESSION OF X VS Y SLOPE: 2.231 INCPT: -.457 CORR COEF: .308
DEPTH: 488.00 TO 896.00 BY .50 FEET



817 POINTS ARE IN THIS PLOT 66 POINTS ARE OUTSIDE PLOT
X AXIS: LOG(TH/K) Y AXIS: UR ....


WELL NAME: DAK
TERRALOG
RMA OF X VS Y SLOPE: 7.232 INCPT: -4.037 CORR COEF: .308
DEPTH: 488.00 TO 896.00 BY .50 FEET
U
REDUCED MAJOR AXIS (RMA)
817 POINTS ARE IN THIS PLOT
66 POINTS ARE OUTSIDE PLOT
X AXIS: LOG(TH/K)
Y AXIS: UR
LOG(TH/K)


COMPARISON OF TWO REGRESSION LINES AND RMA
log(Th/K) on U
RMA
U
U on log(Th/K)
LOG(TH/K)

THE GENERAL REGRESSION MODEL

A dependent (or predicted) variable Y, is regressed on m independent (predictor) variables $X_1, X_2, \dots\dots X_m$.
The n observation sets can be symbolized as :

$$
\text{DATA INPUT} \longrightarrow
\begin{array}{|lllll|}
\hline
Y_1 & X_{11} & X_{12} & \cdots\cdots & X_{1m} \\
Y_2 & X_{21} & X_{22} & \cdots\cdots & X_{2m} \\
y_3 & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\
\cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\
\cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\
Y_n & X_{m1} & X_{n2} & \cdots\cdots & X_{nm} \\
\hline
\end{array}
\quad \text{n sets of observations}
$$

dependent variable **m independent variables**

The regression equation is : $\hat{Y} = a_0 + a_1X_1 + a_2X_2 + \dots\dots + a_mX_m$
The vector of predicted values of Y for all n observation sets can be written in matrix form as :

$$
\begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \cdots \\ \cdots \\ \cdots \\ \hat{Y}_m \end{bmatrix}
=
\begin{bmatrix}
1 & X_{11} & X_{12} & \cdots\cdots & X_{1m} \\
1 & X_{21} & X_{22} & \cdots\cdots & X_{2m} \\
\cdots & \cdots & \cdots & \cdots\cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots\cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots\cdots & \cdots \\
1 & X_{m1} & X_{n2} & \cdots\cdots & X_{nm}
\end{bmatrix}
\begin{bmatrix} a_0 \\ a_1 \\ \cdots \\ \cdots \\ \cdots \\ a_m \end{bmatrix}
$$

which can be symbolized as $\hat{Y} = XA$

Now, the solution is found by minimizing the sum of squares deviations between $\hat{Y}_i$ and Y_i , given by :

$$G = \sum(Y_i - \hat{Y}_i)^2 = \sum(Y_i - (a_0 + a_1X_{1i} + \dots\dots + a_mX_{mi}))^2$$

The partial differentials : $\dfrac{\partial G}{\partial a_0} = 0 \;\dots\; \dfrac{\partial G}{\partial a_1} = 0 \;\dots\dots\; \dfrac{\partial G}{\partial a_m} = 0$

These m equations rearranged in matrix form are :

$$\begin{bmatrix} n & \sum X_1 & \sum X_2 & \cdots\cdots & \sum X_m \\ \sum X_1 & \sum X_1^2 & \sum X_1X_2 & \cdots\cdots & \sum X_1X_m \\ \sum X_2 & \sum X_1X_2 & \sum X_2^2 & \cdots\cdots & \sum X_2X_m \\ \cdots & \cdots & \cdots & \cdots\cdots\cdots & \\ \cdots & \cdots & \cdots & \cdots\cdots\cdots & \\ \sum X_m & \cdots & \cdots & \cdots\cdots & \sum X_m^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \cdots \\ \cdots \\ \cdots \\ a_m \end{bmatrix} = \begin{bmatrix} \sum Y \\ \sum X_1Y \\ \sum X_2Y \\ \cdots \\ \cdots \\ \sum X_mY \end{bmatrix}$$

$$SA = P$$

$$\therefore A = S^{-1}P$$

$$\text{But} \cdots S = X^TX \cdots \text{and} \cdots P = X^TY$$

$$\therefore A = (X^TX)^{-1}X^TY$$

which gives the coefficient unknowns for the general regression equation : $\hat{Y} = a_0 + a_1X_1 + a_2X_2 + \cdots\cdots a_mX_m$

When there is only one independent variable, X1, this is the solution for SIMPLE LINEAR REGRESSION :

$$\hat{Y} = a_0 + a_1X$$

When there are several independent variables, this is the solution for MULTIPLE REGRESSION :

$$\hat{Y} = a_0 + a_1X_1 + a_2X_2 + \cdots\cdots a_mX_m$$

When the independent variables are powers of a single independent variable, this is the solution for POLYNOMIAL REGRESSION : $\hat{Y} = a_0 + a_1X + a_2X^2 + \cdots\cdots a_mX^m$

When Y is measured at geographic locations and two independent variables are polynomial combinations of geographic coordinates, this is the solution for TREND SURFACE ANALYSIS : $\hat{Y} = a_0 + a_1U + a_2V + \cdots\cdots$

When the relationship between dependent and independent variables is of the form :

$$\hat{Y} = aX^b \cdots \text{then} \cdots \log \hat{Y} = \log a + b \cdot \log X$$

and this is a solution for NON-LINEAR REGRESSION.

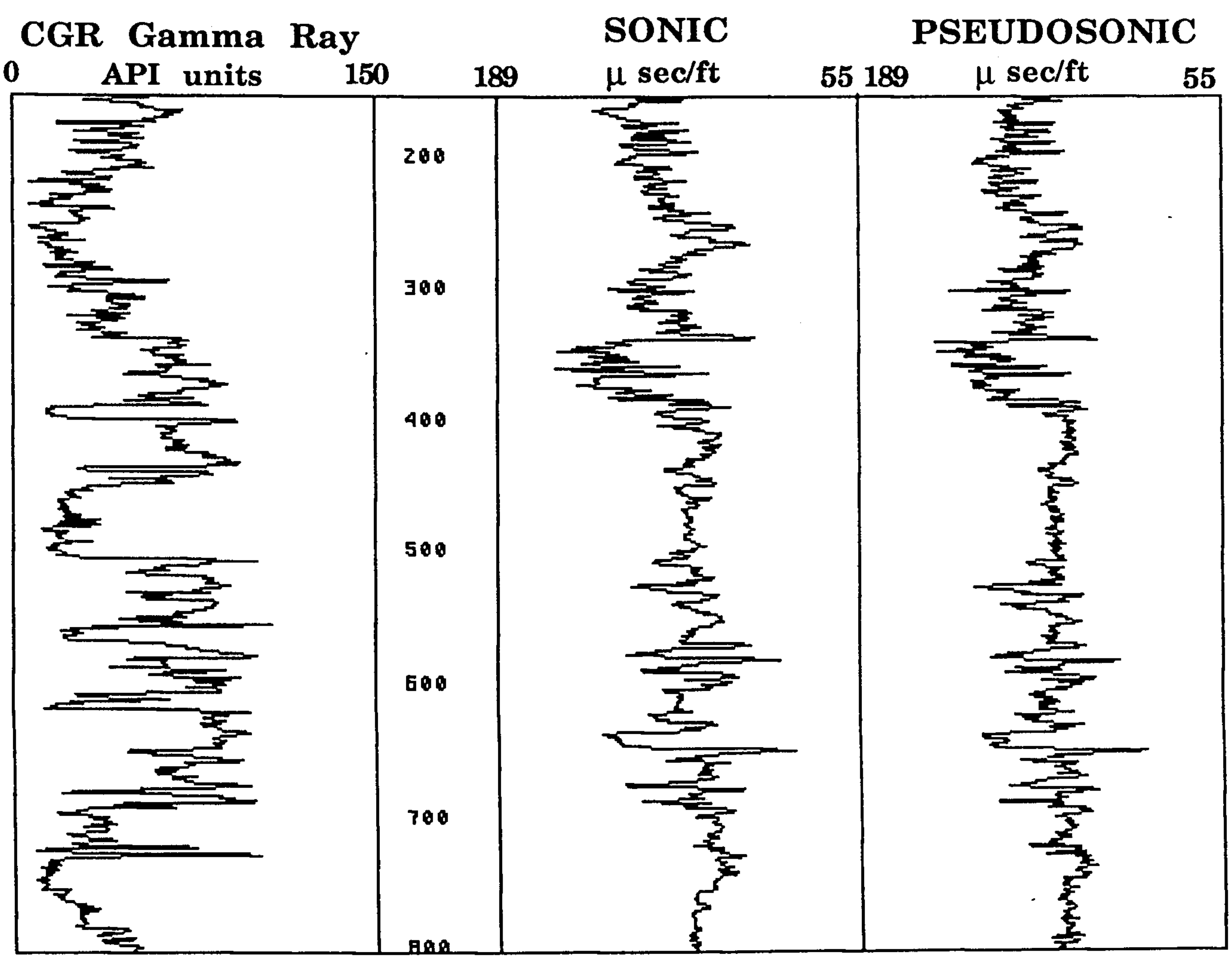
MULTIPLE REGRESSION : PREDICTION

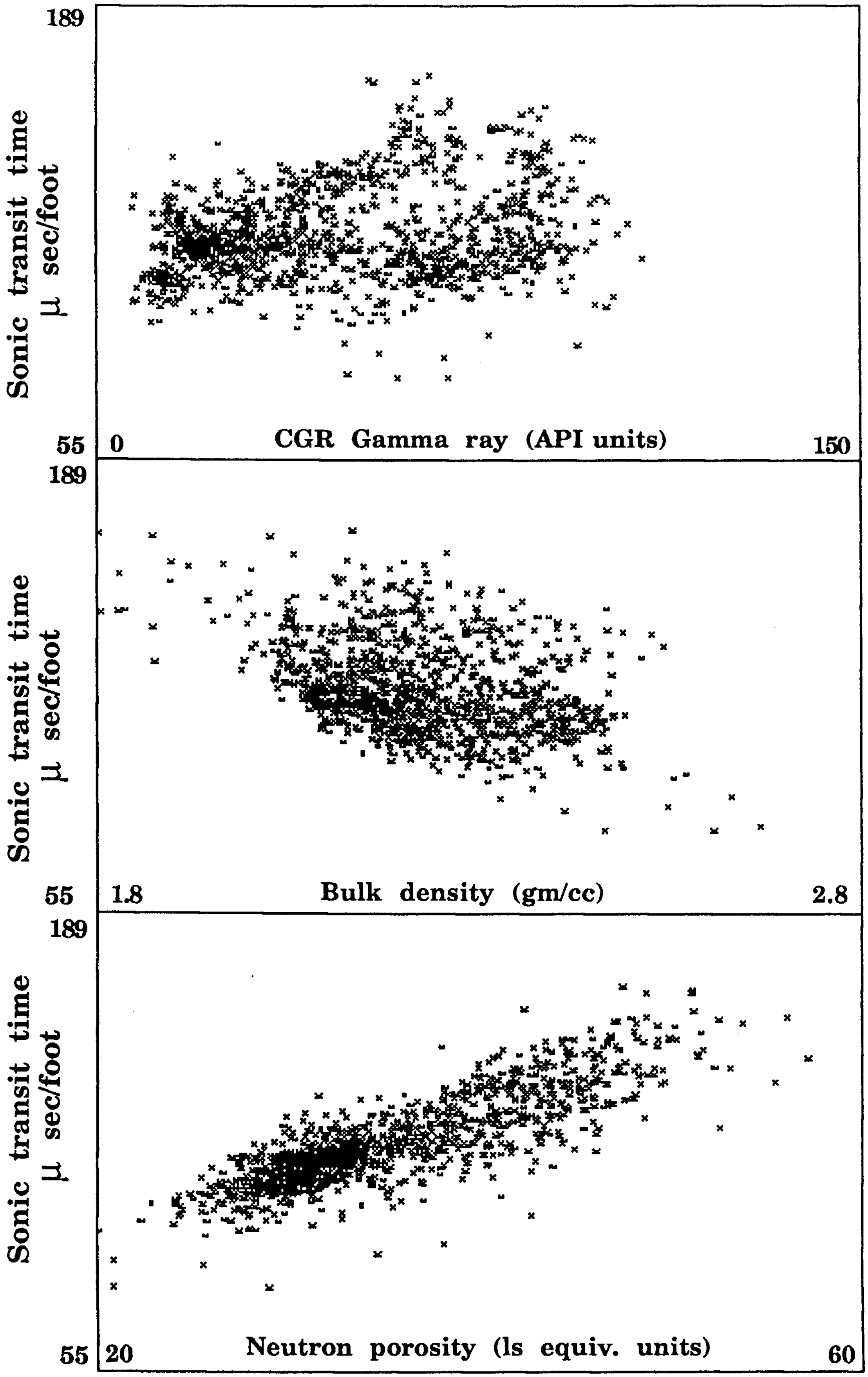
Development of prediction equation to generate pseudo-sonic logs based on regression of sonic transit time on gamma-ray, density, and neutron logs.

$$\hat{\Delta t} = a_0 + a_1 G + a_2 \rho_b + a_3 \Phi_n$$

```
WELL NAME:
LOCATION:
DATE:
DEPTH:          150.00 TO    800.00 BY     .50 FEET
DEPENDENT VARIABLE:        DT

MULTIPLE REGRESSION

CONSTANT                   77.195        .000
CGR                          .029        .054
RHOB                       -7.425       -.068
NPHI                        1.689        .807

           SUM OF SQUARES   DEGREES-OF-FREEDOM    MEAN-SQUARES    F-TEST
  REGRESSION   162579.81            3              54193.269  1139.192
  DEVIATION     61700.43         1297                 47.571
  TOTAL        224280.25         1300 ....
```

$$\hat{\Delta t} = 77.2 + 0.03G - 7.4\rho_b + 1.69\Phi_n$$




189
Sonic transit time
μ sec/foot
55
0
CGR Gamma ray (API units)
150
189
Sonic transit time
μ sec/foot
55
1.8
Bulk density (gm/cc)
2.8
189
Sonic transit time
μ sec/foot
55
20
Neutron porosity (ls equiv. units)
60

MULTIPLE REGRESSION : RESIDUALS

In this example, the variation of uranium is examined, as logged in the Dakota, Kiowa and Cheyenne formations. What constitutes anomalously high uranium or anomalously low uranium ? One approach to this question is to compute a multiple regression of uranium on other log variables. These were selected as apparent grain density, apparent volumetric photoelectric cross-section, neutron porosity, thorium and potassium. So the regression equation is :

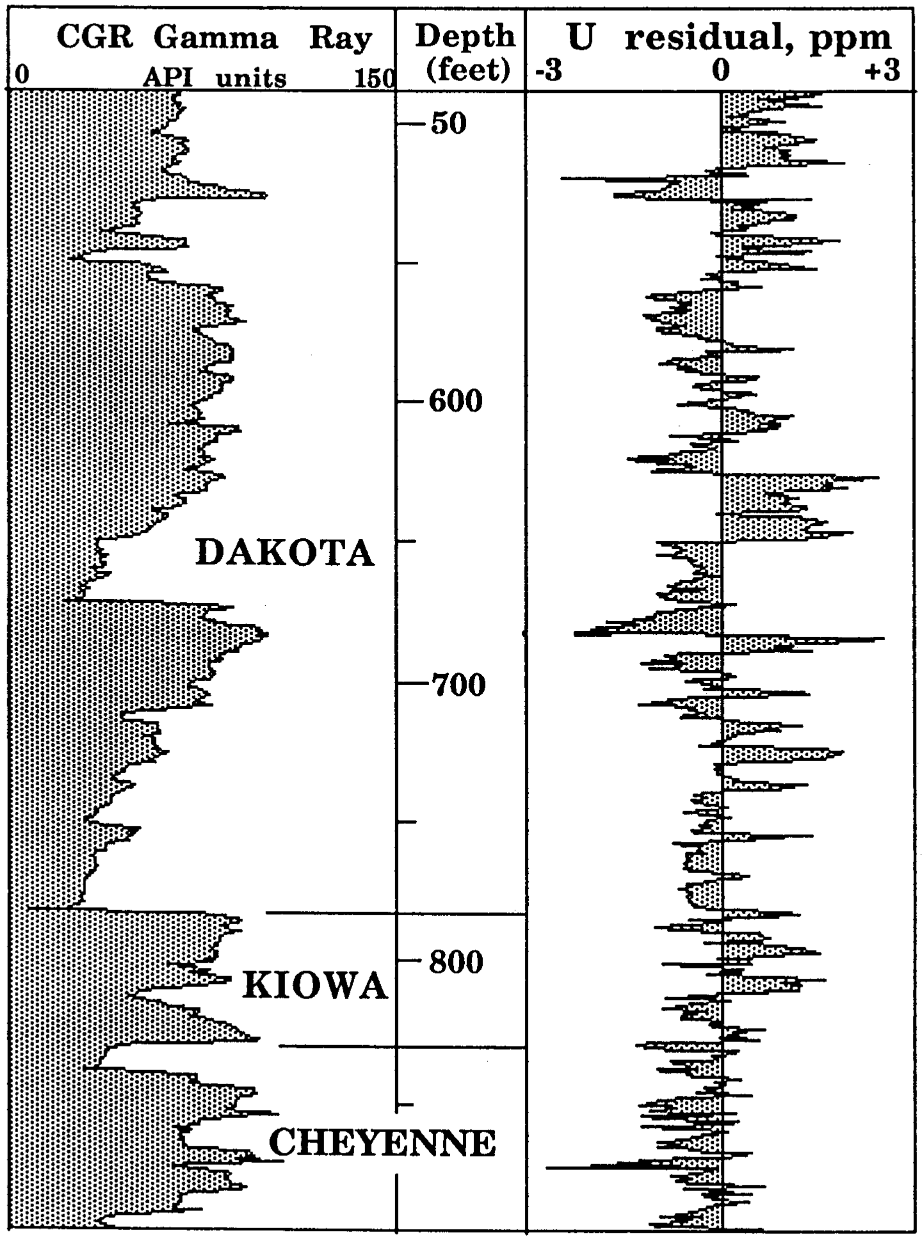
$$\hat{U} = a_0 + a_1\rho_{maa} + a_2U_{maa} + a_3\phi_n + a_4Th + a_5K$$

The results of the regression are :

```
WELL NAME:  DAK                                    TERRALOG
LOCATION:
DATE:
DEPTH:          488.00  TO    896.00  BY    .50  FEET
DEPENDENT VARIABLE:        UR

MULTIPLE REGRESSION

CONSTANT                       -15.625       .000
RHOMAA                           6.110       .699
UMAA                             -.037      -.061
CNL                               .025       .135
TH                               -.057      -.182
K                                -.436      -.218

            SUM OF SQUARES   DEGREES-OF-FREEDOM    MEAN-SQUARES   F-TEST
  REGRESSION     261.16              5                52.233      81.479
  DEVIATION      519.90            811                  .641
  TOTAL          781.06            816 ....
```

Critical F value at 5% and 5 & 811 df = 2.22

The regression is statistically significant and the trend in uranium most strongly linked with apparent grain density. At any depth level, a comparison between the actual uranium and that predicted, is an indication of whether the uranium is higher, lower, or approximately equal to the amount expected on the basis of the other log properties. A log profile of the uranium RESIDUALS highlights anomalous zones of relative enrichment or impoverishment as calibrated by the Dakota/Kiowa/Cheyenne section.


CGR Gamma Ray
0
API units
150
Depth
(feet)
U residual, ppm
-3
0
+3
50
600
700
800
DAKOTA
KIOWA
CHEYENNE

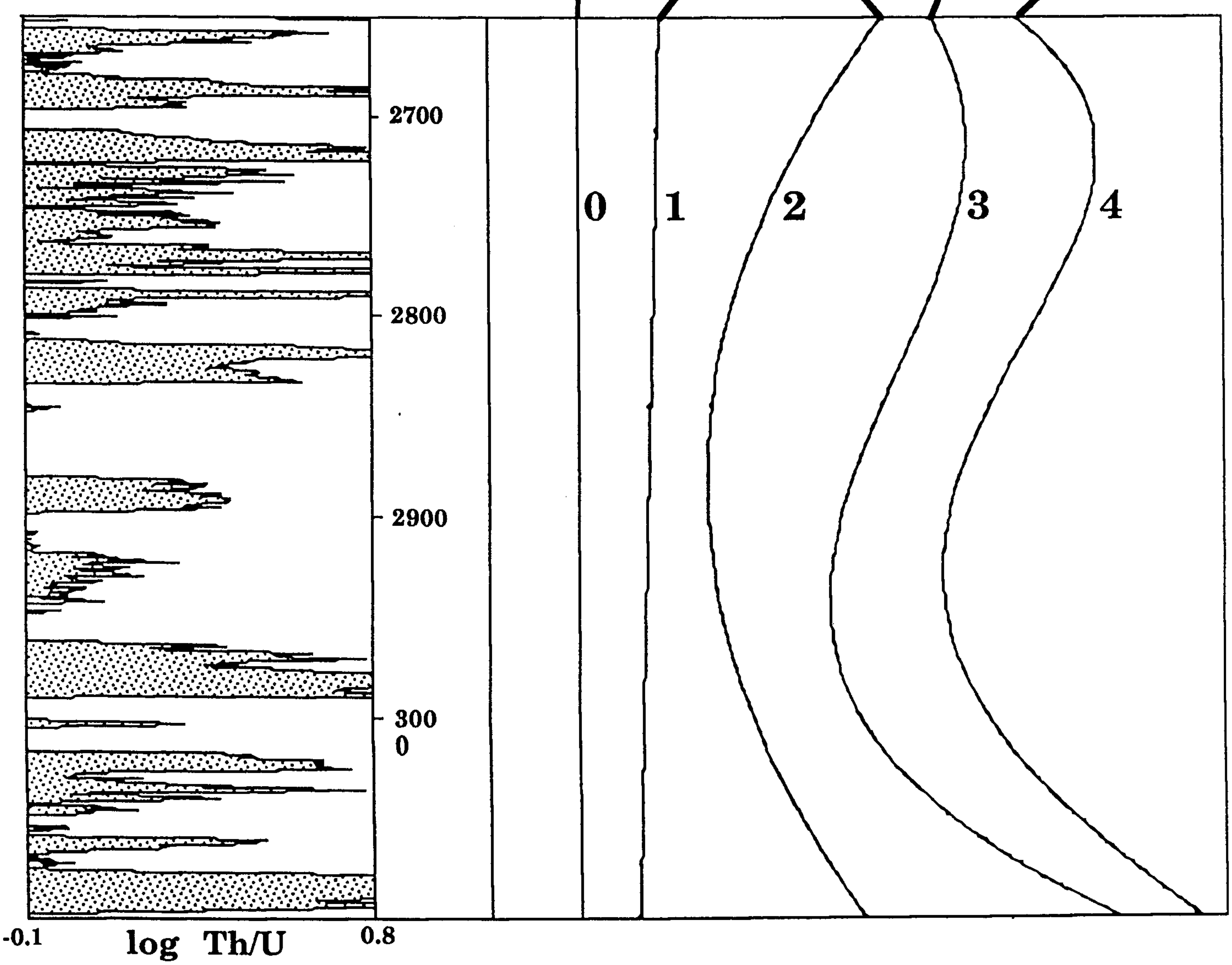
DAKOTA

EXAMPLE : POLYNOMIAL REGRESSION OF log(Th/U) ON DEPTH IN CHASE GROUP AS POTENTIAL INDICATOR OF LONG-TERM TRENDS IN REDOX POTENTIAL

WELL NAME: CHASE — TERRALOG
LOCATION:
DATE:
DEPTH: 2650.00 TO 3100.00 BY .50 FEET
DEPENDENT VARIABLE: LOGTHU
INDEPENDENT VARIABLE: (DEP-2650)/1000

SIMPLE REGRESSION

ORDER OF REGRESSION	0	1	2	3	4
REGRESSION COEF A:	.137	.167	.451	.184	.102
REGRESSION COEF X:		-.131	-3.935	3.228	6.895
REGRESSION COEF X**2:			8.482	-31.469	-68.218
REGRESSION COEF X**3:				59.371	186.605
REGRESSION COEF X**4:					-141.738
NUMBER OF SAMPLES	895	895	895	895	895
SUM OF SQUARES TOTAL	226.7	226.7	226.7	226.7	226.7
SUM OF SQUARES REGRESSION	.0	.2	14.8	24.0	24.6
SUM OF SQUARES DEVIATION	226.7	226.4	211.9	202.6	202.0
GOODNESS OF FIT	.000	.001	.065	.106	.108
COEF. OF MULTIPLE CORR.	.000	.033	.255	.325	.329 ....



CUBIC POLYNOMIAL : $Z = a_0 + a_1 d + a_2 d^2 + a_3 d^3$

where $Z = \log(Th/U)$ and d = depth (standardized as true depth - 2650 / 1000)

From regression analysis : $a_0 = 0.184$, $a_1 = 3.228$, $a_2 = -31.469$, $a_3 = 59.371$

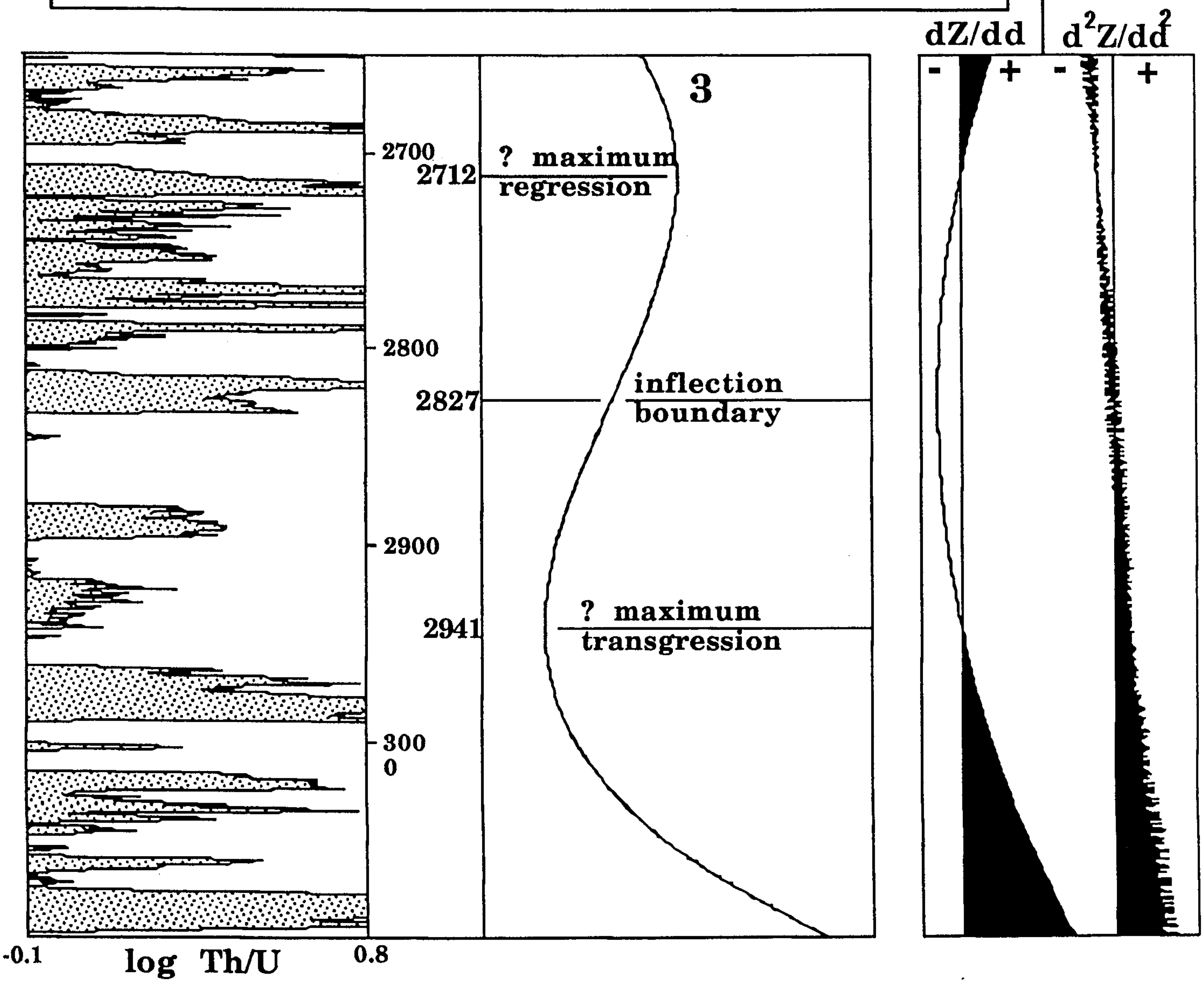
First derivative (slope) = $dZ/dd = a_1 + 2a_2 d + 3a_3 d^2$

$dZ/dd = 0$ at maxima or minima, and gives 2712 and 2941 feet

Second derivative = $d^2Z/dd^2 = 2a_2 + 6a_3 d$

$d^2Z/dd^2 = 0$ at inflection points, and gives one solution at 2827 feet.

OR "Quick - and - dirty"....
Use first difference of cubic trend curve for dZ/dd by applying a FILTER with elements (-1, 1) and second difference for d^2Z/dd^2 using FILTER (1,-2,1)



REGRESSION OF TRENDS IN SEQUENCES OF EVENTS

If "events" occur randomly in time (or space), then their distribution is described by a Poisson model. Since the events are independent, there is no trend with time, and their rate of occurrence is a constant :

$$r = e^{a_0} \dots \text{and} \cdot \text{so} \dots \log_e r = a_0$$

where r is the rate and a is a constant.

The reciprocal of the rate of occurrence per unit time is the average distance , l , between the events in time units, and so :

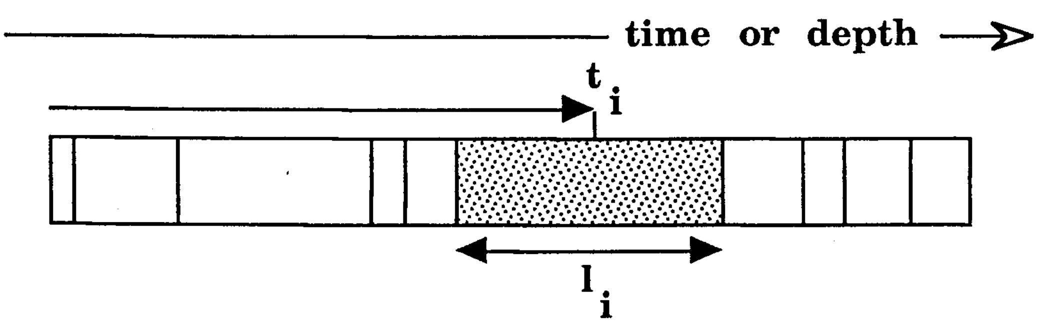
$$a_0 = \log_e r = \log_e\left(\tfrac{1}{l}\right) = -\log_e l$$

When the rate of events changes as a trend function of time, then the changes can be modelled by a polynomial series :

$$r = e^{a_0 + a_1 t + a_2 t^2 + \dots + a_m t^m}$$

$$\log_e r = a_0 + a_1 t + a_2 t^2 + \dots + a_m t^m$$

$$\therefore -\log_e l = a_0 + a_1 t + a_2 t^2 + \dots + a_m t^m$$



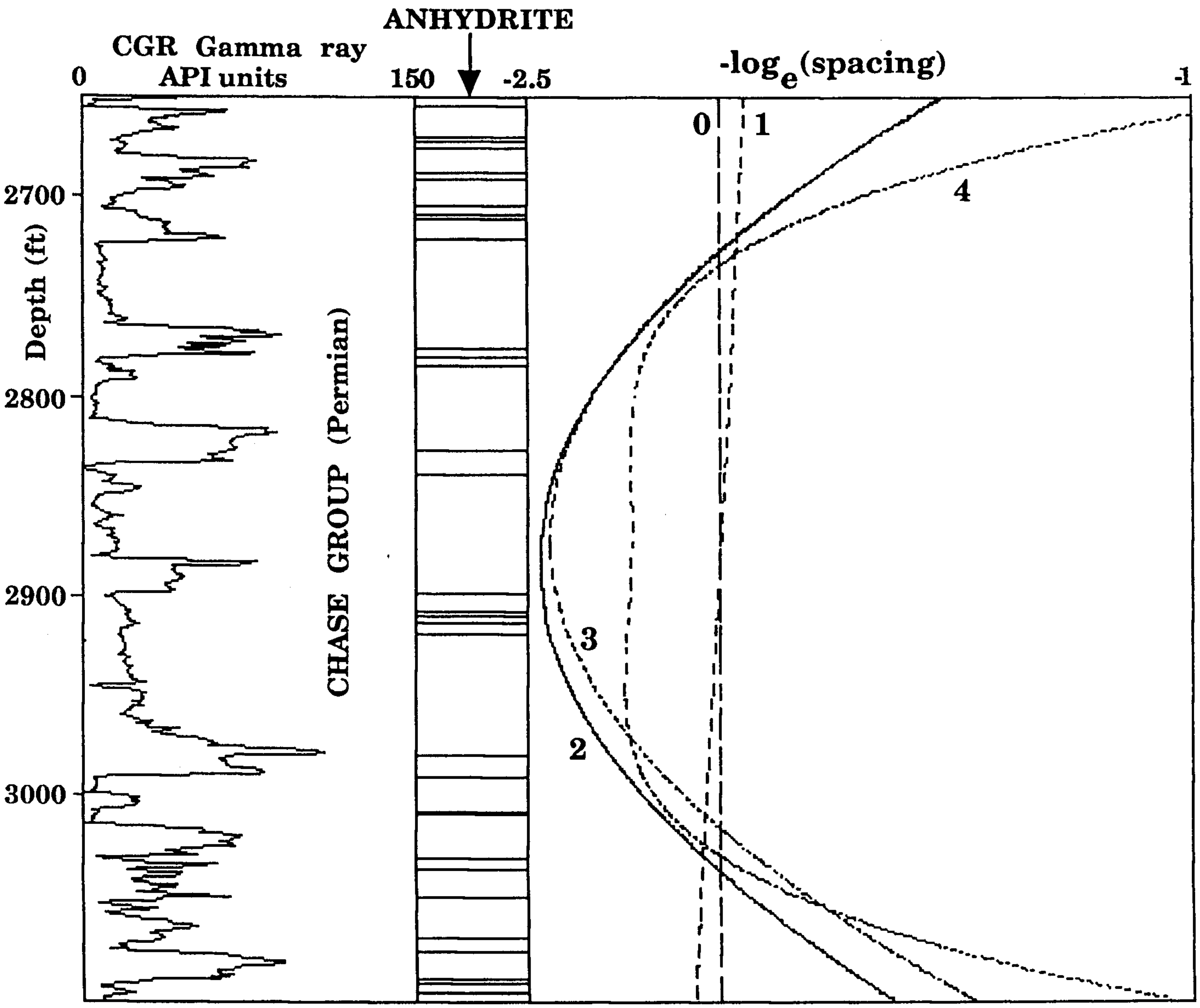
which can be analyzed as a polynomial regression of a data set of the negative natural (or base 10) logarithm of the distance between successive events (-log l) on the elapsed time (or depth) to the midpoint between the two events. The resulting curves graph out polynomial trends in the rate of occurrence as it changes with time. On the next page, linear through fourth order trends are shown for the occurrence of anhydrite events in the Chase Group.

WELL NAME: CHASE TERRALOG
LOCATION:
DATE:
DEPTH:
DEPENDENT VARIABLE: THPOISS
INDEPENDENT VARIABLE: DEPPOISS

CHASE

SIMPLE REGRESSION

ORDER OF REGRESSION	0	1	2	3	4
REGRESSION COEF A:	-2.067	-2.010	-1.564	-1.556	-1.343
REGRESSION COEF X:		-.255	-7.820	-8.078	-18.613
REGRESSION COEF X**2:			16.844	18.277	127.250
REGRESSION COEF X**3:				-2.093	-378.937
REGRESSION COEF X**4:					413.687
NUMBER OF SAMPLES	33	33	33	33	33
SUM OF SQUARES TOTAL	33.3	33.3	33.3	33.3	33.3
SUM OF SQUARES REGRESSION	.0	.0	2.6	2.6	2.8
SUM OF SQUARES DEVIATION	33.3	33.3	30.7	30.7	30.4
GOODNESS OF FIT	.000	.001	.079	.079	.086
COEF. OF MULTIPLE CORR.	.000	.039	.281	.281	.293



(Note that DEPPOISS = (Depth - 2650) / 1000 in regression equations)

NON-LINEAR REGRESSION

When the descriptive equation that links two variables is of the form : $Y = aX^b$

the trend is non-linear, but may be linearized by a logarithmic transformation : $\log Y = \log a + b \log X$

Following the transformation, the data can be analyzed by simple linear regression. However, it should be realized that the squared errors that are minimized are in logarithmic, rather than arithmetic units.

The most well-known example of this equation in petrophysics is the modified Archie equation : $F = \frac{a}{\phi^m}$

An example of the solution of the Archie equation by non-linear regression is shown on the next page.

Other examples include prediction equations that link permeability and porosity such as :

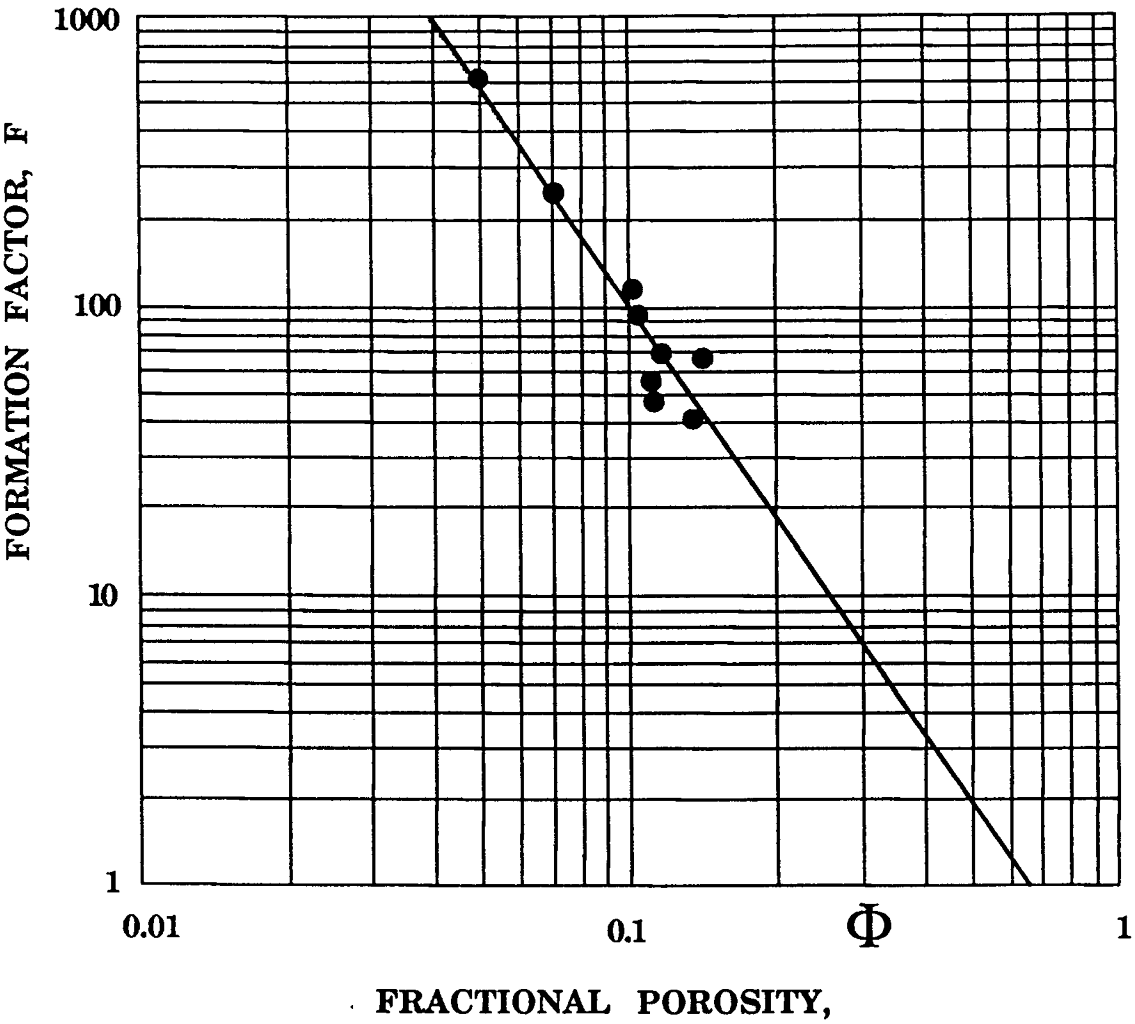
$$\phi = a k^b$$

(Muskat,1949).

or the possible prediction of clay mineral exchange cations in shaly sands by some power of the porosity the equation: $Q_v = d\phi^{-e}$

as suggested by Lavers and others (1974), where the constant d, and exponent e, probably function as dimensional modifiers to convert porosity as volume to a measure of internal surface area.

EXAMPLE : REGRESSION CALCULATION OF ARCHIE EQUATION CONSTANTS FOR ARBUCKLE LIMESTONE, BASED ON CORE MEASUREMENTS OF FORMATION FACTOR AND POROSITY



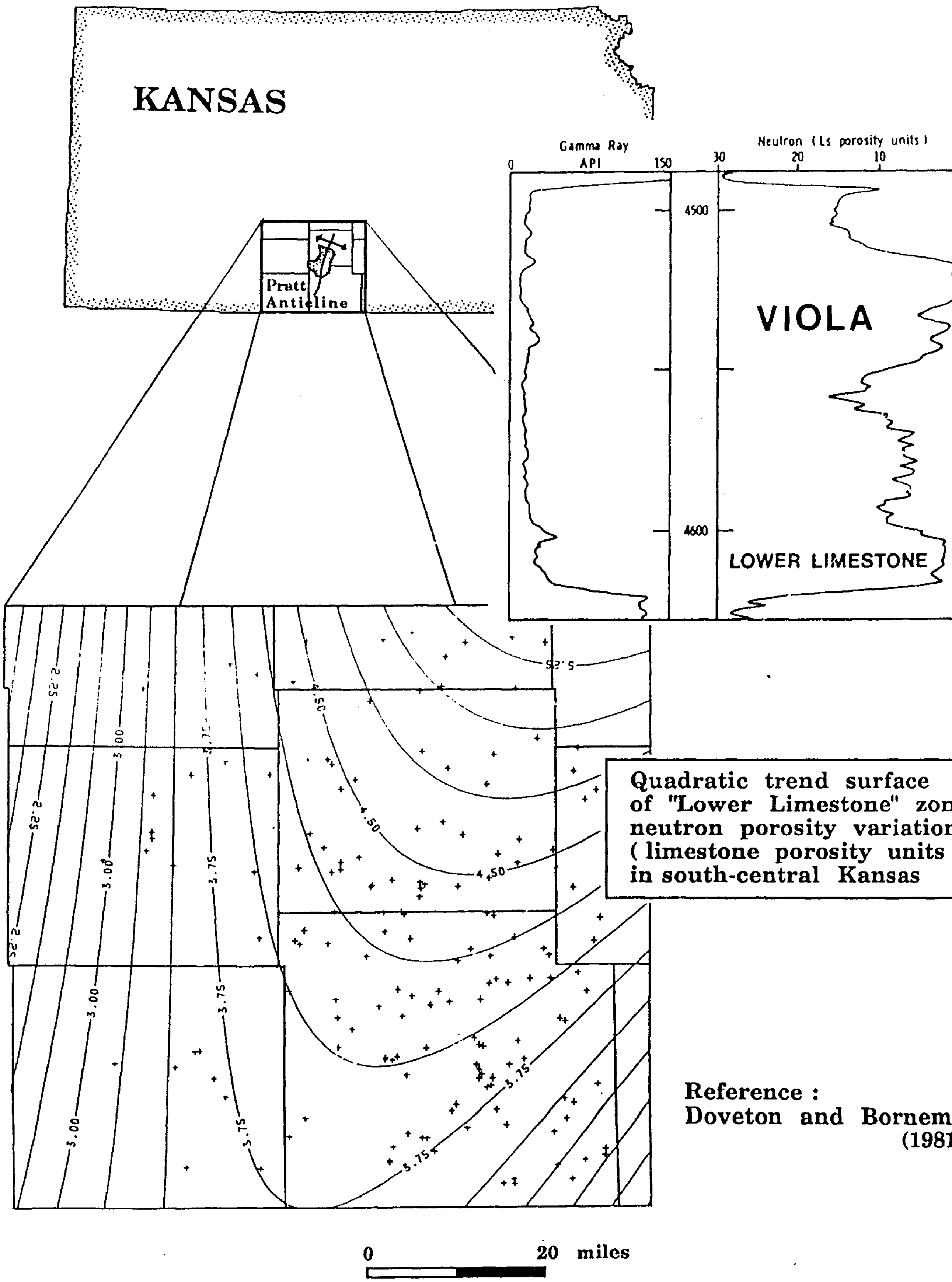
Equation of regression line (F - on - Φ) :

$$\log \hat{F} = -0.445 - 2.444 \log \Phi$$

$$\boxed{\therefore \hat{F} = \frac{0.36}{\Phi^{2.44}}}$$

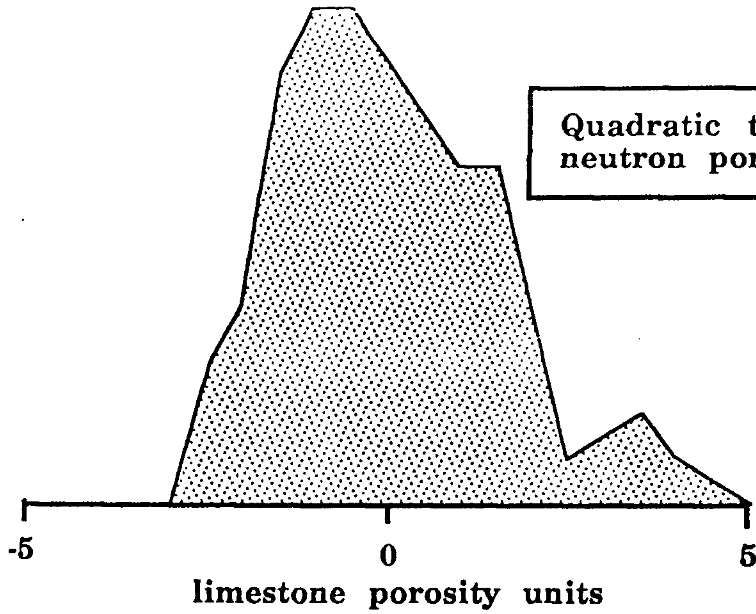
(The reduced major axis (RMA) solution is : $F = \dfrac{0.27}{\Phi^{2.57}}$)

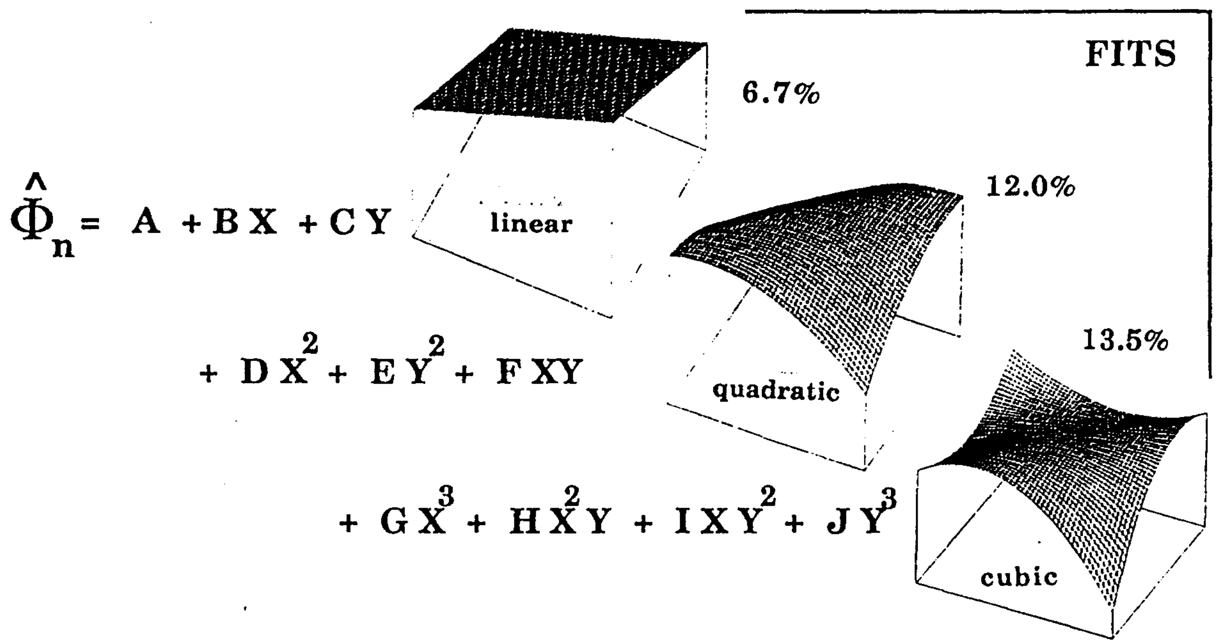
LOG NORMALIZATION BY TREND - SURFACE ANALYSIS



Quadratic trend surface of "Lower Limestone" zone neutron porosity variation (limestone porosity units) in south-central Kansas

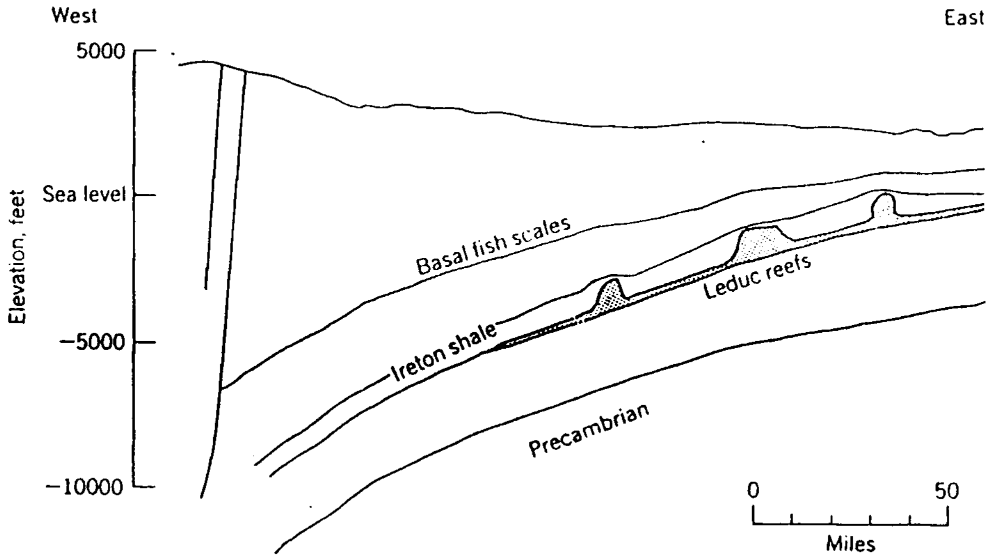
Reference :
Doveton and Bornemann (1981)



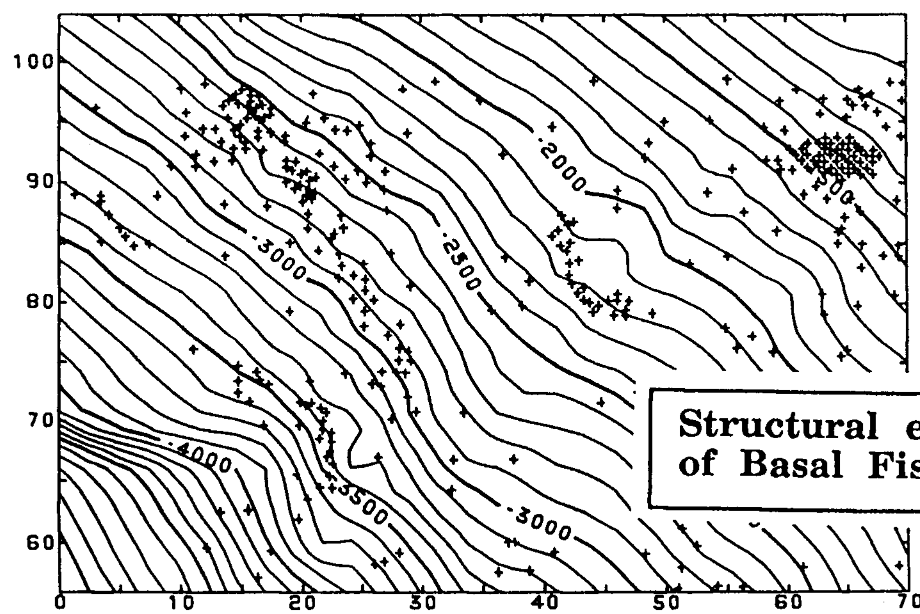


Source of Variation	Sum of Squares	DF	Mean Squares	F Ratio
Linear regression	27.22	2	13.61	5.73*
Linear deviation	377.62	159	2.37	
Quadratic–linear regression	21.41	3	7.14	3.13*
Quadratic deviation	356.22	156	2.28	
Cubic-quadratic regression	5.99	4	1.50	0.65
Cubic deviation	350.23	152	2.30	
Total variation	404.85	161		

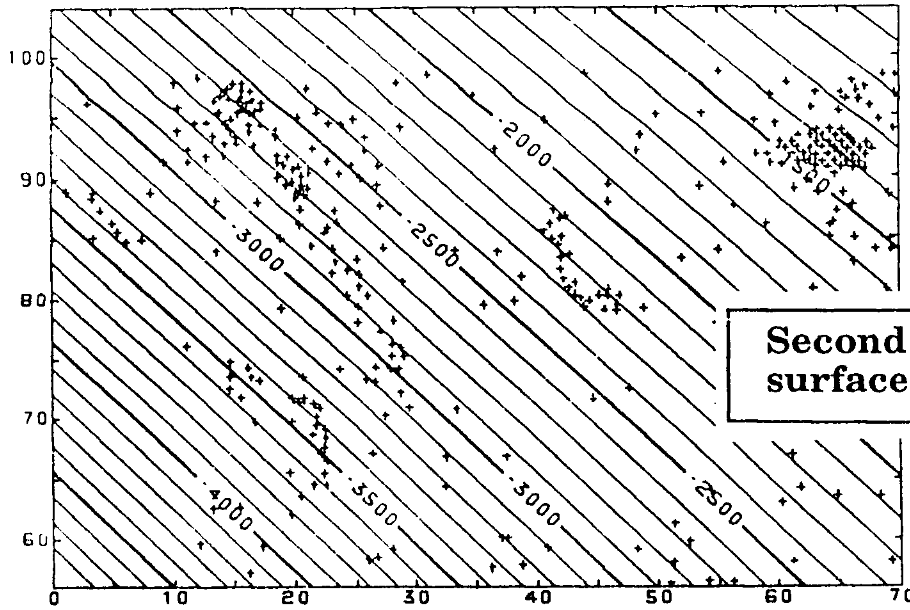
*Significant at 5 percent level.



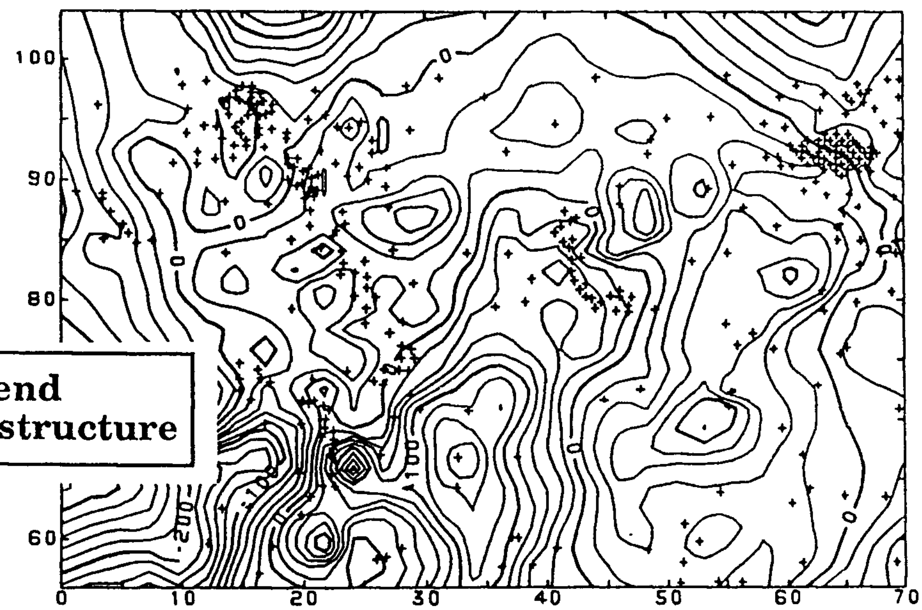
Cross-section of western Alberta, showing structural drape of Basal Fish Scales (Lower Cretaceous) over Leduc reefs (Upper Devonian)



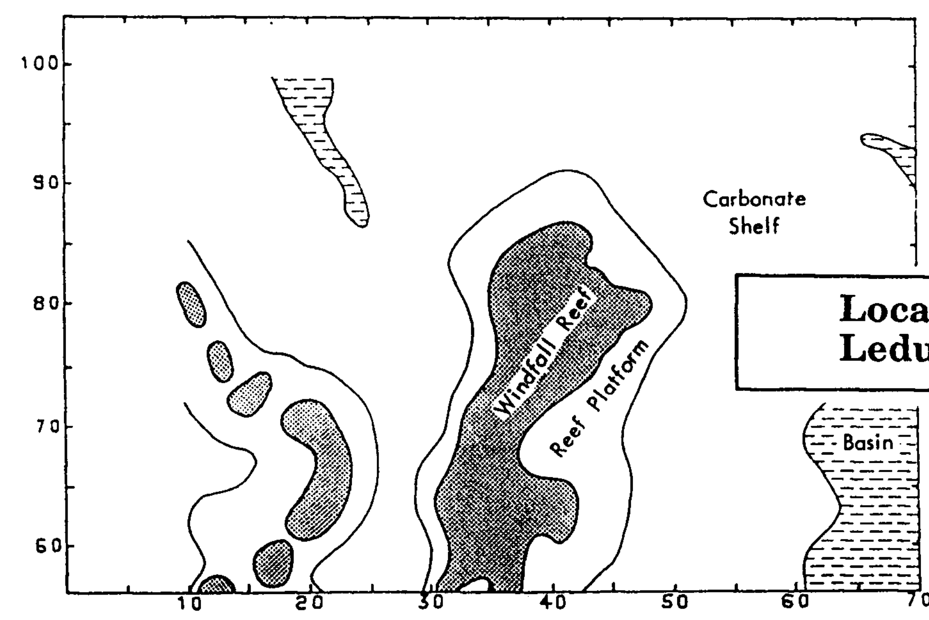
Structural elevation of Basal Fish Scales



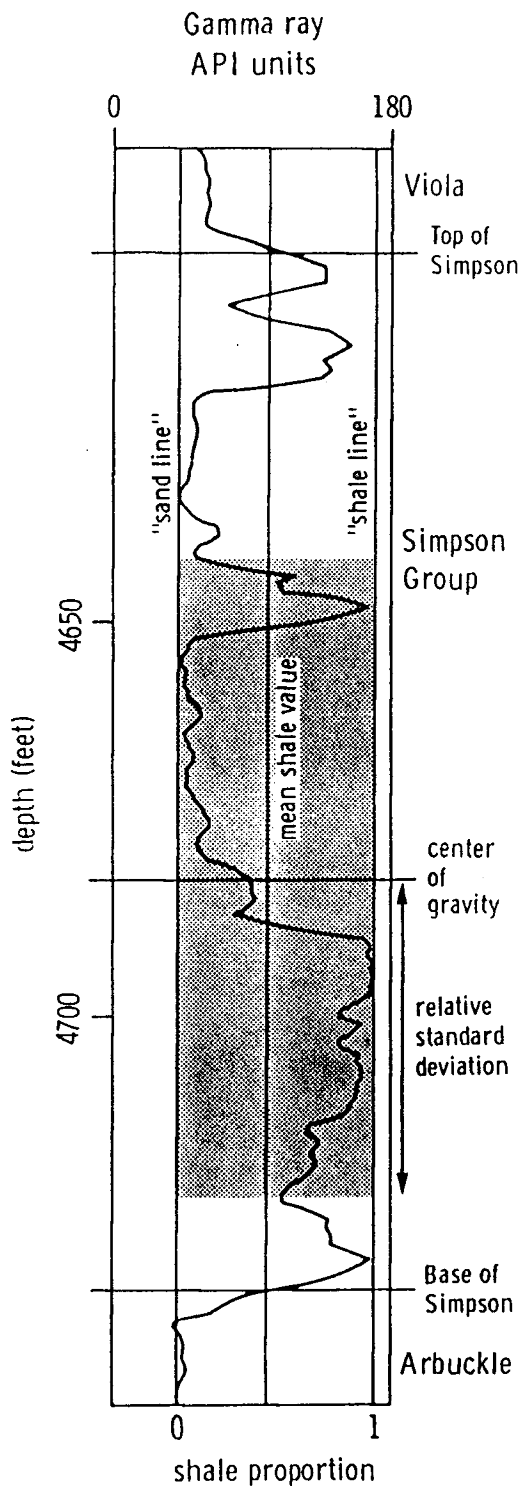
Second degree trend surface of BFS structure



Second degree trend residuals of BFS structure



Location of known Leduc reefs



MOMENTS	
$v_1 = \Sigma Sd / \Sigma S$	center of gravity
$v_2 = \Sigma Sd^2 / \Sigma S$	dispersion
$v_3 = \Sigma Sd^3 / \Sigma S$	skewness
$v_4 = \Sigma Sd^4 / \Sigma S$	kurtosis

$v_m = \Sigma Sd^m / \Sigma S$	

POLYNOMIAL REGRESSION

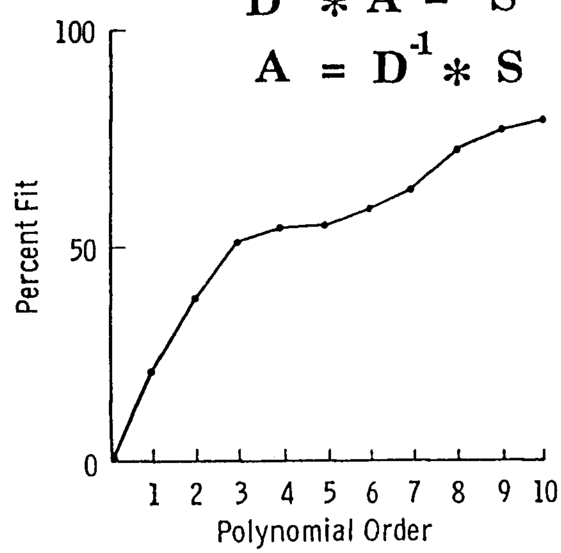
$$\hat{S} = a + bd + cd^2 + \ldots d^m$$

matrix solution

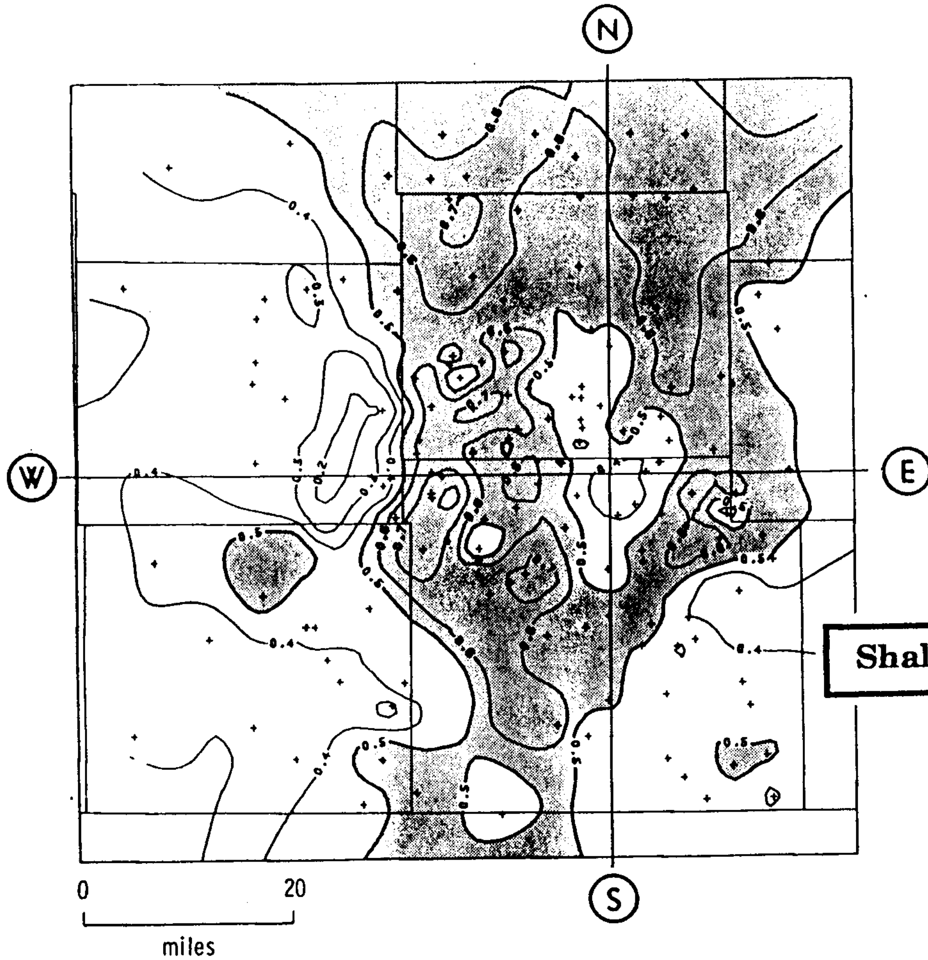
$$\begin{bmatrix} n & \Sigma d & \Sigma d^2 & . & \Sigma d^m \\ \Sigma d & \Sigma d^2 & . & . & . \\ \Sigma d^2 & . & . & . & . \\ . & . & . & . & . \\ \Sigma d^m & . & . & . & \Sigma d^{2m} \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ . \\ . \end{bmatrix} = \begin{bmatrix} \Sigma S \\ \Sigma Sd \\ \Sigma Sd^2 \\ . \\ \Sigma Sd^m \end{bmatrix}$$

$$D * A = S$$

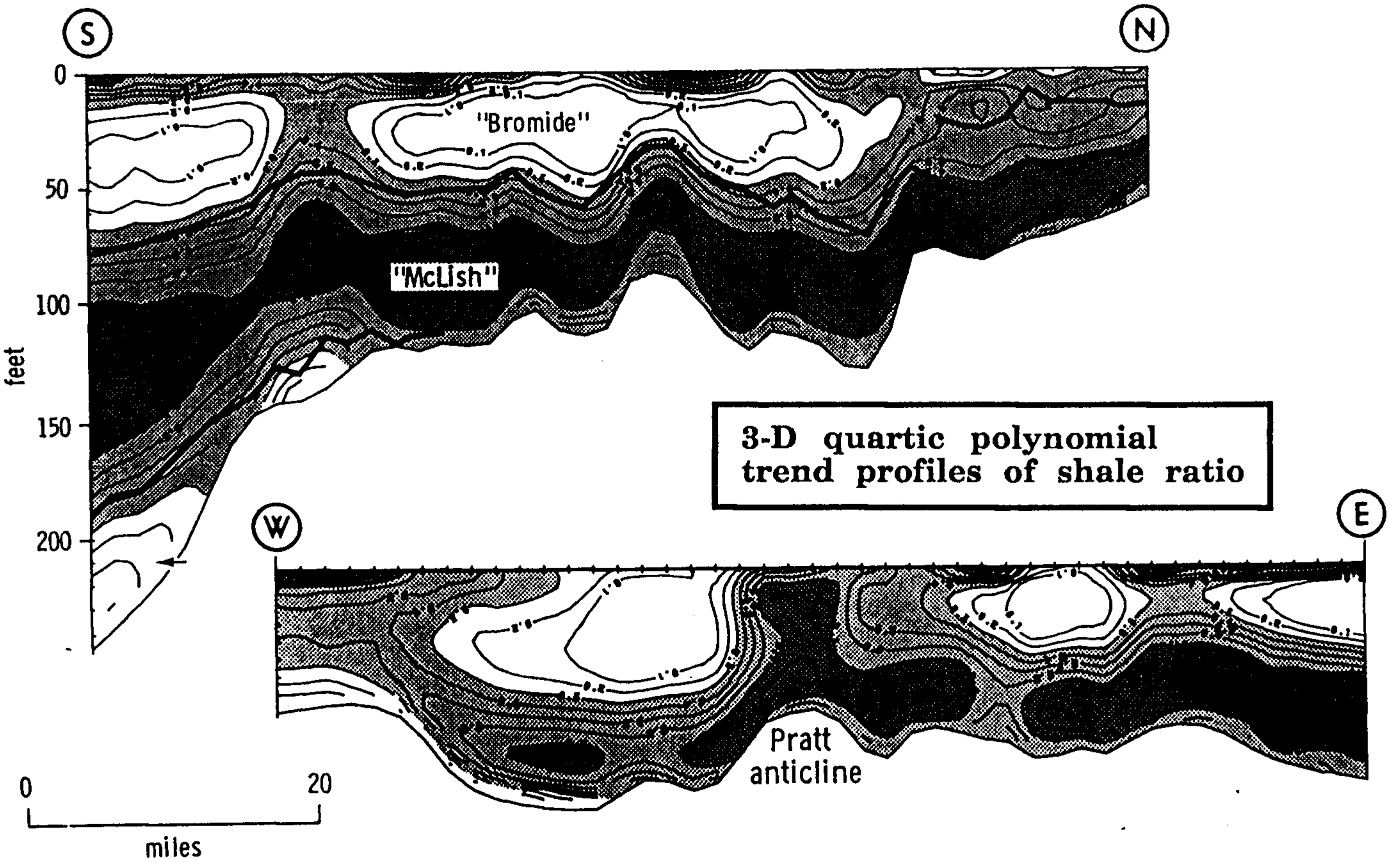
$$A = D^{-1} * S$$



INTERPOLATION OF GAMMA-RAY SHALE MOMENTS BETWEEN WELLS TO GENERATE THREE-DIMENSIONAL TREND MODEL



Shale ratio map



3-D quartic polynomial trend profiles of shale ratio

Reference: Doveton, Zhu, and Davis (1984)

FILTERS

A time series, such as a log, is the INPUT, which when CONVOLVED by a FILTER, generates another time series, which is the OUTPUT :

$$\underset{\textbf{output}}{[O]} = \underset{\textbf{input}}{[I]} \underset{\textbf{convolution operator}}{*} \underset{\textbf{filter}}{[f]}$$

$$\text{if} \ldots [I] = \begin{bmatrix} i_0 & i_1 \end{bmatrix}$$

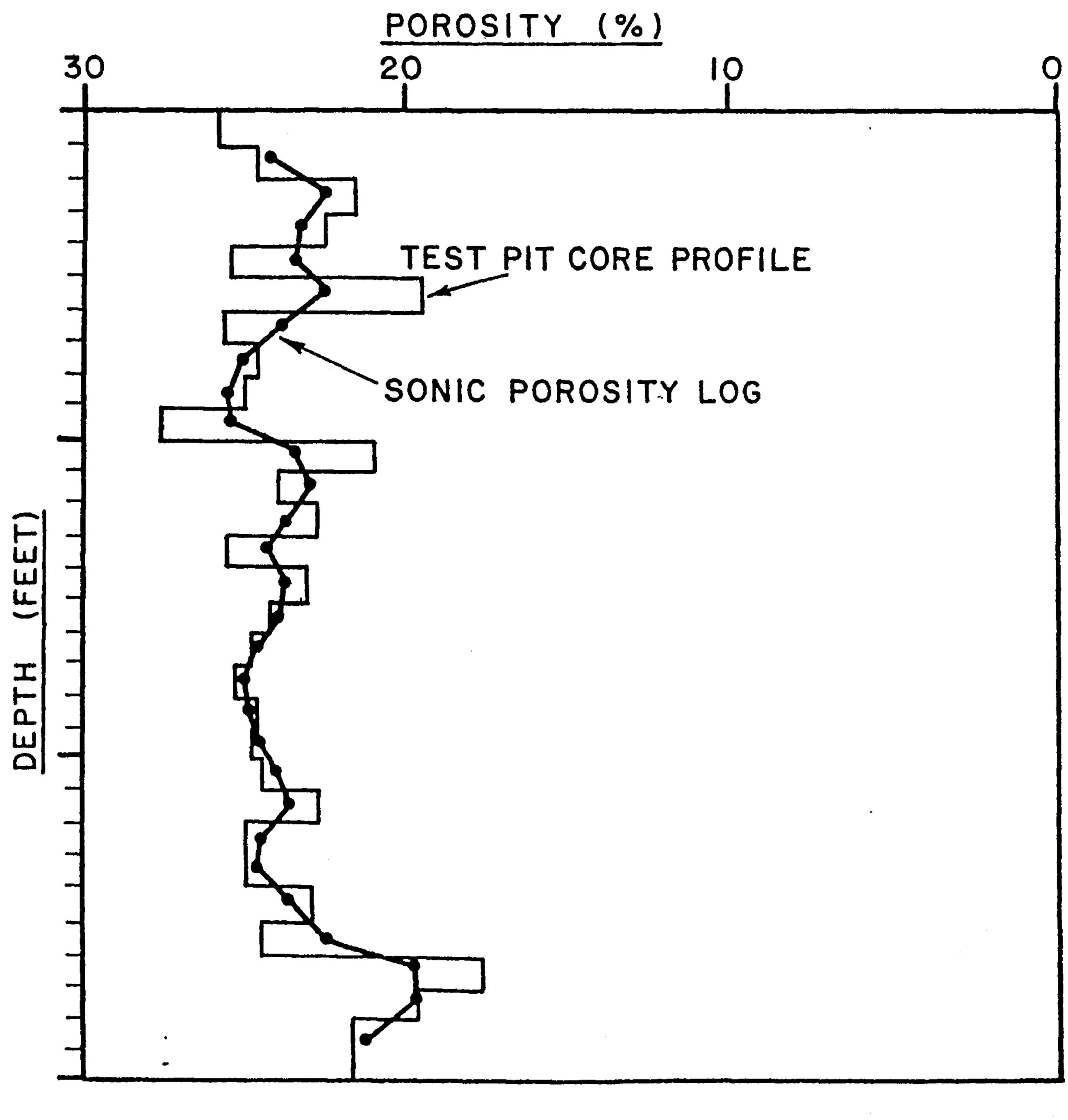
$$\text{and} \ldots [f] = \begin{bmatrix} f_0 & f_1 \end{bmatrix}$$

$$\text{then} \ldots [O] = \begin{bmatrix} f_0 i_0 & f_0 i_1 + f_1 i_0 & f_1 i_1 \end{bmatrix}$$

The most common application of filters in log analysis is to SMOOTH either core or log measurements to ensure compatibilty in common vertical resolution. So, for example, if the transit times recorded by a sonic tool with two - foot span are to be porosity calibrated with core measurements at one - foot sampling, then the core data should be smoothed. Ideally, this would be a (0.5, 0.5) operator, but the output would be located "between" depth increments, so the operator (0.25, 0.5, 0.25) is used (commonly known as the "1-2-1" filter).

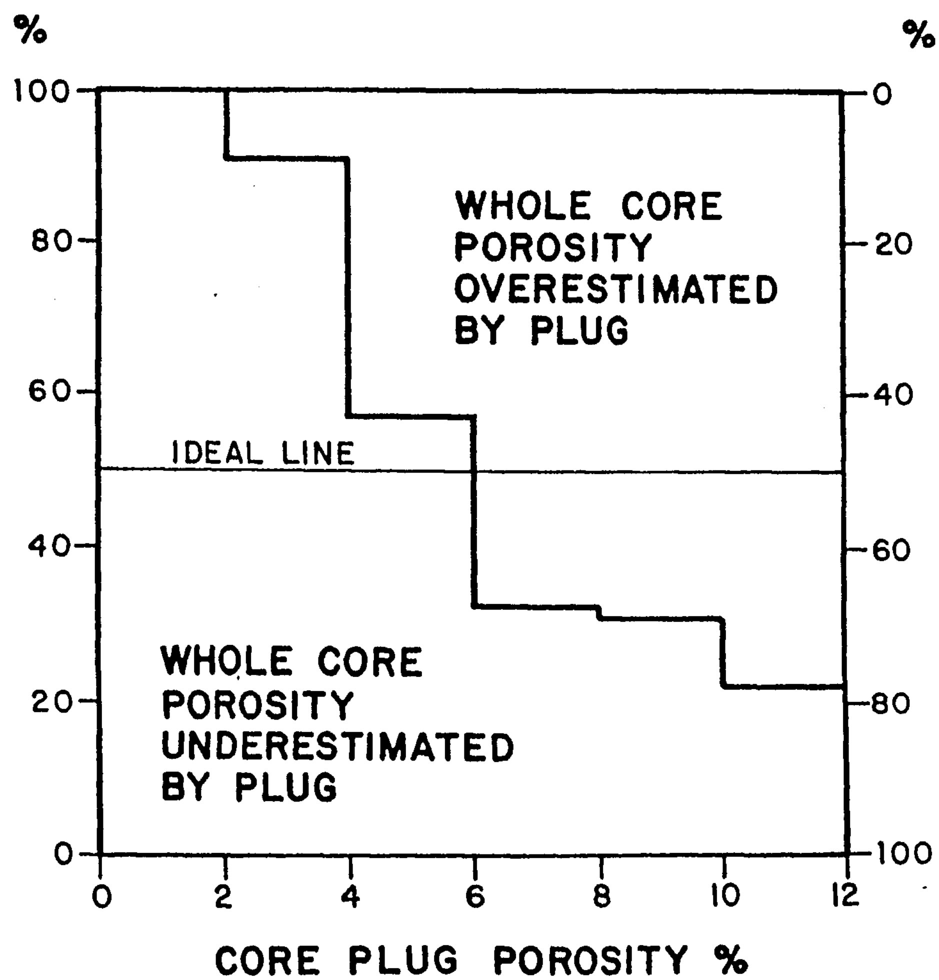
Other applications include "despiking" sonic logs and the computation of first difference (1,-1) and second difference (1, -2, 1) logs.

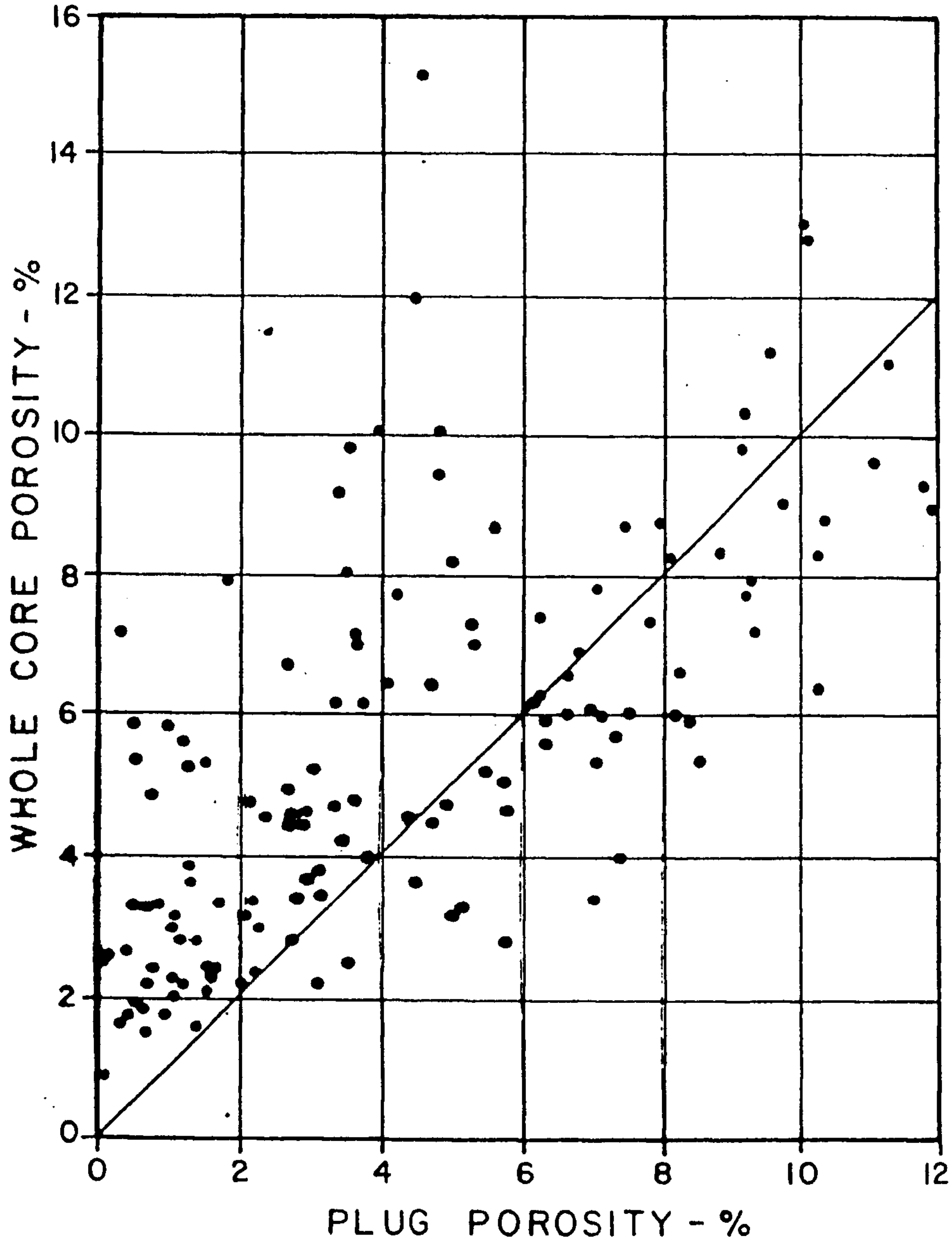
COMPARISON BETWEEN A SEQUENCE OF CORE POROSITIES (ONE FOOT SAMPLING) FROM A SANDSTONE RESERVOIR AND THE POROSITY RESPONSE THAT WOULD BE EXPECTED FROM A SONIC LOGGING TOOL WITH TWO - FOOT SPAN. THE "SONIC LOG" WAS COMPUTED BY CONVOLVING THE CORE POROSITIES WITH A "1-2-1" FILTER



EXAMPLE : DEMONSTRATION OF WHY SMOOTHING IS NECESSARY TO ENSURE COMPATIBILITY BETWEEN CORE AND LOG MEASUREMENTS

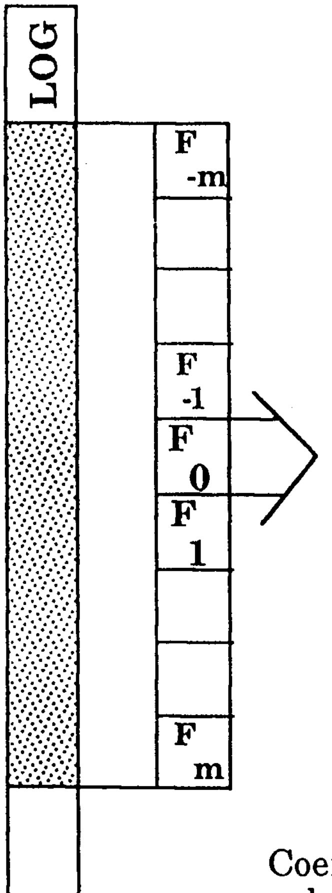
The data shown on these two pages are reported by Baker (1957) and contrast porosities from whole core (6 to 12 inch samples) with plugs (1 inch diameter, 1 inch length) from the same core. Although the data are recorded at one - foot spacing, their character is systematically different. Unless the rock is completely homogeneous within one - foot increments, it is INEVITABLE that the variance of plug porosities will be greater than whole core samples. This will be true regardless of whether the plugs were chosen randomly or as most representative. The plot of one against the other shows the phenomena of both ERROR and BIAS.
This demonstrates the need for smoothing to be used when logging tools or core samples have differing resolutions, if accurate, unbiased results are desired.
Notice that the smoothing of core plug data is conservative (the data will not be "oversmoothed") because the plug variance will be greater than whole core.





CROSSPLOT OF WHOLE - CORE POROSITIES VERSUS POROSITIES FROM PLUGS "SELECTED TO BE REPRESENTATIVE" OF THE SAME CORE. DATA FROM BAKER (1957)

POLYNOMIAL REGRESSION FILTERS



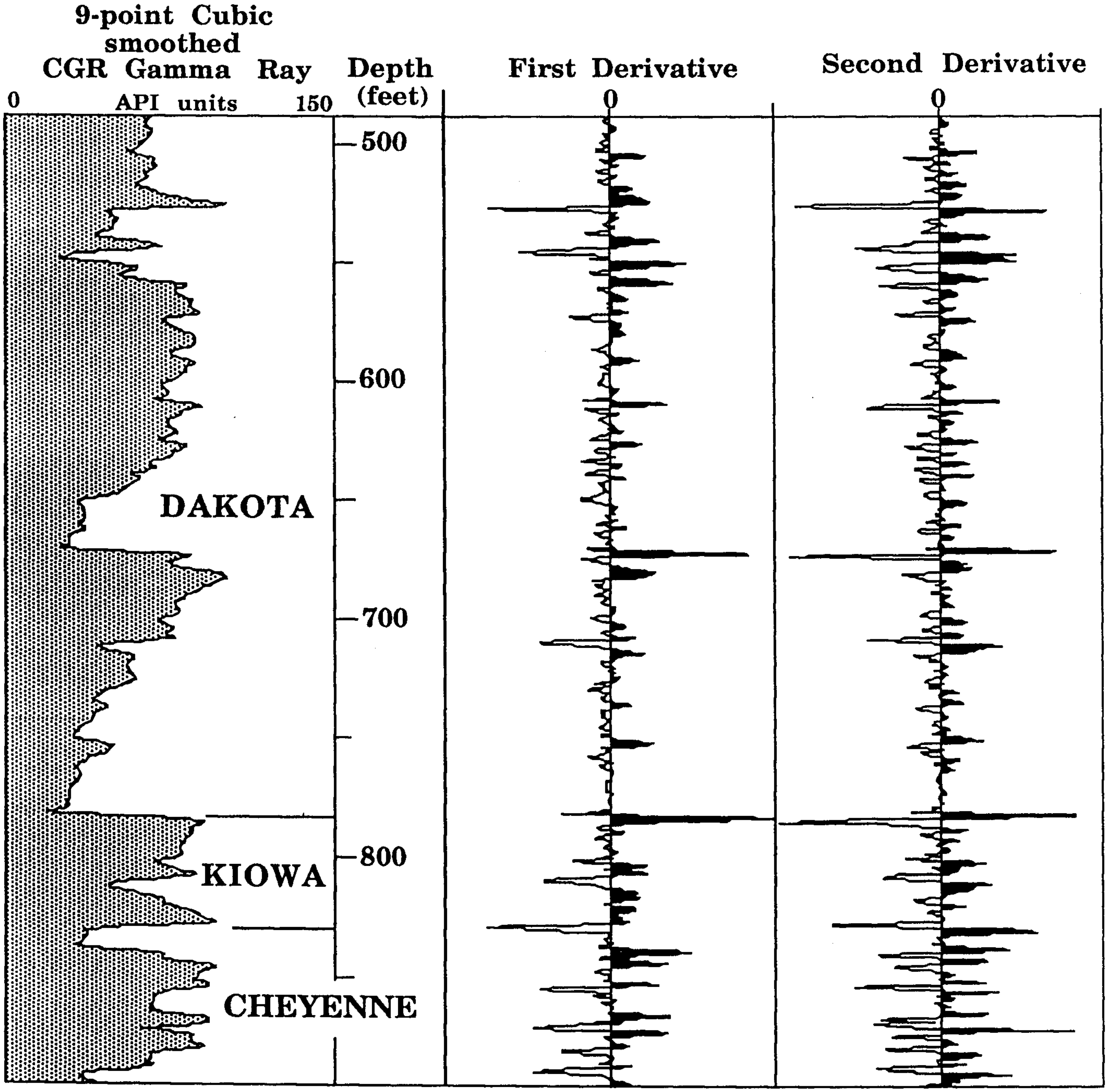
Filters can be designed to make a "running" polynomial regression of the portion of a log matched by the filter as it is moved down the length of the log. The process does not require the procedure described earlier, because the task has some simplifying and helpful features :
(1) the log is digitized at fixed increment, so that the observations of independent variable, X, are spaced equally --- this leads to elimination of some elements of the matrix solution :
(2) the desired output is the polynomial trend characteristic at the midpoint of the filter window, which can be standardized as "zero depth", further simplifying the solution.

Coefficients can be calculated as loadings in polynomial filters to generate either SMOOTHED TREND, FIRST (slope), or SECOND (rate of change of slope) DERIVATIVES of logs.

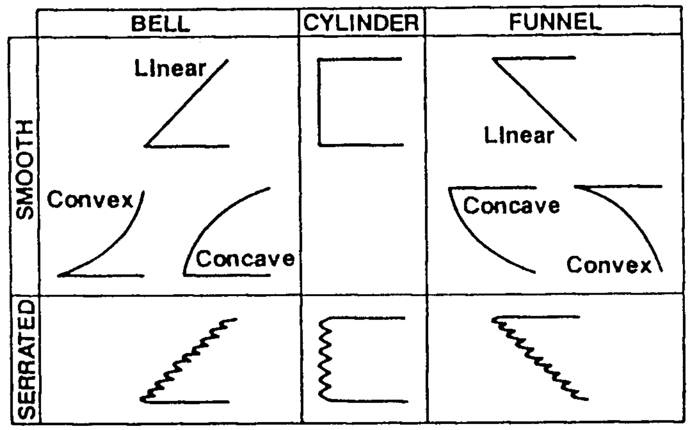
NINE-POINT CUBIC POLYNOMIAL REGRESSION FILTER

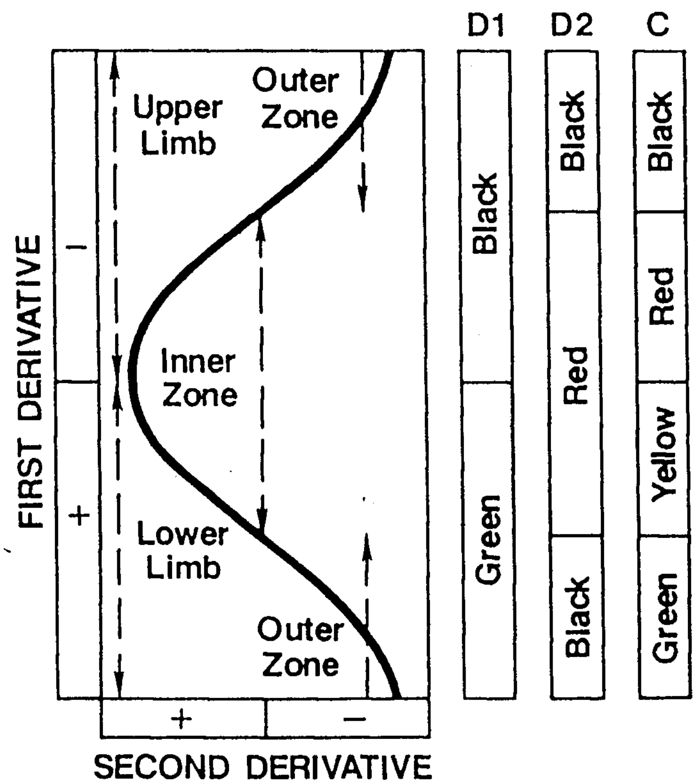
Filter element	Smoothed estimate	First derivative	Second derivative
1	-21	86	28
2	14	-142	7
3	39	-193	-8
4	54	-126	-17
5	59	0	-20
6	54	129	-17
7	39	193	-8
8	14	142	7
9	-21	-86	28
Normalizing factor	231	1188	462

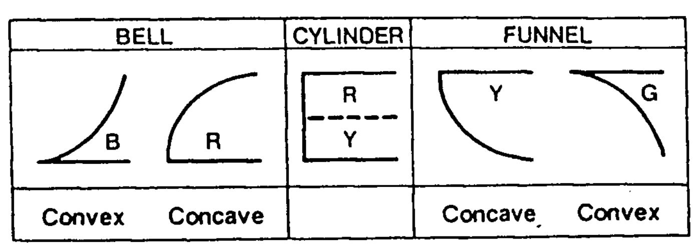
EXAMPLE : SMOOTHED, FIRST DERIVATIVE (SLOPE), AND SECOND DERIVATIVE (RATE OF CHANGE OF SLOPE) OF CGR GAMMA RAY, USING A CUBIC POLYNOMIAL 9 -POINT (4.5 FOOT WINDOW) FILTER



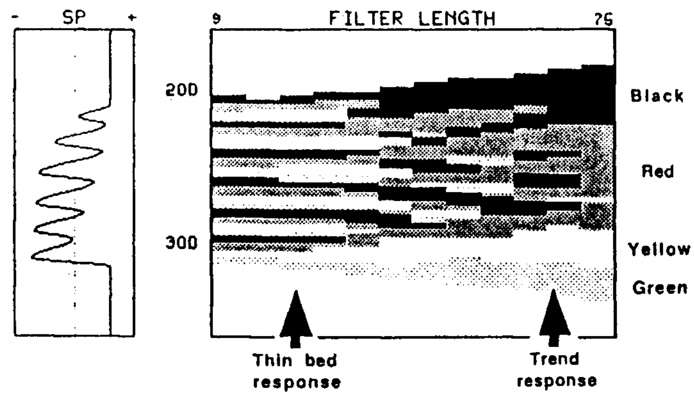
TRANSFORMATION OF LOG SHAPES INTO COLOUR THROUGH THE USE OF THE FIRST DERIVATIVE (D1) AS GREEN, AND THE SECOND DERIVATIVE (D2) AS RED



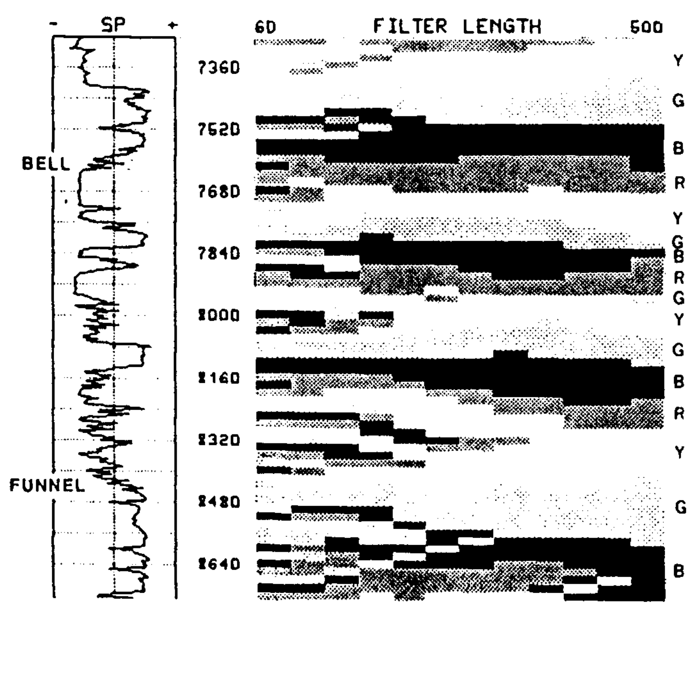




COLOUR PROFILES OF SHAPE GENERATED AT DIFFERENT VERTICAL SCALES THROUGH THE OPERATION OF CUBIC POLYNOMIAL FILTERS TO GENERATE FIRST AND SECOND DERIVATIVES AT DIFFERING WINDOW LENGTHS



Composite colour image from derivatives of a synthetic bell - shaped trend of thin sandstone beds

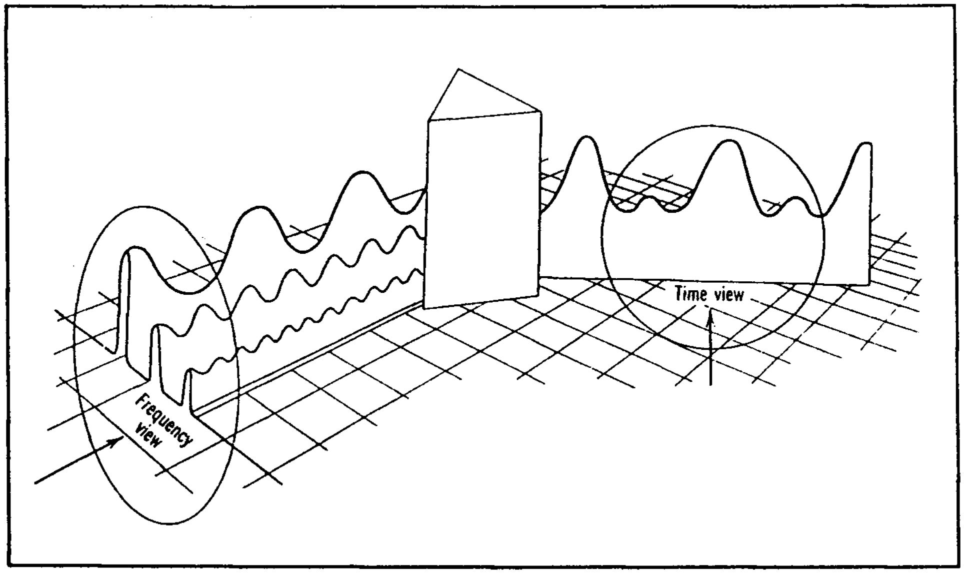


Composite colour derivative image of an SP log in a Tertiary deltaic succession from a well in Louisiana

(from Collins and Doveton, 1988)

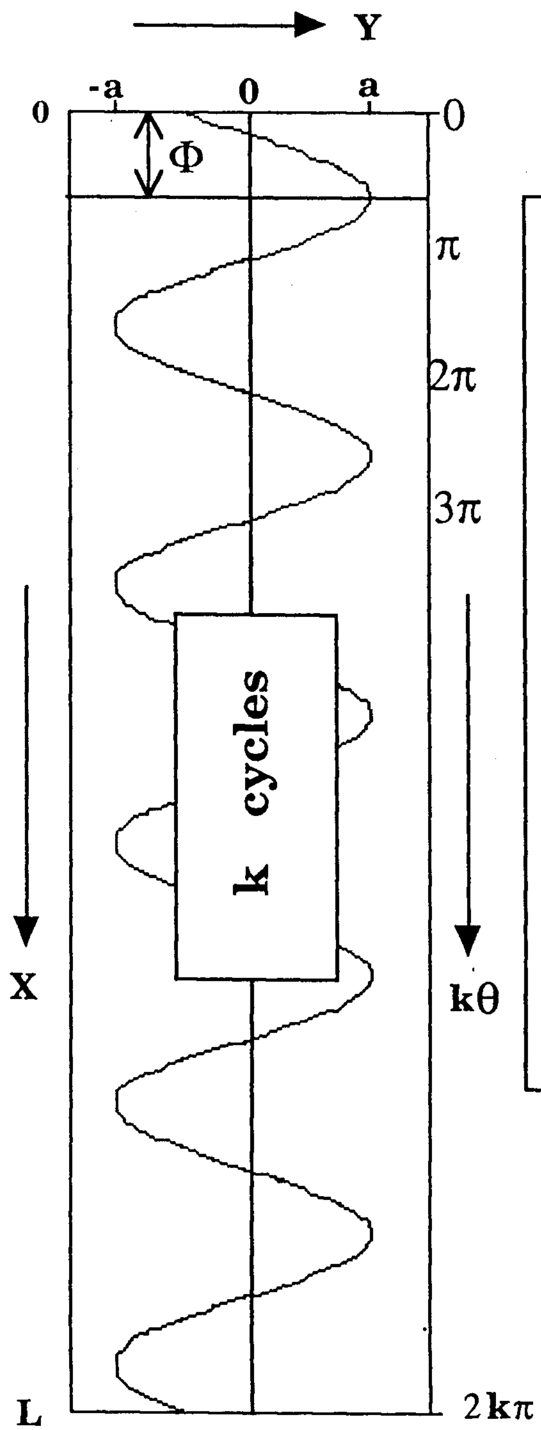
Blank

FOURIER ANALYSIS



FOURIER ANALYSIS

The representation of a time (or depth) record of a variable, Y by a summation of sinusoidal waves of different frequencies.



Conversion of X from time units to angular measure :

$$\theta = \left(\frac{2\pi}{L}\right) \times X$$

CALCULATION OF THE kth HARMONIC :

For k cycles with an initial "shift" of phase angle, Φ :

$$Y = a.\cos(k\theta - \Phi)$$
$$= a.\cos\Phi \cos k\theta + a.\sin\Phi \sin k\theta$$
$$= A \cos k\theta + B \sin k\theta$$

Phase angle :

$$\frac{B}{A} = \frac{a.\sin\Phi}{a.\cos\Phi} = \tan\Phi$$
$$\therefore \text{phase angle}, \Phi = \tan^{-1}\left(\frac{B}{A}\right)$$

Power or variance:

$$A^2 + B^2 = a^2 \sin^2\theta + a^2 \cos^2\theta = a^2$$
$$\therefore \text{amplitude of wave} = \sqrt{A^2 + B^2}$$
$$\text{power}, s^2 = A^2 + B^2$$

FOURIER SERIES :

The complete set of harmonics is a Fourier series :

$$Y_i = \sum_{k=0}^{n/2} A_k \cos k\theta + B_k \sin k\theta$$

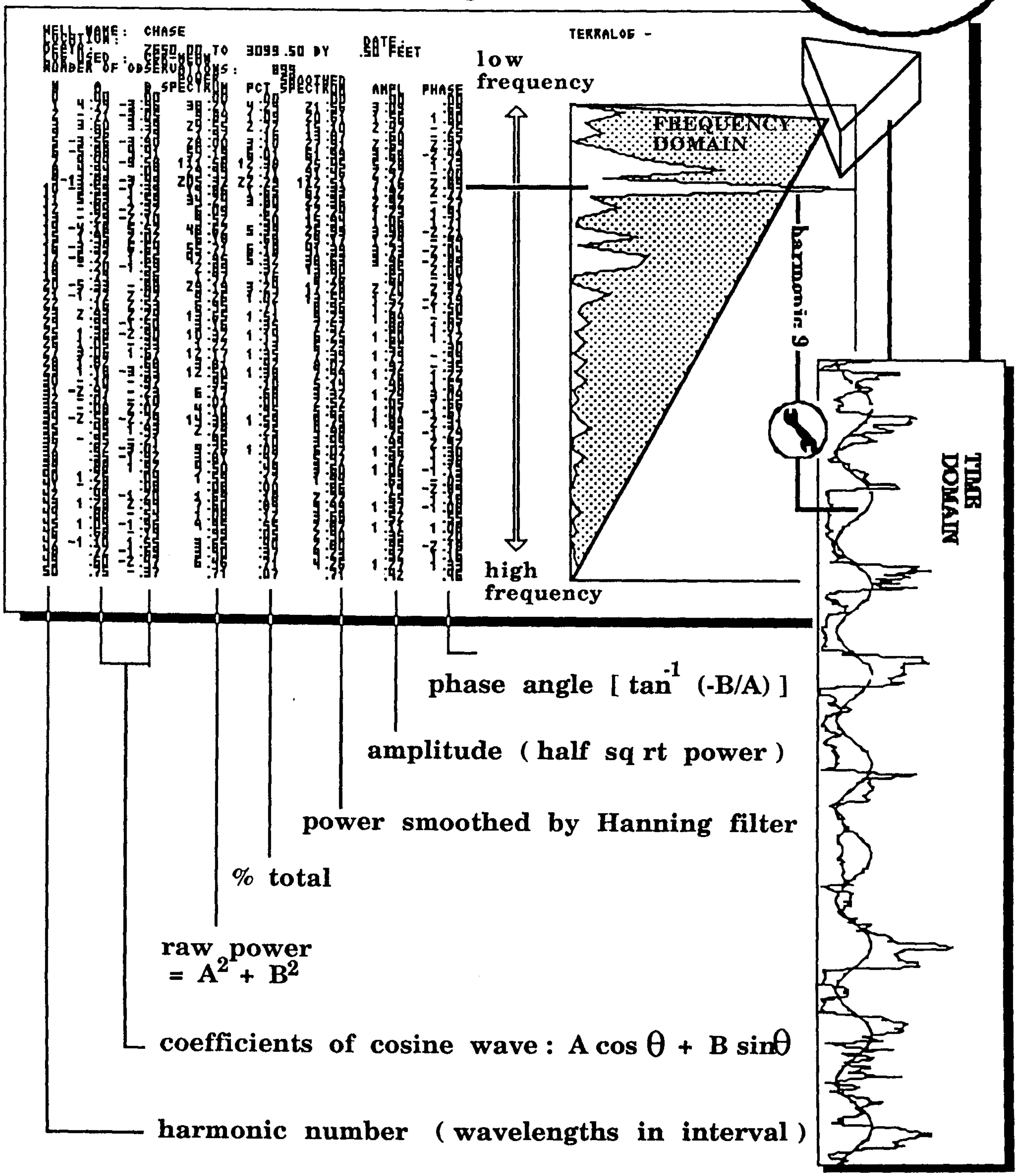
where n is the number of data points and n/2 is the Nyquist frequency

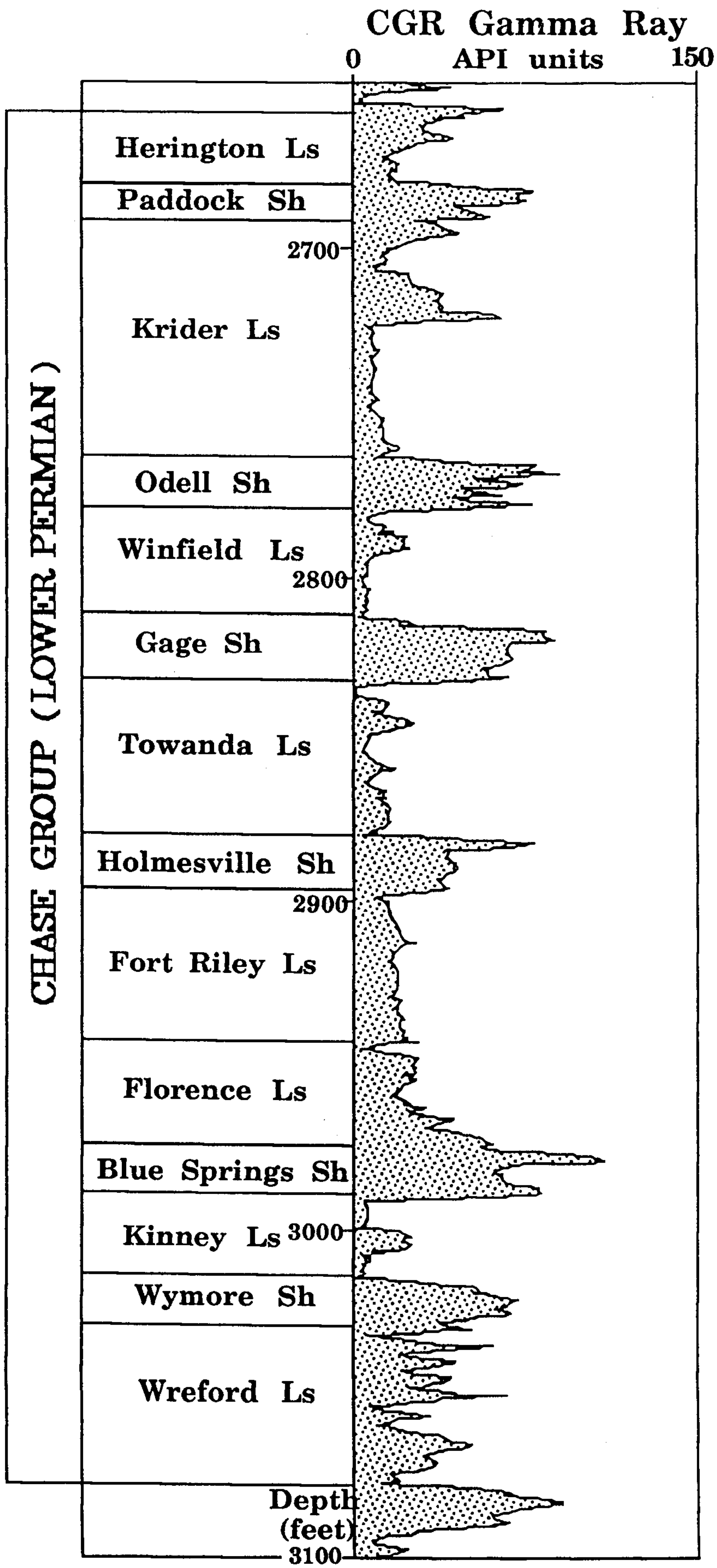
FOURIER ANALYSIS

FOU

INPUT
One log in interval set by depth range

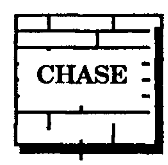
COMPUTES : Discrete power spectrum (periodogram) for harmonic range





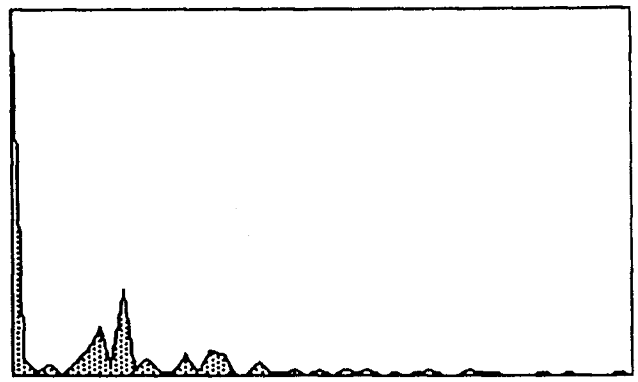
FOURIER ANALYSIS

Example: Evaluation of potential cyclic repetitions in the carbonate - shale composition of the Chase Group



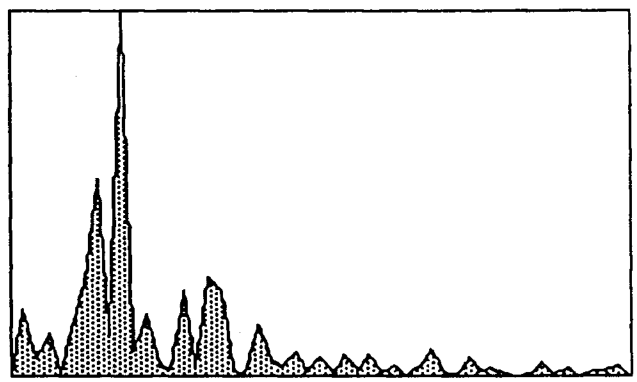
POWER SPECTRA OF CGR GAMMA RAY LOG IN CHASE GROUP

CGR RAW DATA



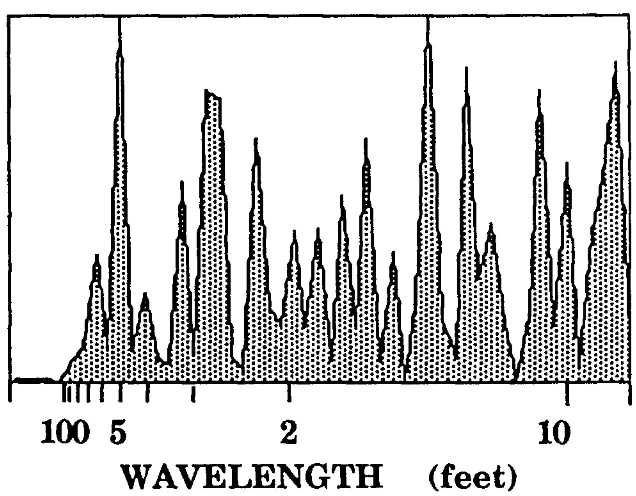
CGR - CGR mean

By standardizing the data, the contribution of the zero harmonic is removed.

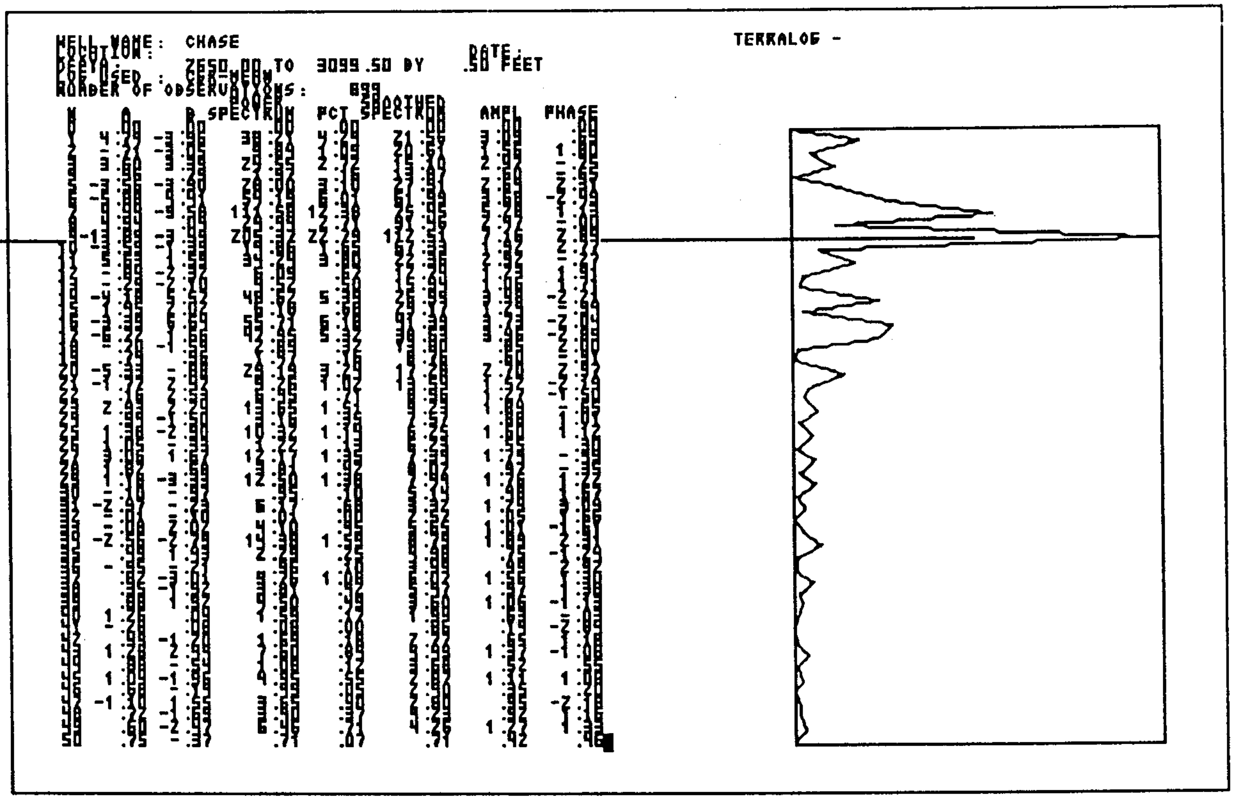


CGR FIRST DIFFERENCE

The use of the first difference results in the removal of long-term trends in the data. The first difference log can be generated with a FILTER and the elements [-1,1]



FOURIER POWER SPECTRUM OF VARIABLE STANDARDIZED TO ZERO MEAN



Harmonic number (h) = 9 **A = -13.89 B = -3.49**

Number of samples (n) = 899 **Depth sampling rate (s) = 0.5 feet**

Depth range (D_0 - D_n) = 2650 - 3099.5 feet

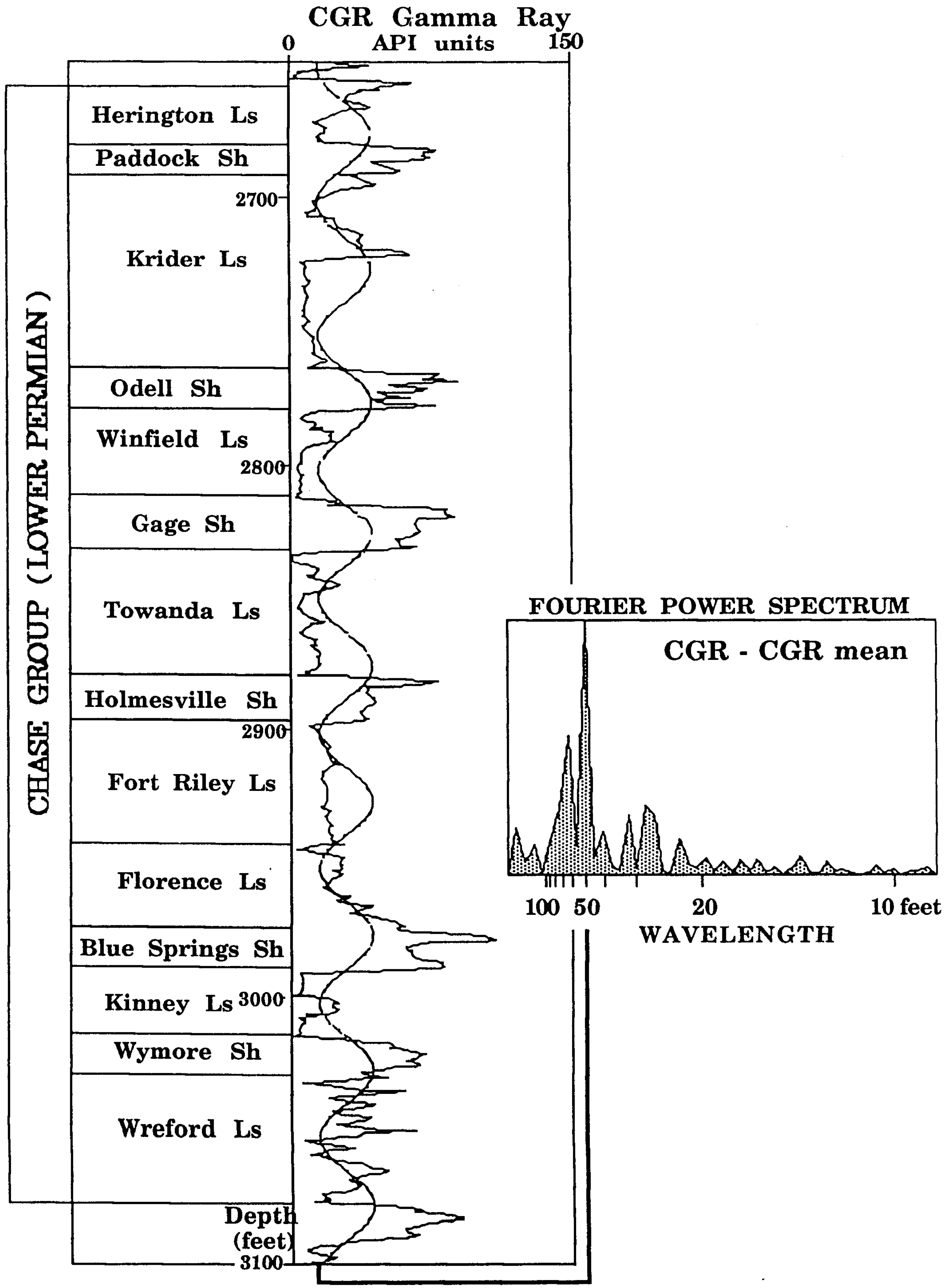
For harmonic h, the depth range is equivalent to h periods of 360 degrees.
The transformation from depth (D) to degrees is given by:

$$\theta = ((D - D_0) * h * 360) / (n * s)$$

When h = 9, $\theta = 7.208\,D - 19101.2$
The log playback of the harmonic is then :

$$W = B \sin\theta + A \cos\theta$$

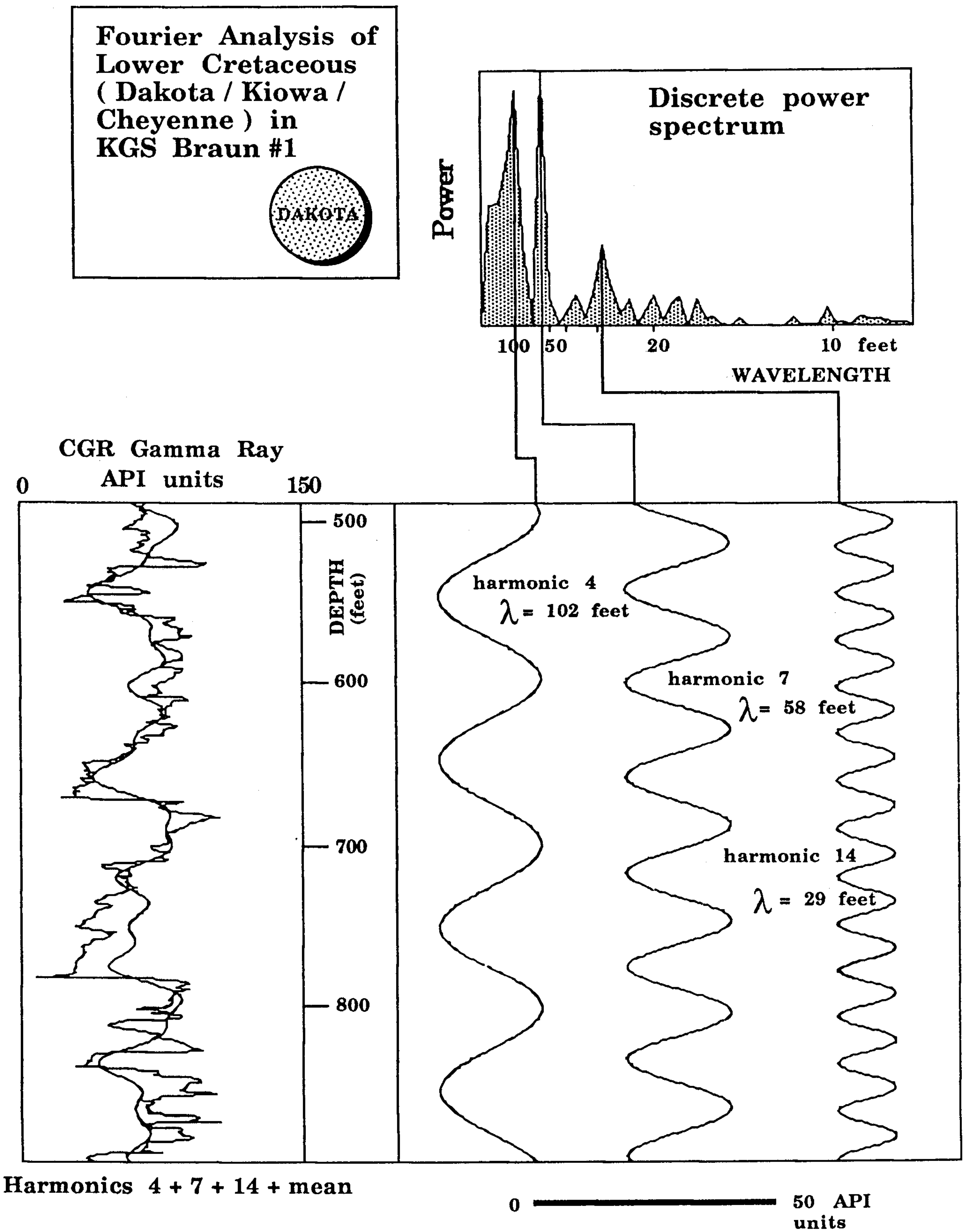
For useful comparison with the original log, the mean value should be added. (The mean value corresponds to the zeroth harmonic.)



CGR Gamma Ray
0 API units 150
CHASE GROUP (LOWER PERMIAN)
Herington Ls
Paddock Sh
2700
Krider Ls
Odell Sh
Winfield Ls
2800
Gage Sh
Towanda Ls
Holmesville Sh
2900
Fort Riley Ls
Florence Ls
Blue Springs Sh
Kinney Ls
3000
Wymore Sh
Wreford Ls
Depth (feet)
3100
FOURIER POWER SPECTRUM
CGR - CGR mean
100 50 20 10 feet
WAVELENGTH

EFFECT OF WAVE SHAPE ON FOURIER ANALYSIS

The Fourier transform converts the original time record into a frequency domain of SINE WAVES. However, the record may be strongly non-sinusoidal, but cyclic, such as can occur with square waves or repetitive saw - tooth features.
In these cases, the power spectrum will often show integer multiple harmonics of the fundamental harmonic, which picks up the basic period.



Fourier Analysis of Lower Cretaceous (Dakota / Kiowa / Cheyenne) in KGS Braun #1
DAKOTA
Discrete power spectrum
Power
100
50
20
10 feet
WAVELENGTH
CGR Gamma Ray
API units
0
150
500
600
700
800
DEPTH (feet)
harmonic 4
λ = 102 feet
harmonic 7
λ = 58 feet
harmonic 14
λ = 29 feet
Harmonics 4 + 7 + 14 + mean
0
50 API units

Blank

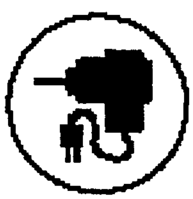
BOTTOM-UP METHODS :

Statistical power tools to locate electrofacies

SUPERVISED :

Discriminant function analysis

UNSUPERVISED :



Principal component analysis

Factor analysis

Cluster analysis

Blank

DEDUCTION

Top-down programming

Petrography

CV = L

INDUCTION

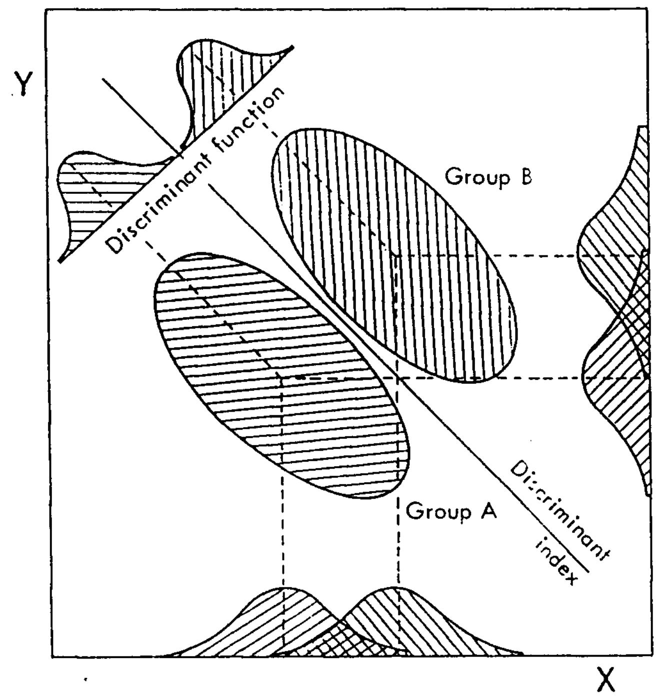
Bottom-up programming

Facies

UNSUPERVISED	Principal Component Analysis	Cluster Analysis
SUPERVISED	Discriminant Function Analysis	

LINEAR DISCRIMINANT FUNCTION ANALYSIS

A SUPERVISED statistical classification technique. A linear discriminant function is computed which best distinguishes two groups of known assignment on the basis of several observational variables. The function is based on the multivariate means of the two groups (their cloud centroids in multivariate space) and their covariance matrices (the clouds' dispersions or "shapes" in multivariate space). The function is located such that the distance between the group data clouds is maximized while, simultaneously the clouds' dispersion is minimized. After calculation of the function, multivariate observations of unknown assignment may be classified as to membership of one or other of the two groups.



Discriminant function : $Z = \lambda_1 X_1 + \lambda_2 X_2 + \dots \dots \lambda_m X_m$

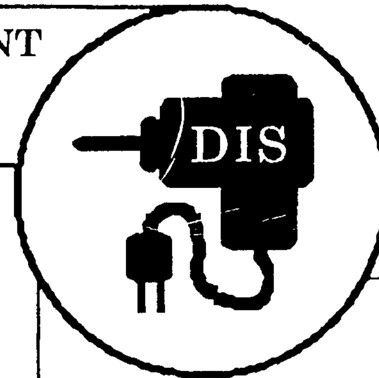
The MAHALANOBIS DISTANCE, D is the generalized distance between the two groups : $D^2 = \lambda_1(\bar{X}1_1 - \bar{X}2_1) + \dots \lambda_m(\bar{X}1_m - \bar{X}2_m)$

An F ratio is calculated from :

$$F = \left[\frac{n_1 + n_2 - m - 1}{(n_1 + n_2 - 2)m}\right]\left[\frac{n_1 n_2}{n_1 + n_2}\right]D^2$$

and the null hypothesis that the two group centroids are equal is evaluated as an F- test with m and (n + n - m -1) degrees of freedom. If judged significant, the function can be used for classification.

DISCRIMINANT FUNCTION ANALYSIS



INPUT:
Two groups
Group interval depth ranges
Log variables

COMPUTES: Discriminant function

$$Z = \lambda_1 X_1 + \lambda_2 X_2 + \lambda_3 X_3 + \ldots\ldots$$

Fine Print:
Parametric method
Normal distribution
Groups have equal variance / covariance matrices.

```
WELL NAME:  CHASE
LOCATION:
DATE:

GROUP A      29 OBSERVATIONS
STARTING DEPTH   2786.000          ENDING DEPTH   2800.000

GROUP B      22 OBSERVATIONS
STARTING DEPTH   2800.000          ENDING DEPTH   2811.000
```

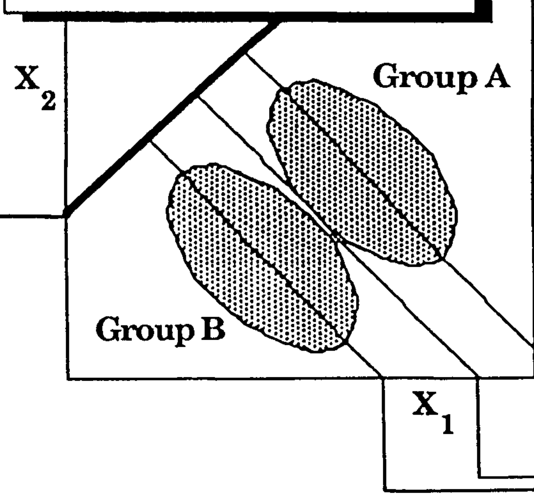
λ

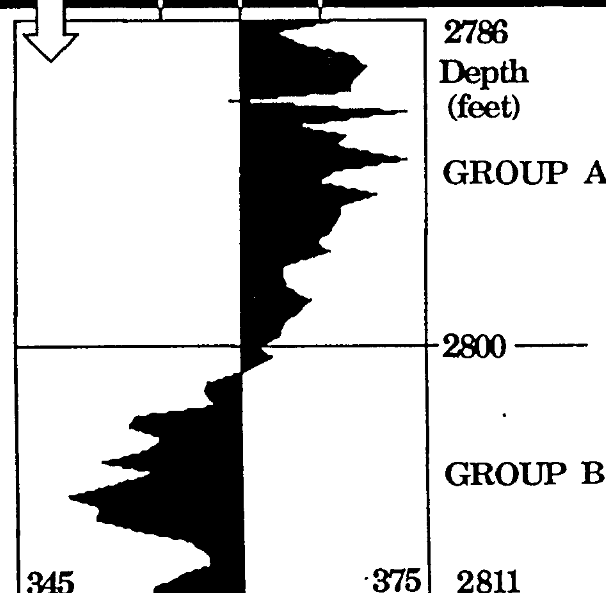
	VECTOR MEAN OF GROUP A	VECTOR MEAN OF GROUP B	DISCRIMINANT COEFFICIENT	RELATIVE CONTRIBUTION
RHOMAA	2.798	2.743	131.896	.611
UMAA	8.907	8.170	-1.987	-.125
CNL%	12.656	13.235	.330	-.016
TH	1.512	.812	5.915	.354
UR	1.050	1.506	-.579	.022
K	.606	.295	5.750	.153

```
DISCRIMINANT INDEX
    GROUP A        367.3459
    TOTAL GROUP    361.4985
    GROUP B        355.6523

STATISTICS
    MAHALANOBIS DISTANCE     11.6937
    T-TEST                  146.2863
    F-TEST                   21.8931
    DEGREES OF FREEDOM       6 AND     44
COMPUTE DISCRIMINANT SCORES - YES OR NO .... Y
```

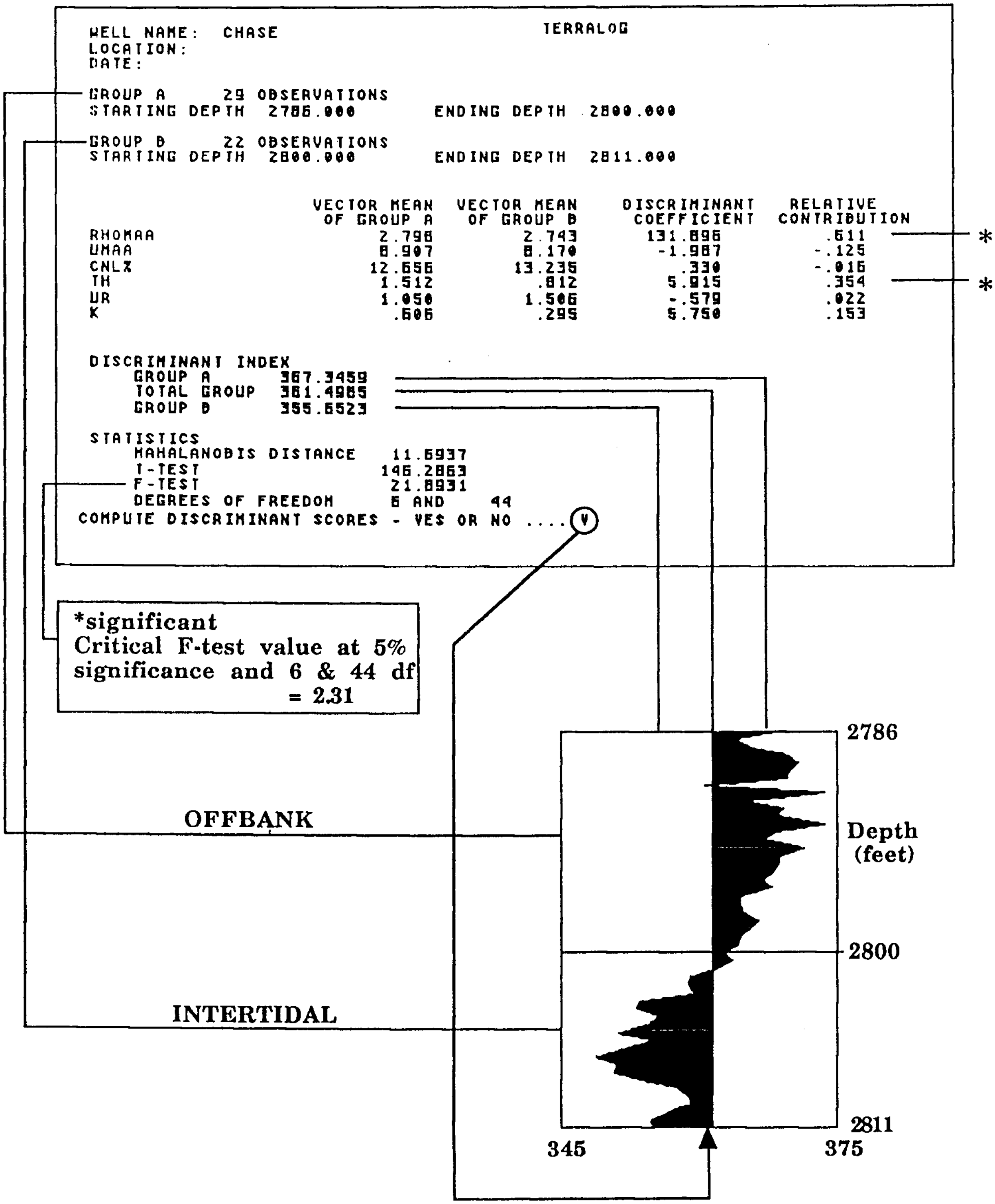
Group multivariate means

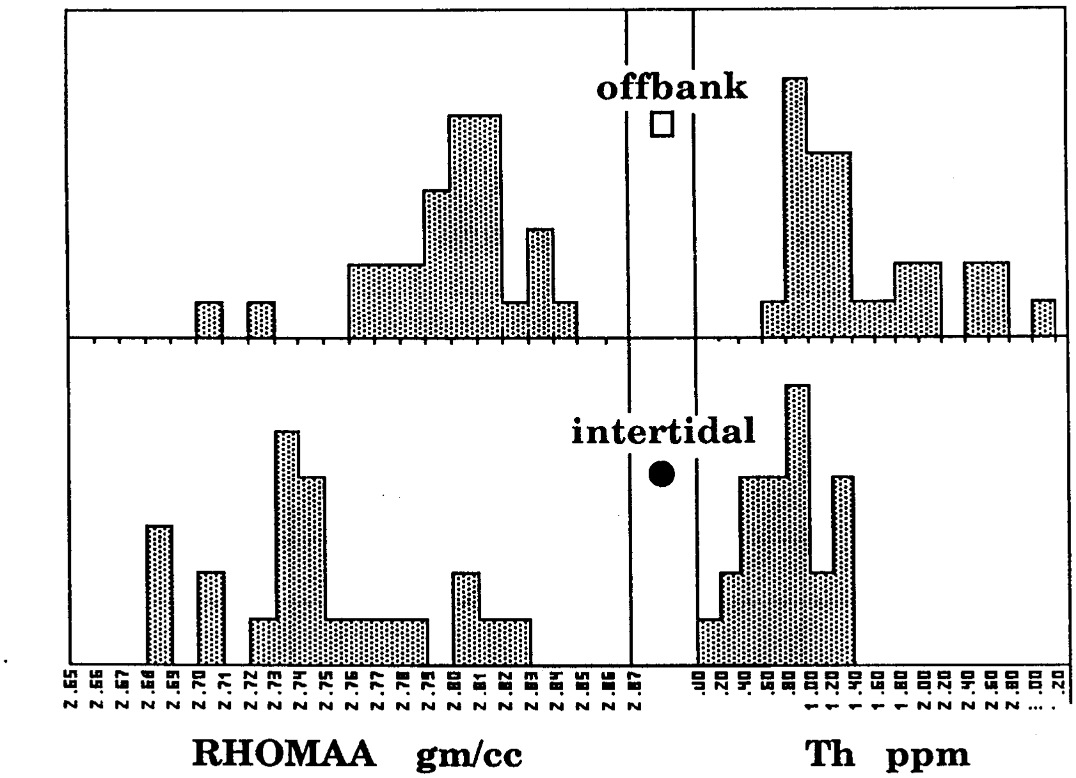
SIGNIFICANCE TEST:
*significant
Critical F-test value at 5% significance and 6 & 44 df = 2.31

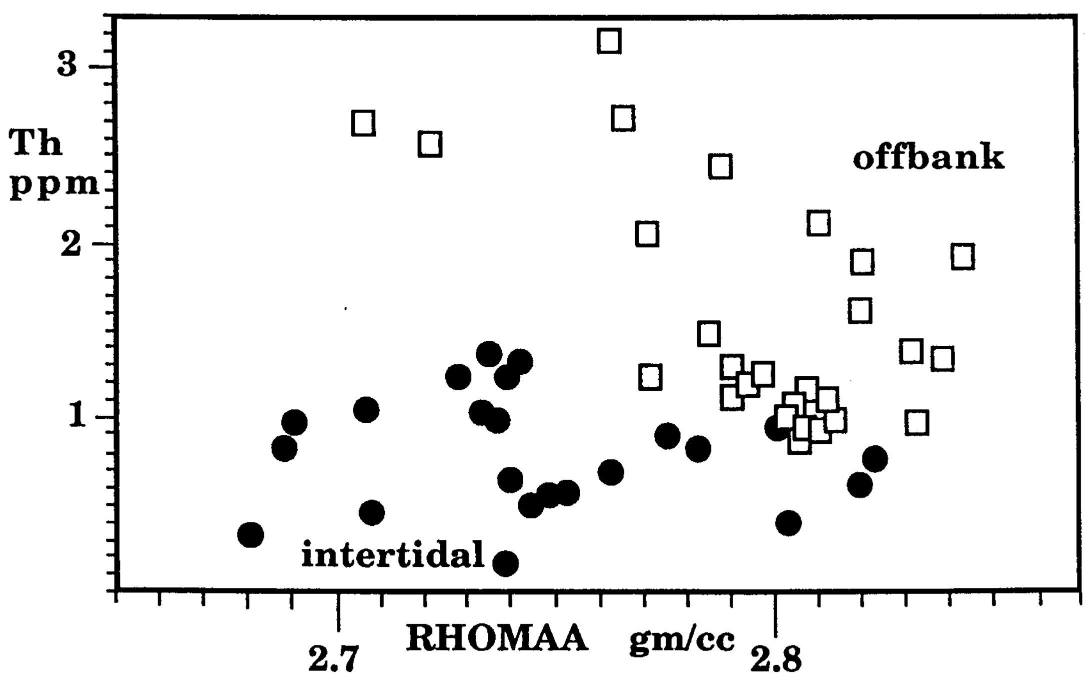




DISCRIMINATION BETWEEN OFFBANK AND INTERTIDAL CARBONATE FACIES IN THE WINFIELD LIMESTONE BASED ON 6 LOG VARIABLES

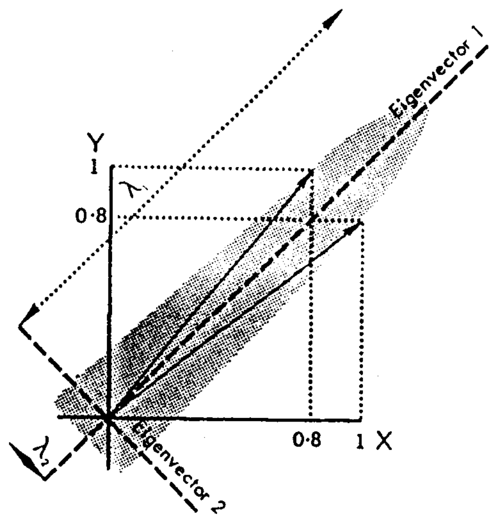



offbank
intertidal
RHOMAA gm/cc
Th ppm


3
Th
ppm
2
1
offbank
intertidal
2.7
RHOMAA gm/cc
2.8

PRINCIPAL COMPONENT ANALYSIS

An UNSUPERVISED multivariate statistical technique. The principal components are the EIGENVECTORS of the covariance or correlation (standardized covariance) matrix. The eigenvectors are the principal axes of the natural variation of the data cloud which is represented by the covariance matrix as a hyperellipsoid whose centroid is the coordinates of the multivariate mean. These principal components are ordered in importance according to their EIGENVALUES which represent their relative degree of "stretch". Each eigenvalue divided by their total sum is the proportion of the total variation accounted for by the associated principal component.

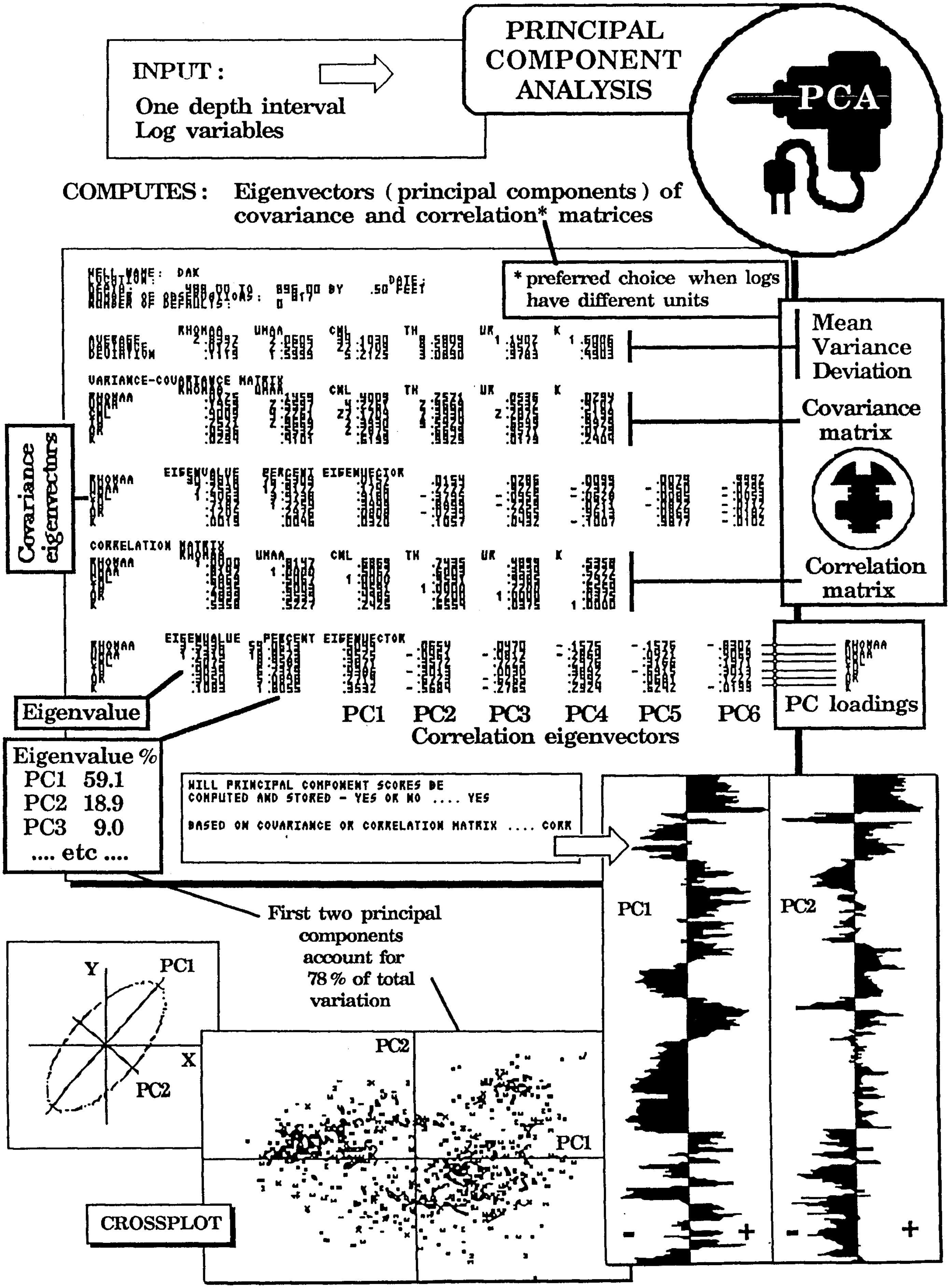


One of the main applications of principal component analysis is to reduce the unwieldy dimensionality of a data set with a large number of variables to a smaller number of dimensions. Principal components does this condensation with the minimum "damage" to the information content, because the principal components "soak up" the major sources of variation from largest to smallest. The majority of variation within data sets with many variables can often be mapped in two dimensions. This is because of the information redundancy in the data set implied by correlations between the variables.

The location of the pth principal component, Y_p with respect to the original variable axes, X_m is given by the coefficients , a_m, so that:

$$Y_p = a_1X_1 + \cdots\cdots a_mX_m$$

The coefficients can often be "read" as clues to the "meaning" of the principal component as an intrinsic process variable. The original obsevations can be converted into PRINCIPAL COMPONENT SCORES, which can be crossplotted or used as new variables in their own right.


PRINCIPAL COMPONENT ANALYSIS
INPUT: One depth interval Log variables
PCA
COMPUTES: Eigenvectors (principal components) of covariance and correlation* matrices
* preferred choice when logs have different units
Mean Variance Deviation
Covariance matrix
Correlation matrix
PC loadings
Covariance eigenvectors
Eigenvalue
Eigenvalue %
PC1 59.1
PC2 18.9
PC3 9.0
.... etc ....
PC1 PC2 PC3 PC4 PC5 PC6
Correlation eigenvectors
WILL PRINCIPAL COMPONENT SCORES BE COMPUTED AND STORED - YES OR NO .... YES
BASED ON COVARIANCE OR CORRELATION MATRIX .... CORR
First two principal components account for 78 % of total variation
Y
PC1
X
PC2
CROSSPLOT
PC1
PC2
- +
- +

PRINCIPAL COMPONENT ANALYSIS OF LOGS IN LOWER CRETACEOUS SANDSTONES AND SHALES

Input log variables :
RHOMAA
UMAA
CNL neutron
Thorium
Uranium
Potassium

Input logs are standardized to eliminate influence of different measurement units. The variance - covariance matrix is then the correlation matrix. The eigenvectors of the correlation matrix are the principal components of the standardized logs, and the eigenvalues express the relative amount of the total variation accounted for by each principal component.

Correlation matrix

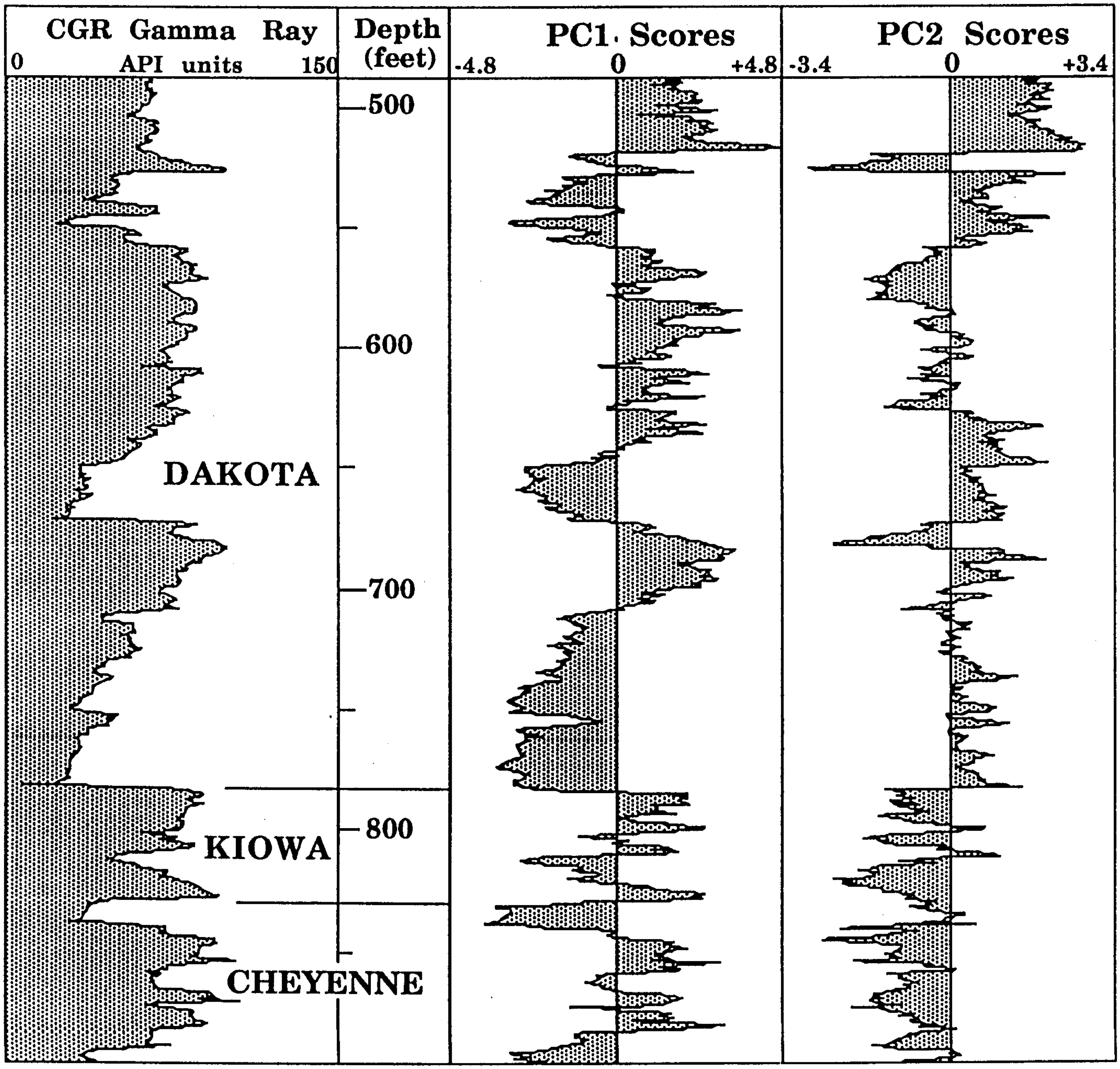
	RHOMAA	UMAA	CNL	TH	UR	K
RHOMAA	1.0000	.8147	.6869	.7435	.4899	.5358
UMAA	.8147	1.0000	.5067	.6003	.3534	.5227
CNL	.6869	.5067	1.0000	.4591	.4485	.2425
TH	.7435	.6003	.4591	1.0000	.2200	.6554
UR	.4899	.3534	.4485	.2200	1.0000	.0375
K	.5358	.5227	.2425	.6554	.0375	1.0000

Principal component eigenvectors

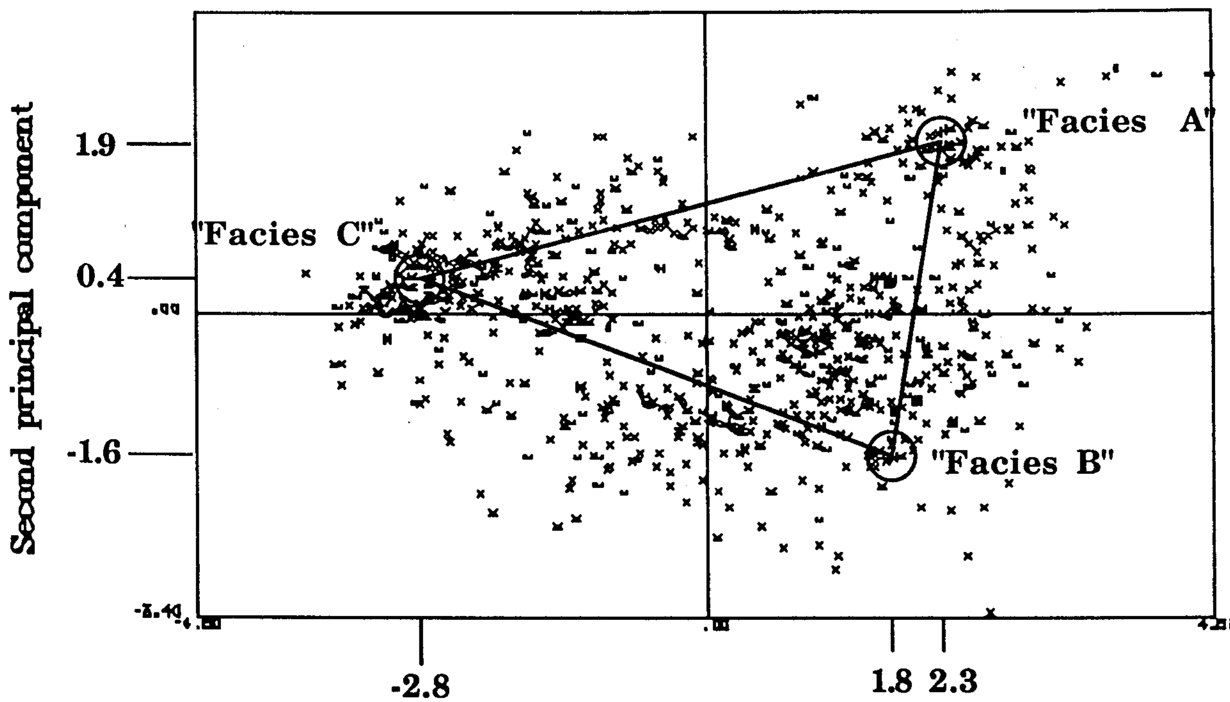
	PC1	PC2	PC3	PC4	PC5	PC6
RHOMAA	.5044	.0654	.0470	-.1575	-.	
UMAA	.4525	-.0461	-.0817	-.7864	.	
CNL	.3871	.3572	.7225	.2476	.	Etc.
TH	.4386	-.3013	.0030	.3842	-.	
UR	.2714	.6723	-.6265	.2492	.	
K	.3532	-.5684	-.2765	.2924	.	
Eigenvalue %	59%	19%	8 %	7 %	5 %	2 %

principal component scores

PRINCIPAL COMPONENT SCORE LOGS



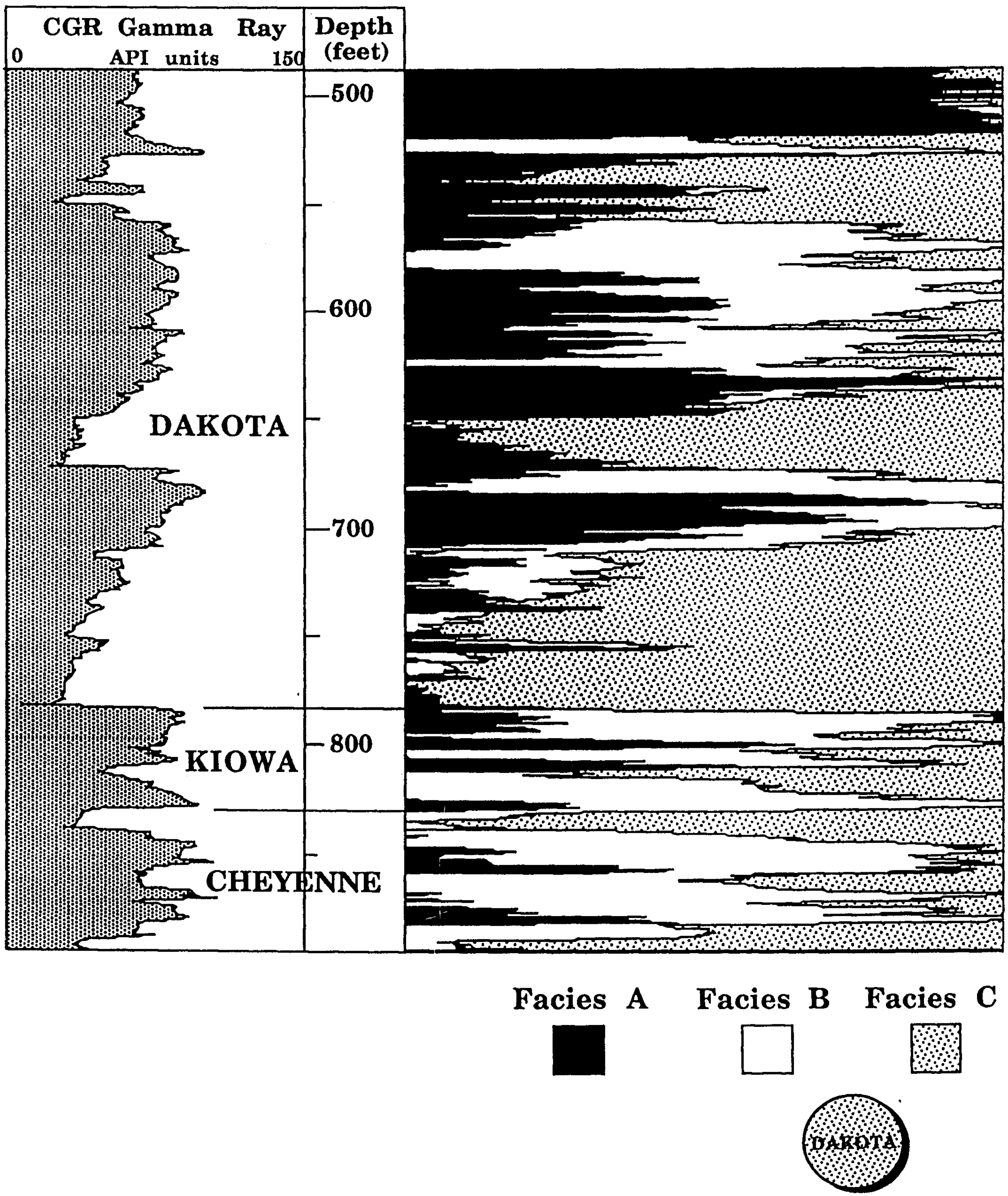
PATTERN RECOGNITION OF ELECTROFACIES
ON CROSSPLOT OF FIRST TWO PRINCIPAL COMPONENTS



PC score coefficients for compositional analysis by matrix algebra solution :

		PC1	PC2
Electrofacies	A	2.3	1.9
	B	1.8	-1.6
	C	-2.8	0.4

Matrix algebra solution of electrofacies located on crossplot of first and second principal components of RHOmaa, Umaa, Φn, Th, U, K



FACTOR ANALYSIS

FACTOR ANALYSIS is a theoretical model that postulates that observed variables are correlated with a lesser number of "hidden" variables or FACTORS which explain the systematic variation in the measured sample. The theory was developed in psychology, where factors were considered to be aptitudes or personality traits which accounted for patterns in results from examinations or questionaires.

The initial phase of factor analysis is programmed as a principal component solution, which generates m eigenvectors for m variables. However, while principal components are simply a geometrical result, factor analysis is a model, in which a FEWER causal variables are said to explain the data. Traditionally, the number and identity of the factors are supposed to be known beforehand and dictate the design of the exam or poll. This is usually not the case in geology. However, the PCA solution of eigenvalues can give insight on the intrinsic dimensionality of the data, and so the number of factors.

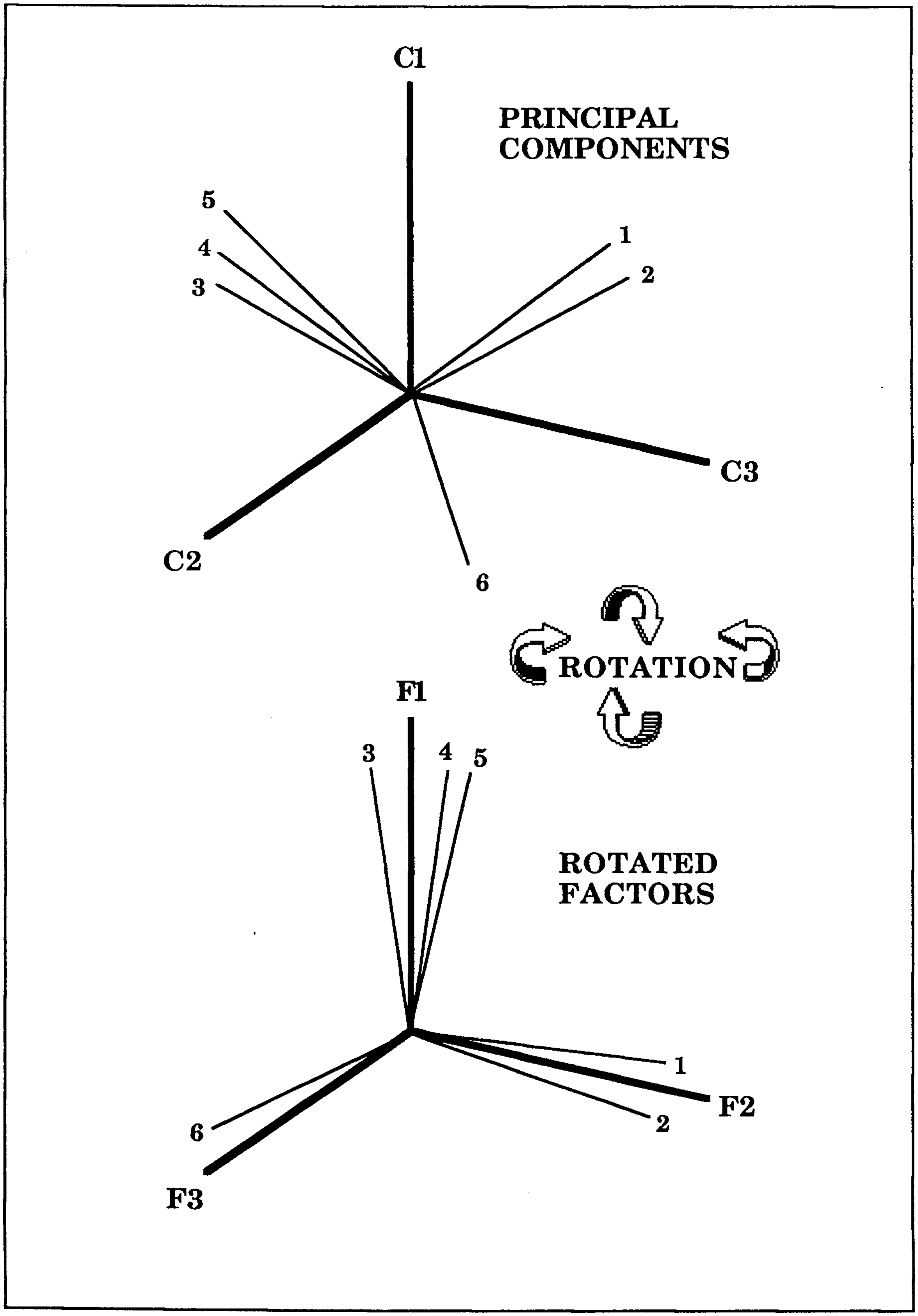
The factor model for k factors can be written as :

$$Z_j = a_{1j}F_1 + a_{2j}F_2 + \dots + a_k F_k + a_j E_j$$

where Z_j is the jth factor, a_k are the FACTOR LOADINGS, and E_j is the residual error. This equation may be used to transform the raw variables for any observation into a FACTOR SCORE.

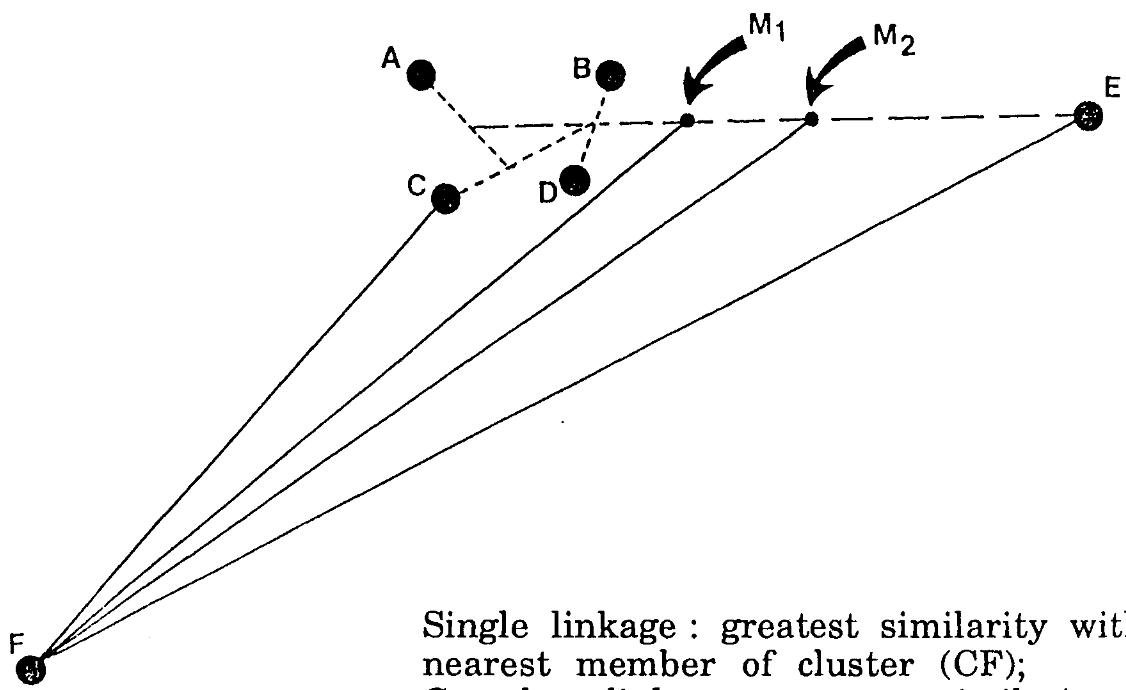
The initial principal components solution is unsatisfactory as a factor solution, both because it has as many components as the original variables, but also because the eigenvectors are generalized composites of the major sources of variation. The aim of factor analysis is to locate "simple" structure, in which each of the variables is expressed as either a strong loading (ideally, +1 or -1), or a weak loading (ideally, zero) with respect to each of the factors. The orthogonal axes are therefore rotated from the eigenvectors in search of a solution with a simple structure.

Once the factor model is solved, the factors may be interpreted as causal variables, based on the factor loadings of the original variables. The factor scores also provide new variables which, hopefully, are more closely linked with the phenomenom of interest than any of the original variables.

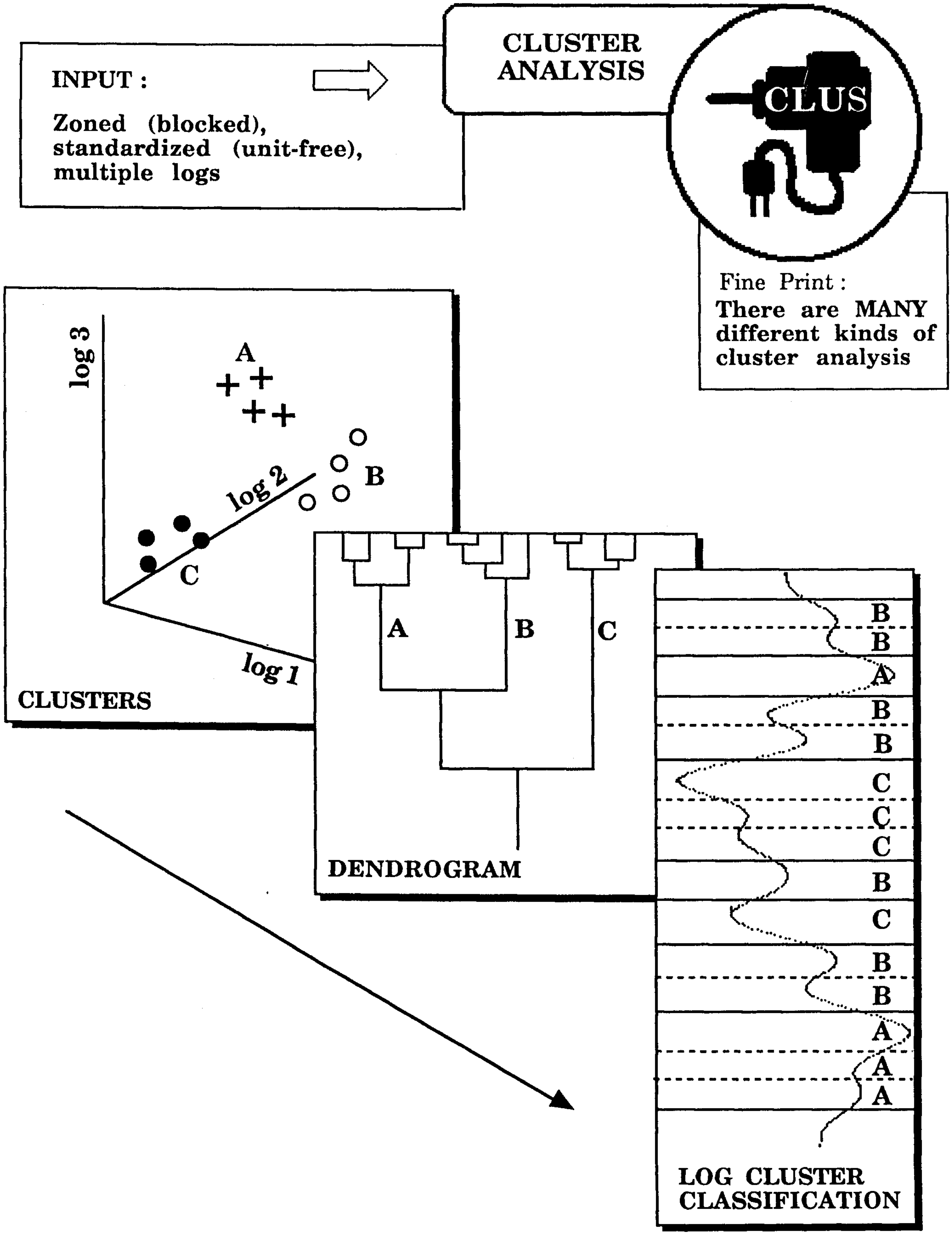

C1
PRINCIPAL COMPONENTS
5
4
3
1
2
C3
C2
6
ROTATION
F1
3
4
5
ROTATED FACTORS
1
F2
2
6
F3

CLUSTER ANALYSIS

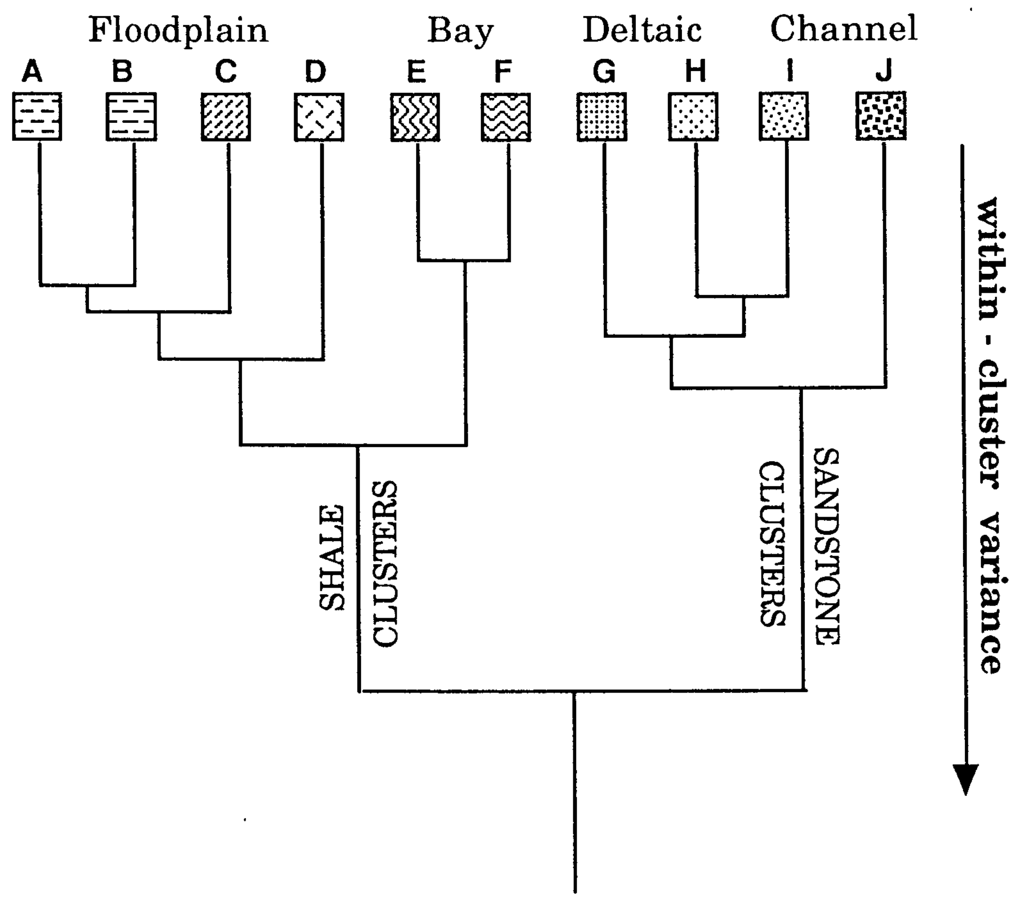
There are a great variety of CLUSTER ANALYSIS methods, each based on different strategies to assign observations to groups which are "homogeneous" and distinct from other groups. The most common techniques used in geology are those of HIERARCHIC CLUSTERING, largely because of their popularity as methods for NUMERICAL TAXONOMY of fossils. First an n x n matrix of similarities between all pairs of the n observations is calculated. Then the pairs with the highest similarities are merged, the matrix recomputed, and the procedure repeated until some "natural" level of clustering is reached, as assessed by visual inspection or some numerical criterion. The result is most commonly shown as a DENDROGRAM. Even in hierarchical clustering, there are different strategies to link observations with clusters in formation as illustrated below :



Single linkage : greatest similarity with nearest member of cluster (CF);
Complete linkage : greatest similarity with most dissimilar member of cluster (EF);
Average linkage : greatest average similarity with members of cluster (M2F);
Centroid linkage : least distance to centroid of cluster (M1F).

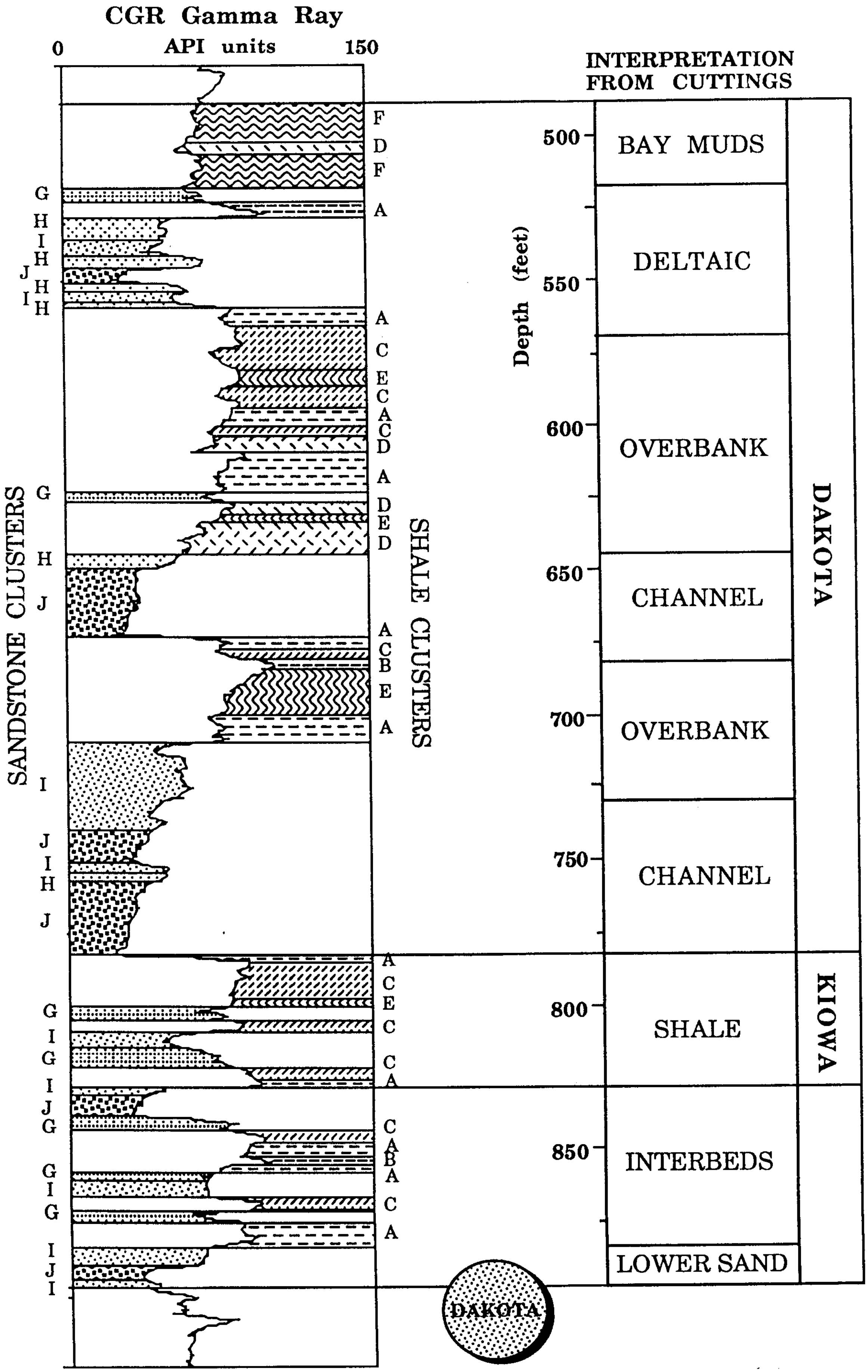

INPUT :
Zoned (blocked), standardized (unit-free), multiple logs
CLUSTER ANALYSIS
CLUS
Fine Print :
There are MANY different kinds of cluster analysis
log 3
log 2
log 1
A
B
C
CLUSTERS
A
B
C
DENDROGRAM
B
B
A
B
B
C
C
C
B
C
B
B
A
A
A
LOG CLUSTER CLASSIFICATION

CLUSTER ANALYSIS OF ZONES IN A SANDSTONE - SHALE SEQUENCE



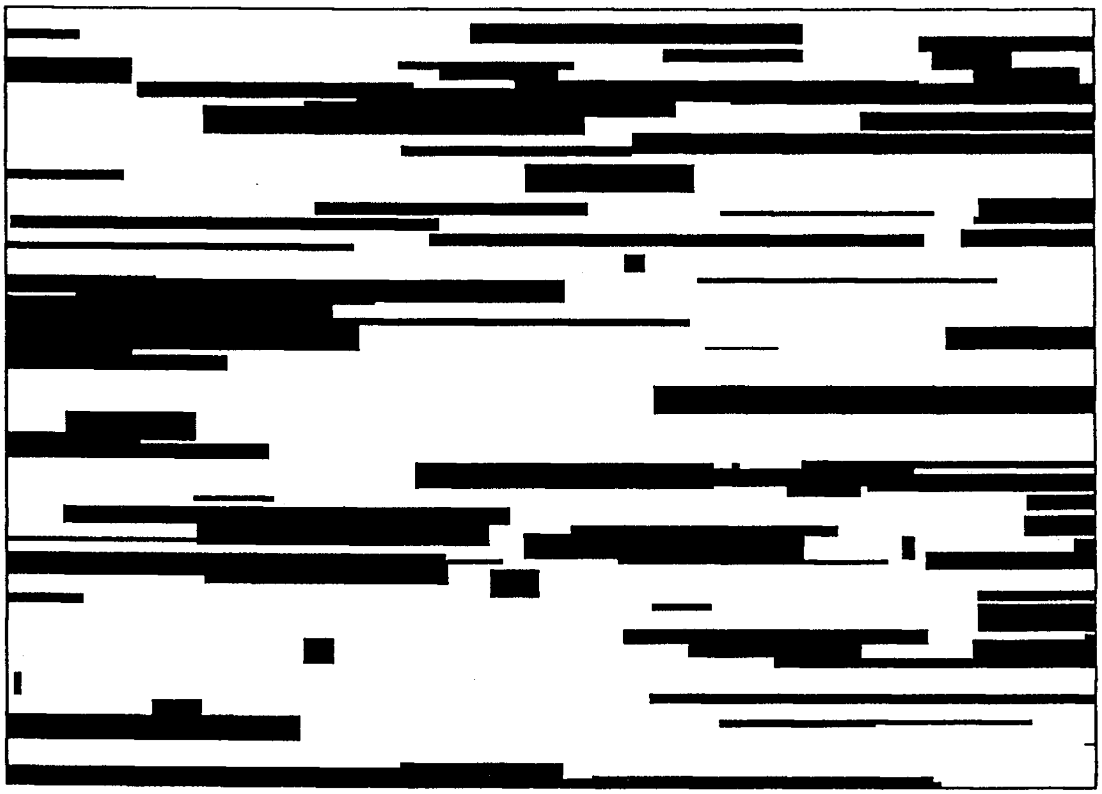
Input log variables for clustering :

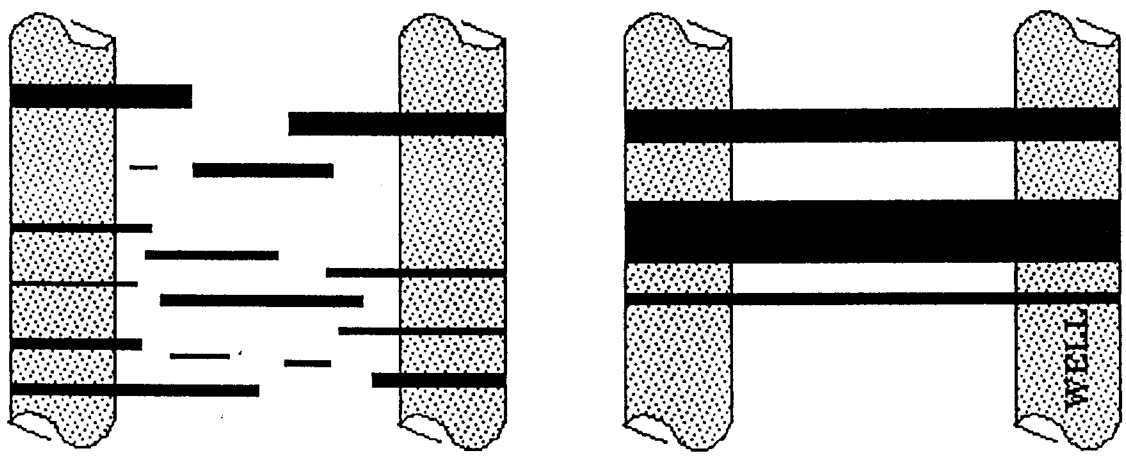
Apparent grain density (RHOmaa)
Apparent matrix photoelectric cross-section (Umaa)
Neutron porosity (ϕ_n)
Potassium (K)
Uranium (U)
Thorium (Th)


CGR Gamma Ray
0
API units
150
INTERPRETATION FROM CUTTINGS
Depth (feet)
500
550
600
650
700
750
800
850
BAY MUDS
DELTAIC
OVERBANK
CHANNEL
OVERBANK
CHANNEL
SHALE
INTERBEDS
LOWER SAND
DAKOTA
KIOWA
SANDSTONE CLUSTERS
SHALE CLUSTERS
DAKOTA

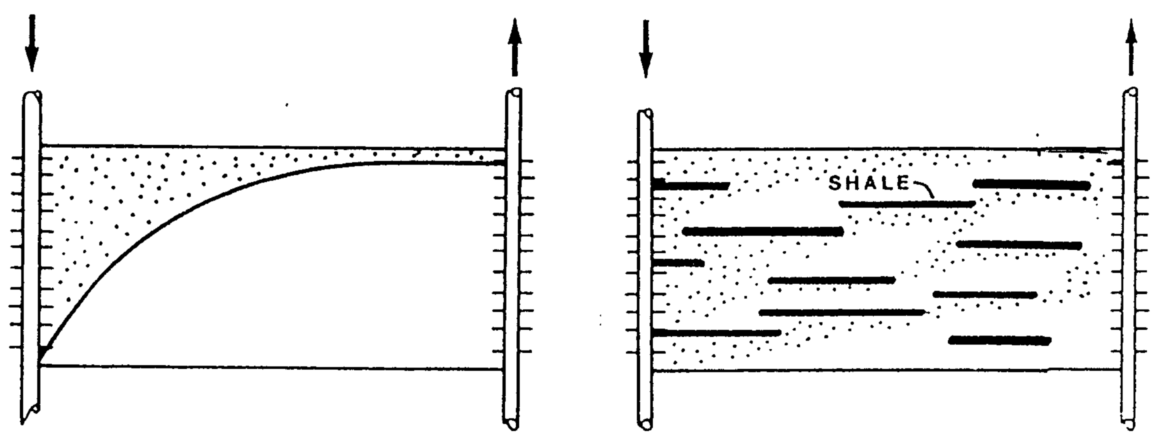
Blank

STOCHASTIC SIMULATION



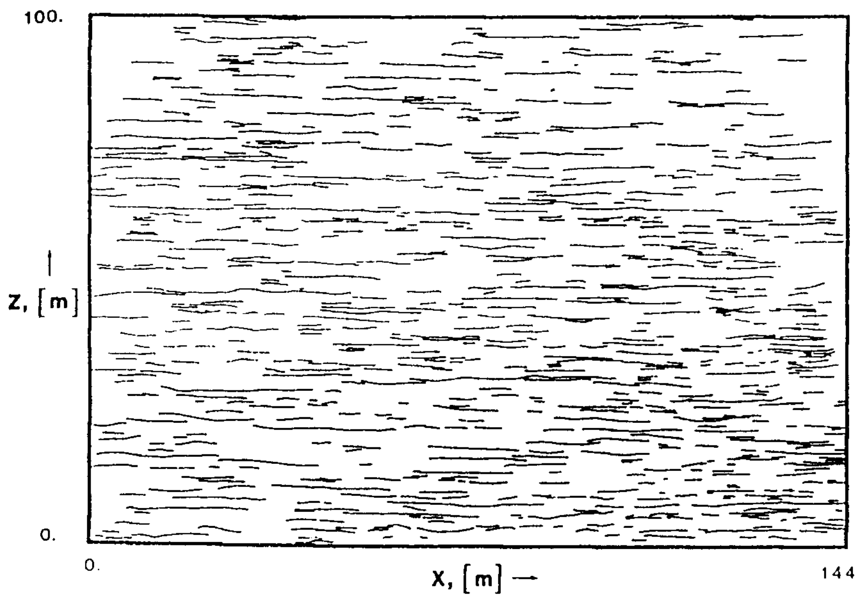


after Haldorsen and Lake (1984)

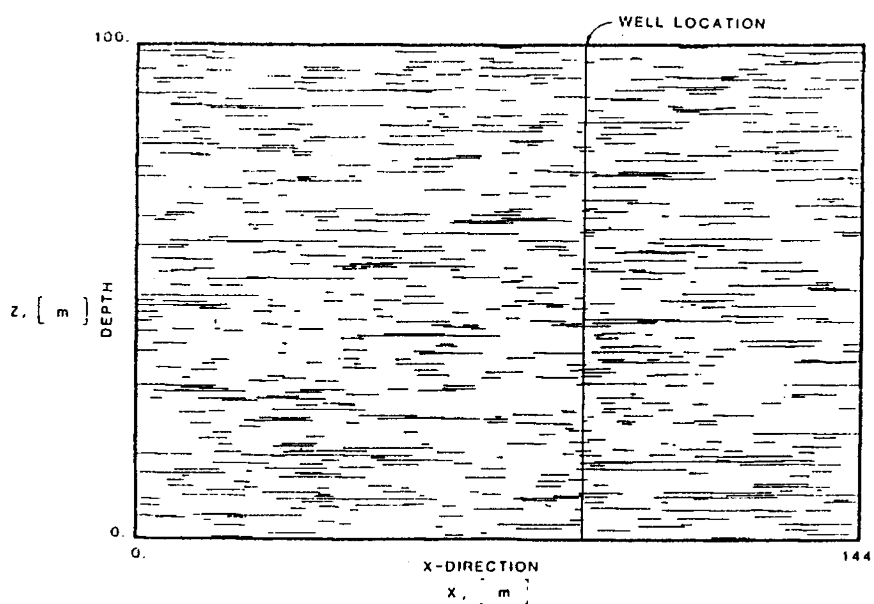


Homogeneous reservoir (left) and reservoir with stochastic shales (right). In horizontal drives, these shales may be advantageous because of increased vertical sweep (as shown in figure). (However, they may reduce vertical sweep in vertical drives.)
From Haldorsen and Chang (1986)

OUTCROP OCCURRENCE AND SIMULATION OF STOCHASTIC SHALES



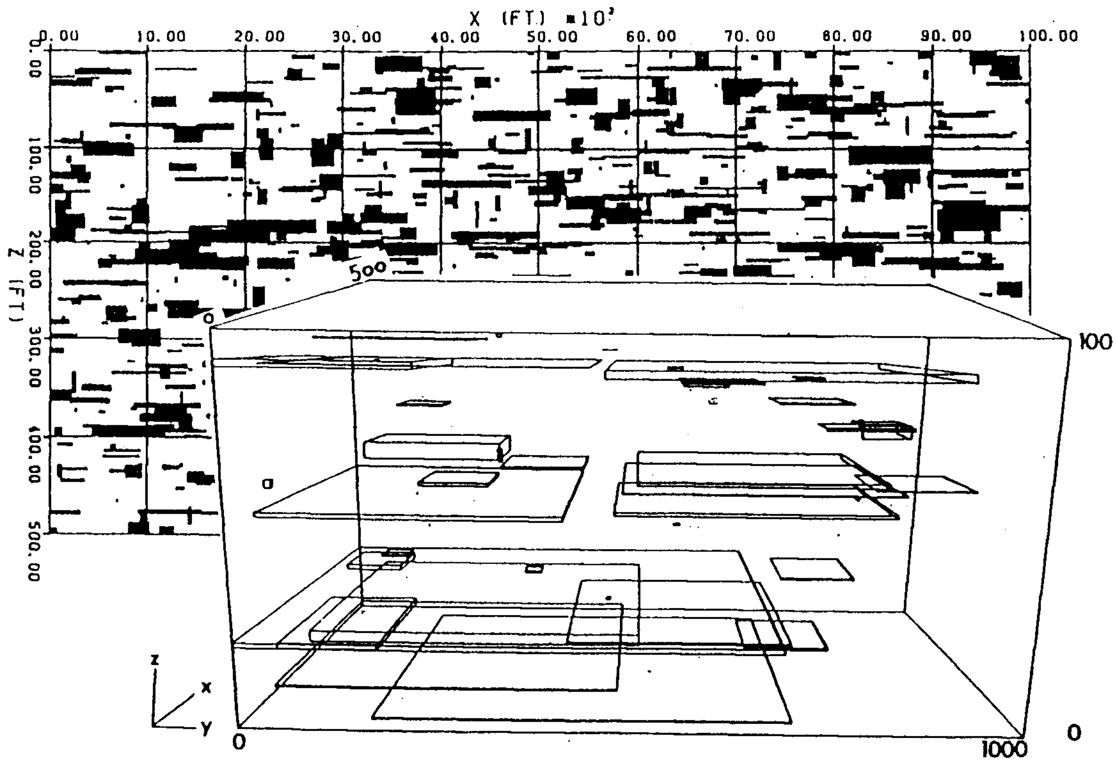
Actual shale distribution in an outcrop of the Assakao (fluvial) Sandstone



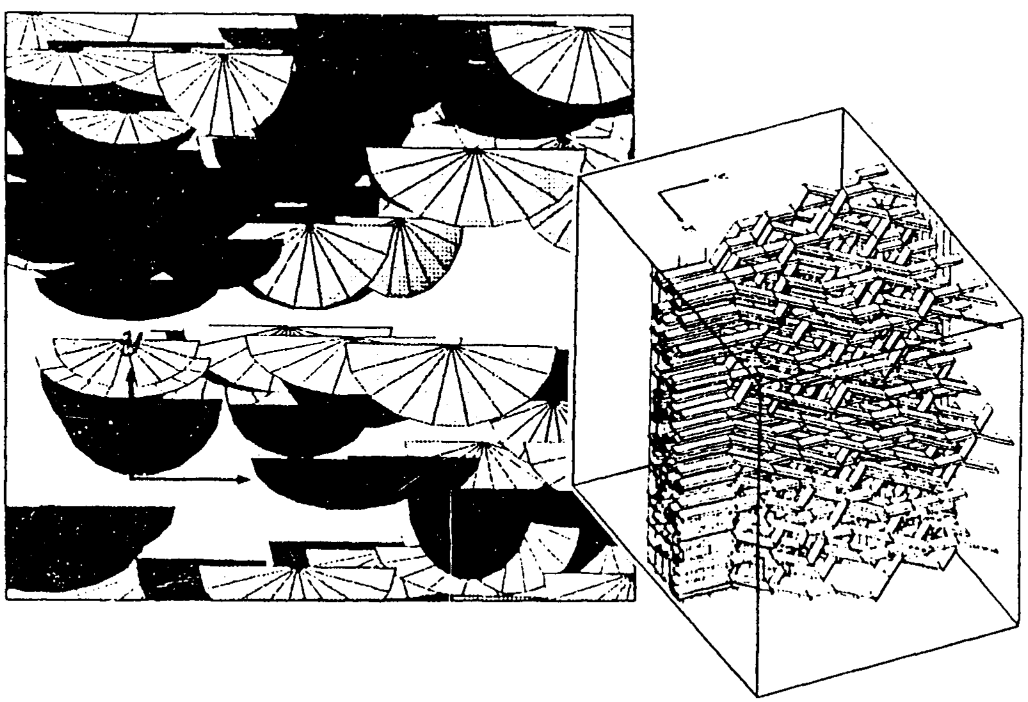
Synthetic cross-section generated from parameters of the Assakao Sandstone outcrop

from Haldorsen and Chang (1986)

2-D AND 3-D STOCHASTIC SIMULATION MODELS OF SHALES AND CHANNEL SANDSTONES

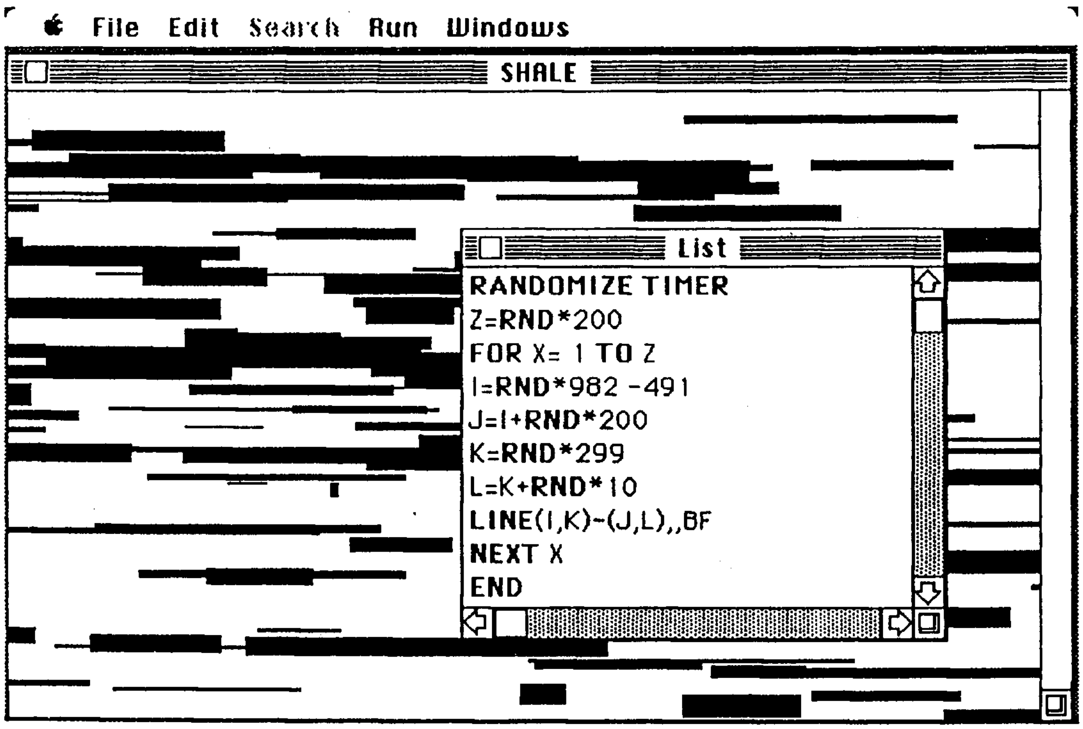


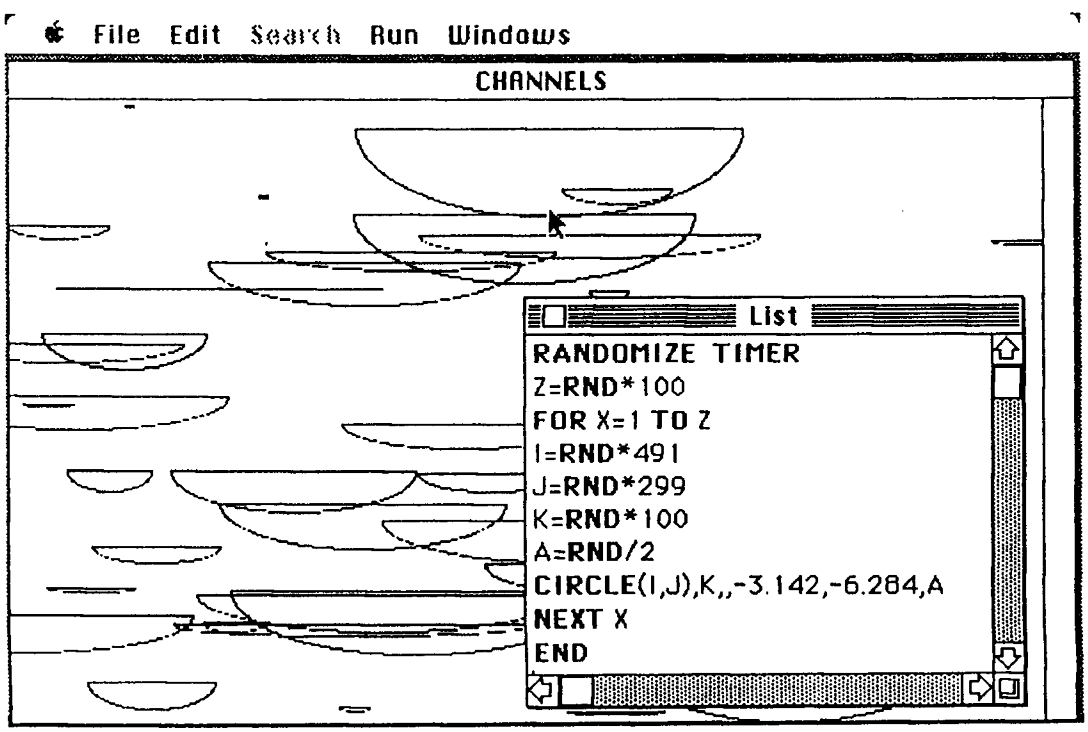
from Haldorsen and Lake (1984)



from Haldorsen and MacDonald (1987)

SIMPLE BASIC PROGRAMS FOR SIMULATION OF RANDOM SHALES AND SANDSTONE CHANNELS





Blank

REFERENCES

Adams,J.A.S., and Weaver,C.E., 1958, Thorium to uranium ratios as indications of sedimentary processes : Example of concept of geochemical facies : *Bull. AAPG,* v. 42, no. 2, p. 387 - 430.

Almouslli,M.O., 1987, *The application of the gamma ray spectralog to the analysis of the Chase Group (Gearyan Stage, Lower Permian system) in Stevens County, Kansas,* Unpubl. MS thesis, Wichita State Univ., 142 pp.

Baker,P.E.,1957, Density logging with gamma rays : *Petroleum Trans. AIME,* v. 210, p. 289 - 294.

Berendsen,P., and Doveton,J.H., 1988, Ironstones from the Lower St. Peter Sandstone (Middle Ordovician) in northeast Kansas : *Geol. Soc. Amer. South-Central Section Mtg.* Abstracts p. 91.

Bigelow,E.L., 1985, Making more intelligent use of log derived dip information : Part V : Stratigraphic interpretation : *The Log Analyst* : v. XXVI, no. 5, p. 25 - 64.

Bornemann,E., and Doveton,J.H., 1983, Lithofacies mapping of Viola limestone in south-central Kansas, based on wireline logs : *Bull. AAPG,* v. 67, no. 4, p. 609 - 623.

Briggs,P.L., 1985, Color display of well logs (abs.) : *Math. Geology,* v. 17, no. 4, p. 481.

Burke,J.A., Campbell,R.L., Jr., and Schmidt,A.W., 1969, The Litho-porosity Cross Plot : *Trans. SPWLA,* Paper Y, 29 pp.

Collins,R.E., 1961, *Flow of Fluids through Porous Materials* : Reinhold, New York, 270 pp.

Choquette,P.W., and Pray,L.C., 1970, Geologic nomenclature and classification of porosity in sedimentary carbonates : *Bull. AAPG,* v. 54, no.2, p. 207 - 250.

Choquette,P.W., and Traut,J.D., 1963, Pennsylvanian carbonate reservoirs, Ismay Field, Utah and Colorado : *in, A Symposium, Shelf Carbonates of the Paradox Basin :* Four Corners Geol. Soc. 4th Field Conf., p. 157 - 184.

Clavier,C., and Rust,D.H., 1976, MID plot: A new lithology technique : *The Log Analyst,* v. XVII, no. 6, p. 16 - 24.

Collins,D.R., and Doveton,J.H.,1988,Colour image transformations of wireline logs as a medium for sedimentary profile analysis : *Bull. Canadian Petr. Geology,* v. 36, no. 2, p. 186 - 190.

Collins,D.R., and Doveton, J.H.,1989, Applying color information theory to the display of lithologic images from wireline logs : *Geobyte,* v. 4, no. 1, p. 16 - 24. (Color reprint of *Trans. SPWLA,* 1988, Paper II.)

Dixon,S.A., Summers,D.M., and Surdam,R.C., 1989, Diagenesis and preservation of porosity in Norphlet Formation (Upper Jurassic), Southern Alabama : *Bull. AAPG,* v.73, no. 6, p. 707 - 728.

Doveton,J.H., 1986, *Log Analysis of Subsurface Geology : Concepts and Computer Methods :* Wiley - Interscience, New York, 273 pp.

Doveton,J.H., and Bornemann,E., 1981, Log normalization by trend surface analysis : The Log Analyst, v. XXII, no. 4, p. 3 - 8.

Doveton,J.H., Zhu Ke-an, and Davis, J.C., 1984, Three-dimensional trend mapping using gamma-ray well logs : Simpson Group, south-central Kansas : *Bull. AAPG,* v. 68, no. 6, p. 690 - 703.

Ebanks,W.J.,Jr., Euwer,R.M., and Nodine-Zeller,D.E., 1977, Mississippian combination trap, Bindley Field, Hodgeman County, Kansas : *Bull. AAPG,* v. 61, no. 3, p. 309 -330.

Edmundson,H., and Raymer,L.L., 1979, Radioactive logging parameters for common minerals : Trans. SPWLA, Paper O, 19 pp.

Ellis,D.V., 1987, *Well Logging for Earth Scientists* : Elsevier, New York, 532 pp.

Elphick,R.Y., 1987, Nuclear log interpretation in hard rock formations : *Geobyte,* v. 2, no. 3, p. 44 - 47.

Fuchtbauer,H., 1967, Influence of different types of diagenesis on sandstone porosity : *Seventh World Petr. Congr.* , v. 2, p. 353 - 368.

Gardner,J.S., and Dumanoir,J.L., 1980, Lithodensity log interpretation : *Trans. SPWLA,* Paper N, 23 pp.

Geehan,G.W., Lawton,T.F., Sakurai,S., Klob,H., Clifton,T.R., Inman,K.F., and Nitzberg,K.E., 1986, Geologic prediction of shale continuity, Prudhoe Bay Field : *in, Reservoir Characterization* (Lake and Carroll, eds.), Academic Press, Orlando, p. 63 - 82

Haldorsen,H.H., and Chang,D.M., 1986, Notes on stochastic shales : From outcrop to simulation model : *in, Reservoir Characterization* (Lake and Carroll, eds.), Academic Press, Orlando, p. 445 - 486.

Haldorsen,H.H., and Lake,L.W. 1984, A new approach to shale management in field-scale models : *Jour. SPE.*

Haldorsen,H.H., and MacDonald,C.J.,1987, Stochastic modelling of underground reservoir facies (SMURF) : *SPE 16751,* 99 - 113.

Harbaugh,J.W., Doveton,J.H., and Davis,J.C.,1977, *Probability Methods in Oil Exploration* : Wiley - Interscience, New York, 269 pp.

Hattin,D.E., 1967, Stratigraphic and paleoecologic significance of macroinvertebrate fossils in the Dakota Formation (Upper Cretaceous) of Kansas : *in, Essays in Paleontology and Stratigraphy,* Raymond C. Moore Comm. Vol. : Univ. Kans., Dept. Geol. Spec. Publ. 2, p. 570 - 589.

Herron,M.M., 1987, Estimating the intrinsic permeability of clastic sediments from geochemical data : Trans. SPWLA, Paper HH, 22 pp.

Kimminau,S., and Schwartz,L., 1987, A review of pore and grain geometric models : *Trans. SPWLA,* Paper AA, 10 pp.

Krumbein,W.C., and Monk,G.D., 1942, Permeability as a function of the size parameters of sedimentary particles : *AIME Tech. Publ.* 1492, p. 153 - 162.

Lavers,B.A.,Smits,L.J.M., and van Barren,C., 1974, Some fundamental problems of formation evaluation in the North Sea : *SPWLA Third European Symposium Trans.*, p. 33 - 47.

Macfarlane,P.A., Doveton,J.H., and Coble,G., 1989, Interpretation of lithologies and depositional environments of Cretaceous and Lower Permian rocks by using a diverse suite of logs from a borehole in central Kansas : *Geology,* v. 17, p. 303 - 306.

Muskat,M., 1949, *Physical Principles of Oil Production* : Mcgraw - Hill, New York, 922 pp.

Neasham,J.W., 1977, The morphology of dispersed clay in sandstone reservoirs and its effect on sandstone shaliness, pore space and fluid flow properties : *Jour. SPE*

Preston,F.W., 1959, Fluid flow characteristics of core : in, Description of a Dakota (Cretaceous) Core from Cheyenne County, Kansas, (Merriam et al) : *Bull. Kans. Geol. Surv.* 134, Part 1, p. 90 - 97.

Ruhovets, N., and Fertl,W.H., 1982, Digital shaly sand analysis based on Waxman-Smits model and log-derived clay typing: *The Log Analyst,* v. XXIII, no. 3, p. 7 - 23.

Schlumberger, 1972, *Log Interpretation : Volume I - Principles* : Schlumberger, New York, 113 pp.

Schlumberger, 1988, *Log Interpretation Charts* : Schlumberger, 150 pp.

Self,G.A., Breard,S.Q., Rael,H.P., Stein,J.A., Thayer,P.A., Traugott,M.O., and Easom,W.D., 1986, *Bull. AAPG,* v. 70, no. 5, p. 501 - 515.

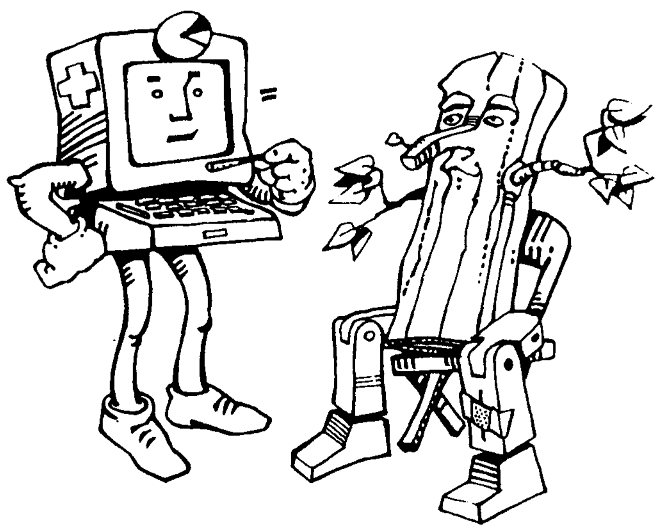
Toomey, D.F., and MitchellR., 1986, Facies relationships and paleodepositional settings of the Herington - to - Winfield stratigraphic interval (Lower Permian), southern Kansas - northern Oklahoma : *Field Guidebook, 4 th Ann. Mtg., Midcontinent Section SEPM,* Ponca City, Okla.

Watney,W.L.,1985, *Origin of Four Upper Pennsylvanian (Missourian) Cyclothems in the Subsurface of Western Kansas : Application to the Search for Accumulation of Petroleum* : Unpubl. PhD diss., Univ. Kansas, 506 pp.

Wilson,M.D., 1982, Origins of clays controlling permeability in tight gas sands : *JPT* (Dec.), p. 2871 - 2876.

Wilson,M.D., and Pittman, E.D., 1977, Authigenic clays in sandstones : recognition and influence on reservoir properties and paleoenvironmental analysis : *J. Sed. Pet.,* v. 47, no. 1, p. 3 - 31.

COMPUTER ANALYSIS
OF WELL LOG DATA



"Temperature normal... sap pressure normal... except for that nasty bark on your shin I'd say you're a well log."

(GEOBYTE)