DEPARTMENT OF THE INTERIOR
U.S. GEOLOGICAL SURVEY

Computer Program for Geochemical Correlation Studies

by

John B. Rapp

U.S. Geological Survey
Menlo Park, California 94025

OPEN FILE REPORT 84-415

INTRODUCTION

This report discusses one of the computer programs which were written for our study of oil and rock correlations of samples from the North Slope, Alaska (Kvenvolden and others). The program produces tables of calculated data which are used for correlations. The program can be used for all combinations (oil/oil,oil/rock,rock/rock) of correlations by simply selecting which samples are inputted. Any group of chemical compounds may be used in the program but we used different classes of hydrocarbons (normal alkanes, isoprenoids, steranes and hopanes) in our study. The report is oriented toward users as opposed to programmers. If the reader needs more detail, the listing of the program is given in the Appendix and contains remark statements to assist the reader. The program is interactive and all input requests are in the form of questions and non-numeric input requests are usually followed by a list of the possible choices. Examples of the output produced by the program are presented in the tables. Choices and limitations of input and output are discussed.

The program is written in BASIC and run on a Honeywell 61/80 computer with the Multics operating system. The reasons for using BASIC are that it is an interactive language and that it is widespread in today's microcomputers. The program is designed to be run on a wide variety of computers including 8 bit microcomputers. Many variables and matrices are re-defined several times in order to conserve memory (for running on microcomputers with small memories).

OVERVIEW

The name of the program is OIL/ROCK.BASIC . The objective of the program is to assist the user in grouping samples (oils and/or rocks) according to their hydrocarbon composition. The data base consists of several files which are constructed before the program is run. Some files contain names of samples or names of hydrocarbon compounds. Other files contain analyses of hydrocarbons in the samples (either concentrations (ppb) or peak heights). The program begins with an initialization section in which the user determines which data file (analyses of which hydrocarbon class-e.g. HOPANE) to be used in the calculations. This data is placed in one matrix, the corresponding names of the hydrocarbon compounds are placed in another martrix and the sample names are placed into a third matrix. The program now branches to one of the subprograms, to the selection of a new data matrix or to the end of the program. This is the major branching point to which all subprograms return. The names of the four subprograms are Ratio, Carbon Preference Index, Odd-Even Predominance, and Dissimilarity Index. The subprogram Ratio calculates and outputs simple ratios (single numerator and single denominator) or multiple ratios (summations in the numerator and/or summations in the denominator). It can be used with compounds from any of the hydrocarbon classes (e.g. STERANES). The subprograms Carbon Preference Index and Odd-Even Predominance are two different methods of calculating a ratio between odd-carbon-numbered normal alkanes and even-carbon-numbered normal alkanes. The subprogram Dissimilarity Index (DI) compares a number of variables in two samples and results in a number which is a measure of the differences between the two samples. The technique is useful in comparing the overall composition (as opposed to only one or two variables) of samples and in grouping samples which have low dissimilarity indices (the output table has been sorted in ascending order and samples with similar DIs are adjacent to each other). This

subprogram may be used with compounds from any of the hydrocarbon classes.

MATRICES

Many matrices are common to all the subprograms. All the matrices are stored within the computer as files and the requested matrices are read by the programs at run time.

One set of matrices contains data (in two dimensions - sample number versus compound number). Matrix ALKANE contains normal alkanes and the isoprenoids pristane and phytane, matrix HOPANE contains tricyclic and pentacyclic terpanes and matrix STERANE contains steranes and diasteranes. The data is quantitated in two ways: 1. concentration (ppb) - where a quantitative standard is available (e.g. matrix ALKANE - ppb contains concentrations of normal alkanes and the isoprenoids pristane and phytane) and 2. peak height - where only a qualitative standard (peaks identified but not quantitated in concentration) is available (e.g. the matrices STERANE, HOPANE and ALKANE-HT).

Another set of matrices contains names. One matrix contains sample names of oils (No. 1-8, 24) and rocks (No. 9-23) from the North Slope of Alaska and was called NS-NAME. Table 1 contains sample numbers and sample names in the Matrix NS-NAME. Other matrices contained compound names and were called the concatenation of the data matrix name and the string "-str" (e.g. STERANE-STR and HOPANE-STR). Table 7 contains the compound numbers and abbreviated names of the compounds in the matrices HOPANE and STERANE (contents of matrices HOPANE-STR and STERANE-STR). Table 8 contains compound numbers and compound names of the terpanes and steranes used in the study. The alkane matrices consisted of n-C_{12} to n-C_{40} (compounds 1 to 29), pris (pristane:compound 30) and phyt (phytane:compound 31).

Several matrices and simple variables are re-defined several times in a single subprogram. This is done for several operations where memory is needed to store intermediate results of calculations. Examples of such operations are sort routines and summations. Normally, the practice of re-defining variables should be avoided in programming but is used here to conserve memory so that these programs might be run on microcomputers with limited memories.

Some of the limitations imposed on the program are the following maximums: 25 samples, and 40 compounds in each data matrix, 20 letters in a sample name, 8 letters in a compound name in all subprograms and 9 ratios, 9 numerators and 9 denominators in the ratio subprogram. Most of these limitations are imposed to fit this one set of data and can easily be expanded to larger values for larger data sets. This entails changing the dimension statement and a few lines of program. Some of the limitations are imposed by the printer width (132 characters). The name lengths are limited in order to get more columns of data per page. Most of the time the number of characters in the column heading name is larger than the number of characters in the data.

OIL/ROCK.BASIC

As stated previously, the objective of the program is to assist the user in grouping samples (oil and rocks) according to their hydrocarbon composition. The output from this program are tables of data ordered by (1.) requested samples and (2.) sorted data. Observing sample similarities, differences and groupings is much easier when sorted data are used. The program is quite versatile in that there is a very large number of possible

combinations of samples and compounds. All the samples (oils and rocks) or only selected ones may be compared and thus all combinations (oil/oil,oil/rock,rock/rock) of correlations may be produced. With the present program, there is a limitation on the combinations of hydrocarbon compounds which may be compared. Only compounds within a single matrix may be compared (e.g. only compounds within HOPANE with each other and not with compounds in STERANE). The program can be easily modified (by combining the matrices) so that any combination of compounds can be compared. The program listing is given in the Appendix. Remark statements (rem) are used in the listing to delineate subprograms and subroutines and to clarify certain statements. This program is composed of an initialization section and four subprograms: Ratio, Carbon Preference Index, Odd-Even Predominance and Dissimilarity Index.

Initialization. After dimensioning the matrices, the program requests the name of the data file to be used. The program then loads that data file into matrix m and defines file No. 2 as the concatenation of the data file name and the string"-str" (this file contains the names of the compounds). File No. 3 is defined as "NS-NAME": the names of the samples. The program then requests whether the user wants to calculate a ratio of compounds, a dissimilarity index, a Carbon Preference Index, or an Odd/Even Predominance or to select a new data matrix or to stop the program This is the major branching point to which all sub-programs return.

Ratio. The sub-program Ratio calculates and outputs simple ratios (single numerator and single denominator) or multiple ratios (summations in the numerator and/or summations in the denominator). The program first asks which samples to include in the output table. It asks the total number of samples and if the sample numbers are contiguous or not (the rock sample numbers are contiguous and the oil sample numbers are contiguous except number 24 which was added later). This is to make sample number entry easier - its easier to enter a range than each individual sample. An example would be if all rock samples were requested - it would be easier to input the two number range (9,23) than to input the fifteen individual sample numbers (9,10,11,....22,23). It then asks which ratios to print in the table. There is a maximum number of ratios of 9 because the table would get too crowded with more. The next question is "Multiple Numerator/Denominators (y/n)? An answer of "n" implies a simple ratio and asks a different set of questions about numerators and denominators.

The output table contains the run date, sample number, sample name, and requested ratios. The heading for simple ratios is the abreviations of the numerators and denominators above the appropriate columns whereas the heading for multiple ratios is just RATIO 1, RATIO 2, etc. because of space limitations. The initial table is printed in the order that the sample numbers were inputted. The program then asks if the user would like a table with the ratios sorted. If the answer is yes then the question is asked "sort according to which ratio." The table is then printed in ascending order of the ratio selected with the other ratios following the same sample order. If the answer is no to the sort question then no further table is produced. If a denominator is zero for a sample, asterisks are printed for the value of the ratio for that sample. Table 2 is an example of the output for simple ratios. The first part is the output in the order of requested samples while the second part is sorted according to the value of pristane/phytane. Note the asterisks for sample eight: none of the requested alkanes was present in the sample. It is easy to observe in the sorted table that samples 1,2,3, and 4 form a group with similar ratios. Table 3 is an example of the output for

multiple ratios. The numerator is the sum of four diasteranes and the denominator is the sum of four steranes. The second part of the table contains the ratios sorted in ascending order. The group formed by samples 1,2,3,4, and 7 is much more easily delineated in the sorted table than in the unsorted table. This grouping is also supported by information reported earlier (Kvenvolden and others).

Carbon Preference Index. This subprogram calculates the carbon preference index (CPI) of the normal alkanes of samples according to an extension of the method of Bray and Evans (1961). CPI is a "ratio" of odd-carbon-numbered to even-carbon-numbered normal alkanes and is calculated as follows:

$$CPI = \frac{\sum_{i=j}^{k} C_{2i+1}}{2}\left[\frac{1}{\sum_{i=j}^{k} C_{2i}} + \frac{1}{\sum_{i=j}^{k} C_{2i+2}}\right]$$

Where C_{2i} = the concentration measure (weight percent, mole percent, concentration, peak height, etc) of a normal alkane with carbon number 2i.

j = one half the lowest carbon numbered alkane in the range.

k = (one half the highest carbon numbered alkane in the range) minus 1.

The range of the original method was limited to n-C_{24} to n-C_{34} (j=12, k=16) whereas the range of this subprogram is user defined. CPI is useful as a maturity indicator (i.e. recent sediments have high values whereas ancient sediments have values approaching 1).

The program first checks which matrix is being used. It can only be used with the "ALKANE-" matrices, otherwise an error is printed. Next, the range of hydrocarbons to be used in the calculation is requested. If the carbon numbers entered aren't both even then an error is noted and the question is repeated. Because of the structure of the alkane matrix, the CPI range is limited to n-C_{12} to n-C_{40}. The structure can easily be modified to extend the range.

The subprogram's output is a table containing the run date, name of the matrix used in the calculations, a heading and a sorted list of CPIs. If all the compounds in the selected range are present in the sample, the CPI is printed. If any of the first three (lowest carbon numbered) compounds in the selected range are missing, CPI is not printed but the lowest missing compound's carbon number is printed in the "missing alkanes" column. If any compound beyond the first three is missing, the lowest missing compound's carbon number is printed in the "missing alkanes" column and a "pseudo CPI" is calculated using summations accumulated to that point and is printed in parenthesis in the CPI column. The CPIs are sorted and printed in ascending order in the table. The "pseudo CPI" should not be used with or compared to the real CPIs in the table. It is only printed in order to give the user an estimate of CPI in range from the lowest carbon numbered compound in the selected range to the compound just before the missing compound. Table 4 is an example of various CPI outputs. The range is n-C_{16} to n-C_{28}. The values for oils are printed first. Sample 8 doesn't contain n-C_{16} so no CPI is printed. Sample 6 doesn't contain n-C_{20} so a pseudo CPI is printed in

parenthesis. The values for rocks are printed next. Note that sample 10 is higher than the rest, indicating less maturity.

Odd-Even Predominance. This subprogram calculates the odd-even predominance (OEP) of the normal alkanes of samples according to the method of Scalan and Smith (1969). The calculation uses five consecutive members of the n-alkane homologous series. OEP is a ratio of odd-carbon-numbered to even-carbon-numbered normal alkanes and is calculated as follows:

$$OEP = \left[\frac{C_i + 6C_{i+2} + C_{i+4}}{4C_{i+1} + 4C_{i+3}}\right]^{(-1)^{i+1}}$$

where C_i = the concentration measure (weight percent, mole percent, concentration, peak height, etc) of a normal alkane with carbon number i.

The OEP value is assigned to the alkane in the center of the group (alkane containing i + 2 carbon atoms). It is used to relate samples and as a maturity indicator (young samples have high values and mature samples have values around 1.0). It is similar to CPI but with more filtering - large variations in the raw data have less effect on OEP than on CPI. The program first checks that the current data matrix is "ALKANE-." It then asks which carbon number the user wishes the OEP to center around and which samples are to be used.

The program's output is a table containing the run date, name of the matrix used, a heading containing the carbon number about which the OEP was calculated and a sorted list of OEPs. If any of the values of the five consecutive alkanes in a sample is missing then no OEP is printed but the sample number, sample name, and the carbon number of the lowest missing alkane are printed at the top of the table. The rest of the table contains OEP values in ascending order. This ordering is useful in grouping samples since samples with similar OEPs are adjacent to each other. Table 5 is on OEP output centering at $n\text{-}C_{25}$. Oil and rock OEPs were printed separately for clarity but could easily have been printed together. Samples 6 and 8 didn't contain $n\text{-}C_{23}$ so their OEPs are not printed. Because the rock OEP listed is sorted, it is easy to see (1.) that generally the Pebble and Torok samples have higher OEPs than the rest of the samples and (2.) that sample 10 is very different from the rest of the samples (1.57 versus 1.29 for the adjacent sample).

Dissimilarity Index. Dissimilarity Index (DI) (Montalvo, 1978) is the result of a mathematical calculation which compares a number of selected variables in two samples (somewhat analogous to correlation coefficient which compares many samples for two variables -comparing samples versus comparing variables). The technique is useful in comparing the overall composition (as opposed to only one or two variables) of two samples and in grouping samples which have low dissimilarity indices. The variables in each sample are first normalized such that the summation of all the selected variables in that sample equals 1.0 .

$$V_{xy} = \frac{C_{xy}}{\sum_{i=1}^{t} C_{xi}}$$

where x = sample number
y = compound number
V_{xy} = normalized variable for sample x and compound y
C_{xy} = concentration measure (weight %, mole %, concentration (ppb), peak height, etc) of compound y in sample x

The absolute difference between the normalized variable in the two samples is summed for all the compounds. The summation is divided by 2.0 to give the dissimilarity index for the sample pair:

$$DI_{x_1x_2} = \frac{\sum_{i=1}^{t} \left| V_{x_1i} - V_{x_2i} \right|}{2}$$

where $DI_{x_1x_2}$ = dissimilarity index for the sample pair x_1 and x_2
t = total number of compounds used
V_{x_1i} = normalized variable for compound i in sample x_1
V_{x_2i} = normalized variable for compound i in sample x_2

The limits for dissimilarity index are 0 for a perfect match (identical composition) and 1.0 for a complete mismatch (all compounds present in one sample are absent in the other sample and vice-versa). The program first asks about the samples to be run. The total number of samples to be printed in the table is requested. Next, sample numbers are requested - contiguous ranges of sample numbers should be inputted before sample numbers which aren't contiguous (ranges are easier to input than entering every individual sample number). Information on which compounds to use is inputted next. If all compounds in a matrix are to be used then a matrix c is filled in ascending order with all the compound numbers. If all compounds in a matrix aren't to be used then the ranges and compound numbers are inputted, sorted in ascending order and placed into matrix c. All the requested samples are checked to see if at least one of the requested compounds is present and if not, a message is printed stating none of the compounds are present in that sample (the sample is then excluded from further calculations and printouts).

The program outputs a table of dissimilarity indices. A reference sample number is requested (can be any sample -isn't limited to those requested above). The DIs printed are all comparisons between this reference sample and the above requested (comparison) samples. The reference may be an oil and the comparison samples may be oils (oil/oil comparisons) or rocks (oil/rock comparisons). The reference may be a rock sample and the comparison samples may be rocks (rock/rock comparisons) or oils (oil/rock comparisons). The current date, compound matrix name, reference sample number and name are printed at the beginning of the table. The table consists of a sorted list of

DIs in ascending order together with the corresponding sample number and name. The difference between adjacent DIs is printed under the heading of DELTA. Printing the DIs in ascending order is useful because members of the same group as the reference will generally be together at the top of the list. Printing the delta values is sometimes helpful in determining boundaries between groups. Table 6 contains the DIs of 39 hopanes in all the rock samples compared to oil sample 2. Note the large delta value (.249) between samples 14 and 17. It shows the division between the Shublik/Kingak group, which is more closely associated with oil sample 2, and the Pebble/Torok group. The two exceptions (samples 18 and 23) have been discussed previously (Kvenvolden and others).

REFERENCES

Bray, E.E. and Evans, E.D., 1961, Distribution of n-paraffins as a clue to recognition of source beds, Geochemica et Cosmochemica Acta. 22, pp. 2-15.

Kvenvolden, K.A., Rapp, J.B., and Bourell, J.H., Comparison of molecular markers in crude oils and rocks from the north slope of Alaska, America Association of Petroleum Geologists Special Publication, in press.

Montalvo, J.G., 1978, The analytical chemist as a third party fact finder, Analytical Chemistry, 50(14), pp. 1330-1336A.

Scalan, R.S. and Smith, J.E., 1970, An improved measure of the odd-even predominance in the normal alkanes of sediment extracts and petroleum, Geochemica et Cosmochimica Acta, 34, pp. 611-620.

TABLE CAPTIONS

Table 1. Contents of matrix NS-NAME (sample numbers and sample names). (numbers following rock names are depths (in feet) below the surface)

Table 2. Input and output of (simple) Ratio subprogram.pris/n-C_{17}, phyt/n-C_{18} and pris/phyt in oil samples: 1. in order of requested samples and 2. sorted according to ascending value of pris/phyt.

Table 3. Input and output of (multiple) Ratio subprogram. Diasteranes/Steranes in oil samples: 1. in order of requested samples and 2. sorted in ascending order.

Table 4. Input and output of Carbon Preference Index subprogram. Carbon Preference Index (in the n-C_{16} to n-C_{28} range) of oil and rock samples.

Table 5. Input and output of Odd-Even Predominance subprogram. Odd-Even Predominance centering at n-C_{25} of oil and rock samples.

Table 6. Input and output of Dissimilarity Index subprogram. Dissimilarity Indices of 39 hopanes in rock samples compared to oil sample 2.

Table 7. Contents of matrices HOPANE-STR and STERANE-STR (compound numbers and abbreviated compound names).

Table 8. Compound numbers and names of the terpanes and steranes used in the study.

SAMPLES

OILS	1	Prudhoe - Put River
	2	South Barrow 20
	3	South Barrow 19
	4	Fish Creek 1
	5	Simpson Core Test 4
	6	Seabee 1 DST 3
	7	Dalton 1 DST 2
	8	Cape Simpson SL
	24	Umiat 4
ROCKS	9	Pebble shale 6361
	10	Pebble shale 1977
	11	Kingak Shale 14577
	12	Fort. Mtn. Fm 12014
	13	Sadlerochit Gp 13831
	14	Shublik Fm 12273
	15	Kingak Shale 11704
	16	Torok Fm 7054
	17	Pebble shale 6405
	18	Shublik Fm 10273
	19	Pebble shale 7371
	20	Torok Fm 3792
	21	Shublik Fm 9042
	22	Torok Fm 4994
	23	Kingak Shale 7390

Table 1

```
run
Name of matrix file=? alkane-ht
Ratio(r),Dissimilarity Index(d),Cpi/Oep(c),New Matrix(m),or Stop(s)? r
Number of samples in table =? 9
Input contiguous ranges before non-contig. samples
Are samples contiguous(c) or not(n)? c
What range(low,high)? 1,8
Are samples contiguous(c) or not(n)? n
Sample Number =? 24
Number of ratios in table(9 max.)=? 3
Multiple Numerator/Denominators(y/n)? n
Ratio  1  numerator compound number=? 30
Ratio  1  denominator compound number=? 6
Ratio  2  numerator compound number=? 31
Ratio  2  denominator compound number=? 7
Ratio  3  numerator compound number=? 30
Ratio  3  denominator compound number=? 31

date=01/16/84  matrix is alkane-ht
```

Sample Number	Sample Name	pris n-C17	phyt n-C18	pris phyt
1	Prudhoe - Put River	0.388	0.360	1.444
2	South Barrow 20	0.419	0.349	1.556
3	South Barrow 19	0.273	0.223	1.565
4	Fish Creek 1	0.382	0.267	1.750
5	Simpson Core Test 4	0.453	0.180	3.545
6	Seabee 1 DST 3	0.438	0.200	14.000
7	Dalton 1 DST 2	0.343	0.400	1.000
8	Cape Simpson SL	**	**	**
24	Umiat 4	0.458	0.248	2.200

```
Print table with ratios sorted(y/n)? y
Sort according to which ratio? 3

date=01/16/84  matrix is alkane-ht
```

Sample Number	Sample Name	pris n-C17	phyt n-C18	pris phyt
7	Dalton 1 DST 2	0.343	0.400	1.000
1	Prudhoe - Put River	0.388	0.360	1.444
2	South Barrow 20	0.419	0.349	1.556
3	South Barrow 19	0.273	0.223	1.565
4	Fish Creek 1	0.382	0.267	1.750
24	Umiat 4	0.458	0.248	2.200
5	Simpson Core Test 4	0.453	0.180	3.545
6	Seabee 1 DST 3	0.438	0.200	14.000
8	Cape Simpson SL	**	**	**

```
Another sort(y/n)? n
```

Table 2

```
Name of matrix file=? sterane
Ratio(r),Dissimilarity Index(d),Cpi/Oep(c),New Matrix(m),or Stop(s)? r
Number of samples in table =? 9
Input contiguous ranges before non-contig. samples
Are samples contiguous(c) or not(n)? c
What range(low,high)? 1,8
Are samples contiguous(c) or not(n)? n
Sample Number =? 24
Number of ratios in table(9 max.)=? 1
Multiple Numerator/Denominators(y/n)? y
Maximum Numerators or Denominators is 9 each.
Number of numerators in ratio 1 =? 4
   Numerator 1  =? 1
   Numerator 2  =? 2
   Numerator 3  =? 10
   Numerator 4  =? 13
Number of denominators in ratio 1 =? 4
   Denominator 1  =? 9
   Denominator 2  =? 12
   Denominator 3  =? 20
   Denominator 4  =? 23
Ratio 1
         Numerators=27baSD+27baRD+29baSD+29baRD
         Denominators=27aaaSS+27aaaRS+29aaaSS+29aaaRS

date=01/16/84  matrix is sterane

Sample     Sample                RATIO
Number      Name                   1

   1   Prudhoe - Put River        1.736
   2   South Barrow 20            1.389
   3   South Barrow 19            1.518
   4   Fish Creek 1               1.842
   5   Simpson Core Test 4        2.230
   6   Seabee 1 DST 3             **
   7   Dalton 1 DST 2             1.495
   8   Cape Simpson SL            2.349
  24   Umiat 4                    3.325


Print table with ratios sorted (y/n)? y
Sort according to which ratio? 1

date=01/16/84  matrix is sterane

Sample     Sample                RATIO
Number      Name                   1

   2   South Barrow 20            1.389
   7   Dalton 1 DST 2             1.495
   3   South Barrow 19            1.518
   1   Prudhoe - Put River        1.736
   4   Fish Creek 1               1.842
   5   Simpson Core Test 4        2.230
   8   Cape Simpson SL            2.349
  24   Umiat 4                    3.325
   6   Seabee 1 DST 3             **


Another sort (y/n)? n
```

Table 3

```
Ratio(r),Dissimilarity Index(d),CPi/OEP(c),New Matrix(m),or Stop(s)? m
Name of matrix file=? alkane-ht
Ratio(r),Dissimilarity Index(d),CPi/OEP(c),New Matrix(m),or Stop(s)? c
CPI(c) or OEP(p)? c
What range?C(even) to C(even)? 16,28
Number of samples in table =? 9
Input contiguous ranges before non-contig. samples
Are samples contiguous(c) or not(n)? c
What range(low,high)? 1,8
Are samples contiguous(c) or not(n)? n
Sample Number =? 24

date=01/16/84  matrix is alkane-ht

Sample    Sample                  CPI
                                n_C16_28   Missing
Number      Name               Center=22   Alkanes

   8   Cape Simpson SL                        16
  24   Umiat 4                    0.99
   4   Fish Creek 1               1.00
   3   South Barrow 19            1.04
   7   Dalton 1 DST 2             1.05
   1   Prudhoe - Put River        1.05
   5   Simpson Core Test 4        1.07
   2   South Barrow 20            1.08
   6   Seabee 1 DST 3           ( 3.36)         20

Other samples with same range (y/n)? y
Number of samples in table =? 15
Input contiguous ranges before non-contig. samples
Are samples contiguous(c) or not(n)? c
What range(low,high)? 9,23

date=01/16/84  matrix is alkane-ht

Sample    Sample                  CPI
                                n_C16_28   Missing
Number      Name               Center=22   Alkanes

  21   Shublik Fm 9042            0.80
  11   Kingak Shale 14577         0.85
  12   Fort. Mtn. Fm 12014        0.94
  14   Shublik Fm 12273           1.00
  13   Sadlerochit Gp 13831       1.02
  18   Shublik Fm 10273           1.03
  23   Kingak Shale 7390          1.04
  19   Pebble shale 7371          1.04
  17   Pebble shale 6405          1.06
  15   Kingak Shale 11704         1.09
   9   Pebble shale 6361          1.11
  20   Torok Fm 3792              1.13
  16   Torok Fm 7054              1.14
  22   Torok Fm 4994              1.15
  10   Pebble shale 1977          1.27

Other samples with same range (y/n)? n
```

Table 4

```
Ratio(r),Dissimilarity Index(d),Cpi/Oep(c),New Matrix(m),or Stop(s)? c
CPI(c) or OEP(p)? p
OEP centering around what carbon number? 25
Number of samples in table =? 9
Input contiguous ranges before non-contig. samples
Are samples contiguous(c) or not(n)? c
What range(low,high)? 1,8
Are samples contiguous(c) or not(n)? n
Sample Number =? 24

date=01/16/84  matrix is alkane-ht

Sample      Sample                  OEP        Missing
Number       Name                Center=25    Alkanes

   8    Cape Simpson SL                          23
   6    Seabee 1 DST 3                           23
   7    Dalton 1 DST 2            0.74
   3    South Barrow 19           0.93
   1    Prudhoe - Put River       0.94
   2    South Barrow 20           1.00
  24    Umiat 4                   1.00
   5    Simpson Core Test 4       1.01
   4    Fish Creek 1              1.06

Other samples with same range(y/n)? y
Number of samples in table =? 15
Input contiguous ranges before non-contig. samples
Are samples contiguous(c) or not(n)? c
What range(low,high)? 9,23

date=01/16/84  matrix is alkane-ht

Sample      Sample                  OEP        Missing
Number       Name                Center=25    Alkanes

  18    Shublik Fm 10273          0.72
  12    Fort. Mtn. Fm 12014       0.88
  21    Shublik Fm 9042           0.95
  11    Kingak Shale 14577        0.95
  13    Sadlerochit Gp 13831      0.97
  14    Shublik Fm 12273          0.97
  23    Kingak Shale 7390         1.02
  17    Pebble shale 6405         1.04
  19    Pebble shale 7371         1.05
  15    Kingak Shale 11704        1.09
   9    Pebble shale 6361         1.15
  16    Torok Fm 7054             1.20
  22    Torok Fm 4994             1.23
  20    Torok Fm 3792             1.29
  10    Pebble shale 1977         1.57

Other samples with same range(y/n)? n
```

Table 5

```
Ratio(r),Dissimilarity Index(d),CPi/OEP(c),New Matrix(m),or Stop(s)? m
Name of matrix file=? hopane
Ratio(r),Dissimilarity Index(d),CPi/OEP(c),New Matrix(m),or Stop(s)? d
Number of samples in table =? 15
Input contiguous ranges before non-contig. samples
Are samples contiguous(c) or not(n)? c
What range(low,high)? 9,23
Use all(a) compounds in matrix or not(n)? a
Total number of compounds to be used=? 39

date=01/16/84  matrix is hopane

Reference Sample Number(May be other than inputted samples)=? 2

DISSIMILARITY INDEX OF SAMPLE  2    South Barrow 20

Sample      Sample            Diss.    Delta
Number       Name             Index

  21    Shublik Fm 9042       0.094
  15    Kingak Shale 11704    0.111    0.016
  11    Kingak Shale 14577    0.111    0.001
  12    Fort. Mtn. Fm 12014   0.152    0.041
  14    Shublik Fm 12273      0.197    0.045
  17    Pebble shale 6405     0.445    0.249
  23    Kingak Shale 7390     0.451    0.006
  13    Sadlerochit Gp 13831  0.470    0.018
  19    Pebble shale 7371     0.521    0.051
  22    Torok Fm 4994         0.536    0.015
   9    Pebble shale 6361     0.542    0.006
  18    Shublik Fm 10273      0.550    0.008
  20    Torok Fm 3792         0.556    0.006
  16    Torok Fm 7054         0.558    0.003
  10    Pebble shale 1977     0.658    0.100


Another Reference Sample for D.I.(y/n)? n
Ratio(r),Dissimilarity Index(d),CPi/OEP(c),New Matrix(m),or Stop(s)?
```

Table 6

TERPANES

1	C19 TC
2	C20 TC
3	C21 TC
4	C22 TC
5	C23 TC
6	C24 TC
7	C25 TC
8	A
9	C26TC(S)
10	C26TC(R)
11	B'
12	B
13	C28TC(S)
14	C28TC(R)
15	C29TC(S)
16	C29TC(R)
17	Ts
18	Tm'
19	Tm
20	C
21	C'
22	D
23	D'
24	C29H
25	E
26	C29M
27	C30H
28	C30M
29	C31H(S)
30	C31H(R)
31	Gam?
32	C31M
33	C32H(S)
34	C32H(R)
35	C32M
36	C33H(S)
37	C33H(R)
38	C34H(S)
39	C34H(R)

STERANES

1	27baSD
2	27baRD
3	27abSD
4	27abRD
5	28baSD
6	A
7	B
8	28baRD
9	27aaaSS
10	29baSD
11	C
12	27aaaRS
13	29baRD
14	D
15	E
16	28aaaSS
17	28abbRS
18	28abbSS
19	28aaaRS
20	29aaaSS
21	29abbRS
22	29abbSS
23	29aaaRS

Table 7

Terpanes (m/z 191)

Peak No.	Compound	
1.	C19 tricyclic terpane	
2.	C20 tricyclic terpane	
3.	C21 tricyclic terpane	
4.	C22 tricyclic terpane	
5.	C23 tricyclic terpane	
6.	C24 tricyclic terpane	
7.	C25 tricyclic terpane	
8.	A C24 tetracyclic terpane	
9.	C26 tricyclic terpane (S?)	
10.	C26 tricyclic terpane? (R?)	
11.	B' tricyclic terpane?	
12.	B tricyclic terpane?	
13.	C28 tricyclic terpane (S?)	
14.	C28 tricyclic terpane (R?)	
15.	C29 tricyclic terpane (S?)	
16.	C29 tricyclic terpane (R?)	
17.	18α(4),21β(H)-22,29,30-trisnorhopane (Ts)	(C27)
18.	Tm' (C30) triterpane?	
19.	17α(H),21β(H)-22,29,30-trisnorhopane (Tm)	(C27)
20.	C (C28) triterpane?	
21.	C' triterpane?	
22.	D triterpane?	
23.	D' (C29) triterpane?	
24.	17α(H),21β(H)-30-normoretane	(C29)
25.	E triterpane?	
26.	17β(H),21α(H)-30-normoretane	(C29)
27.	17α(H),21β(H)-hopane	(C30)
28.	17β(H),21α(H)-moretane	(C30)
29.	17α(H),21β(H)-30-homohopane (22S)	(C31)
30.	17α(H),21β(H)-30-homohopane (22R)	(C31)
31.	Gammacerane?	(C30)
32.	17β(H),21α(H)-30-homomoratane	(C31)
33.	17α(H),21β(H)-30,31-bishomohopane (22S)	(C32)
34.	17α(H),21β(H)-30,31-bishomohopane (22R)	(C32)
35.	17β(H),21α(H)-30,31-bishomomoretane	(C32)
36.	17β(H),21β(H)-30,31,32-trishomohopane (22S)	(C33)
37.	17β(H),21β(H)-30,31,32-trishomohopane (22R)	(C33)
38.	17β(H),21β(H)-30,31,32,33?-tetrakishomohopane (22S)	(C34)
39.	17β(H),21β(H)-30,31,32,33?-tetrakishomohopane (22R)	(C34)

Steranes (m/z 217)

Peak No.	Compound	
1.	13β(H),17α(H) diacholestane (20S)	(C27)
2.	13β(H),17α(H) diacholestane (20R)	(C27)
3.	13α(H),17β(H) diacholestane (20S)	(C27)
4.	13α(H),17β(H) diacholestane (20R)	(C27)
5.	13β(H),17α(H) diaergostane (20S)	(C28)
6.	A sterane?	
7.	B sterane?	
8.	13β(H),17α(H) diaergostane (20R)	(C28)
9.	5α(H),14α(H),17α(H) cholestane (20S)	(C29)
10.	13β(H),17α(H) diastigmastane (20S)?	(C29)
11.	C Sterane?	
12.	5α(H),14α(H),17α(H) cholestane (20R)	(C27)
13.	13β(H),17α(H) diastigmastane (20R)	(C29)
14.	D sterane?	
15.	E sterane?	
16.	5α(H),14α(H),17α(H) ergostane (20R)	(C28)
17.	5α(H),14β(H),17β(H) ergostane (20R)	(C28)
18.	5α(H),14β(H),17β(H) ergostane (20S)	(C28)
19.	5α(H),14α(H),17α(H) ergostane (20R)	(C28)
20.	5α(H),14α(H),17α(H) stigmastane (20S)	(C29)
21.	5α(H),14β(H),17β(H) stigmastane (20R)	(C29)
22.	5α(H),14β(H),17β(H) stigmastane (20S)	(C29)
23.	5α(H),14α(H),17α(H) stigmastane (20R)	(C29)

Table 8

APPENDIX

Listing of Computer Program OIL/ROCK.BASIC

```
10 marsin 132
20 dim c(40),d(9),f(25,2),m(25,40),n(9),r(25,9),s(25)
30 print"Name of matrix file=";
40 input f$
50 file #1:f$
60 reset #1:0
70 mat read #1:m
80 let c$=f$&"-str"
90 file #2:c$
100 file #3:"ns-name"
110 reset #3:0
120 print "Ratio(r),Dissimilarity Index(d),Cpi/Oep(c),New Matrix(m),or Stop(s)"
\c;
130 input p$
140 if seg$(p$,1,1)="r" then 460
150 if seg$(p$,1,1)="c" then 1390
160 if seg$(p$,1,1)="d" then 2410
170 if seg$(p$,1,1)="s" then 4820
180 if seg$(p$,1,1)="m" then 30
190 go to 120
200 print"Number of samples in table =";
210 input s
220 print"Input contiguous ranges before non-contig. samples"
230 let x=1
240 print"Are samples contiguous(c) or not(n)";
250 input m$
260 if m$="c" then 290
270 if m$="n" then 400
280 go to 240
290 rem contiguous samples
300 print"What range(low,high)";
310 input l1,h1
320 if l1<h1 then 340
330 go to 300
340 for y=l1 to h1
350 let s(x)=y
360 let x=x+1
370 next y
380 if x<=s then 240
390 go to 450
400 rem non-contiguous samples
410 print"Sample Number =";
420 input s(x)
430 let x=x+1
440 if x<=s then 410
450 return
460 rem Ratio Program
470 gosub 200
480 print"Number of ratios in table(9 max.)=";
490 input r
500 if r>9 then 480
510 mat r=zer
520 print"Multiple Numerator/Denominators(y/n)";
530 input p$
540 if p$="y" then 3810
550 if p$="n" then 570
560 go to 520
570 for x=1 to r
580 print"Ratio ";x;" numerator compound number=";
590 input n(x)
600 print"Ratio ";x;" denominator compound number=";
```

```
610 input d(x)
620 next x
630 for x=1 to r
640 for y=1 to s
650 if m(s(y),d(x))<>0 then 680
660 let r(s(y),x)=9999
670 go to 690
680 let r(s(y),x)=m(s(y),n(x))/m(s(y),d(x))
690 next y
700 next x
710 gosub 910
720 gosub 1240
730 print"Print table with ratios sorted(y/n)";
740 input p$
750 if p$="n" then 120
760 gosub 1160
770 gosub 3390
780 gosub 910
790 gosub 1240
800 print"Another sort(y/n)";
810 input p$
820 if p$="y" then 760
830 if p$<>"n" then 800
840 print"Use same ratios with different samples(y/n)";
850 input p$
860 if p$="n" then 120
870 if p$="y" then 890
880 go to 840
890 gosub 200
900 go to 630
910 rem Print SIMPLE RATIO heading
920 print
930 print"date=";dat$,"matrix is ";f$
940 print
950 print"Sample     Sample              ";
960 for x=1 to r
970 reset #2:n(x)-1
980 read #2:n$
990 print using"  <#######",n$;
1000 next x
1010 print
1020 print tab(30);
1030 for x=1 to r
1040 print"  ------  ";
1050 next x
1060 print
1070 print "Number       Name                ";
1080 for x=1 to r
1090 reset #2:d(x)-1
1100 read #2:n$
1110 print using"  <#######",n$;
1120 next x
1130 print
1140 print
1150 return
1160 print"Sort accordins to which ratio";
1170 input m
1180 if m<=r then 1200
1190 go to 1160
1200 for x=1 to s
```

```
1210 let f(x,2)=r(s(x),m)
1220 next x
1230 return
1240 for n=1 to s
1250 reset #3:s(n)-1
1260 read #3:s$
1270 print using"  -#    <####################    ",s(n),s$;
1280 for x=1 to r
1290 if r(s(n),x)<>9999 then 1320
1300 print"     **     ";
1310 go to 1330
1320 print using" -###.### ",r(s(n),x);
1330 next x
1340 print
1350 next n
1360 print
1370 print
1380 return
1390 rem CPI/OEP Calculations
1400 if seg$(f$,1,6)="alkane" then 1430
1410 print "WRONG MATRIX"
1420 go to 30
1430 print"CPI(c) or OEP(p)";
1440 input p$
1450 if p$="p" then 2040
1460 if p$="c" then 1480
1470 go to 1430
1480 print"What range?C(even) to C(even)";
1490 input l,h
1500 if 2*int(l/2)<l then 1530
1510 if 2*int(h/2)<h then 1530
1520 go to 1550
1530 print"Input even numbers only"
1540 go to 1480
1550 mat f=zer
1560 gosub 200
1570 print
1580 print"date=";dat$,"matrix is ";f$
1590 print
1600 print"Sample";tab(10);"Sample";tab(33);"CPI"
1610 print tab(31);
1620 print using"n_C-#_-#   Missing",l,h
1630 print"Number      Name";tab(30);
1640 print using"Center=-#   Alkanes",(l+h)/2
1650 print
1660 for n=1 to s
1670 for i=l to l+2
1680 if m(s(n),i-11)<>0 then 1710
1690 let f(s(n),1)=i
1700 go to 1830
1710 next i
1720 let a=b=c=0
1730 for i=l to h-2 step 2
1740 if m(s(n),i-9)=0 then 1810
1750 let a=a+m(s(n),i-10)
1760 let b=b+m(s(n),i-11)
1770 let c=c+m(s(n),i-9)
1780 next i
1790 let f(n,2)=a/2*(1/b+1/c)
1800 go to 1830
```

```
1810 let f(s(n),1)=i+2
1820 let f(n,2)=a/2*(1/b+1/c)
1830 next n
1840 gosub 3390
1850 for n=1 to s
1860 reset #3:s(n)-1
1870 read #3:s$
1880 print using"  -#    <####################    ",s(n),s$;
1890 if f(s(n),1)=0 then 1930
1900 if f(n,2)=0 then 1950
1910 print using"(-#.##)        -#",f(n,2),f(s(n),1)
1920 go to 1970
1930 print using" -#.##",f(n,2)
1940 go to 1970
1950 print tab(44);
1960 print using"-#",f(s(n),1)
1970 next n
1980 print
1990 print "Other samples with same range (y/n)";
2000 input p$
2010 if p$="y" then 1550
2020 if p$="n" then 120
2030 go to 1990
2040 print"OEP centering around what carbon number";
2050 input l
2060 let l=l-2
2070 mat f=zer
2080 gosub 200
2090 print
2100 print"date=";dat$,"matrix is ";f$
2110 print
2120 print"Sample";tab(11);"Sample";tab(33);"OEP";tab(42);"Missing"
2130 print"Number       Name";tab(30);
2140 print using"Center=-#    Alkanes",l+2
2150 print
2160 for n=1 to s
2170 for i=1 to l+4
2180 if m(s(n),i-11)<>0 then 2210
2190 let f(s(n),1)=i
2200 go to 2230
2210 next i
2220 let f(n,2)=((m(s(n),l-11)+6*m(s(n),l-9)+m(s(n),l-7))/(4*m(s(n),l-10)+4*m(s
\c(n),l-8)))^(-1^(l+1))
2230 next n
2240 gosub 3390
2250 for n=1 to s
2260 reset #3:s(n)-1
2270 read #3:s$
2280 print using"  -#    <####################    ",s(n),s$;
2290 if f(s(n),1)=0 then 2330
2300 print tab(44);
2310 print using"-#",f(s(n),1)
2320 go to 2340
2330 print using" -#.##",f(n,2)
2340 next n
2350 print
2360 print"Other samples with same range(y/n)";
2370 input p$
2380 if p$="y" then 2070
2390 if p$="n" then 120
2400 go to 2360
```

```
2410 rem dissimilarity index
2420 gosub 200
2430 print"Use all(a) compounds in matrix or not(n)";
2440 input p$
2450 print"Total number of compounds to be used=";
2460 input n
2470 if p$="a" then 2500
2480 if p$="n" then 2540
2490 go to 2430
2500 for x=1 to n
2510 let c(x)=x
2520 next x
2530 go to 2850
2540 let x=1
2550 print"Input contiguous ranges before non-contig. compounds"
2560 print"Contiguous Compounds (y/n)";
2570 input p$
2580 if p$="y" then 2660
2590 if p$="n" then 2610
2600 go to 2560
2610 print"Compound Number=";
2620 input c(x)
2630 let x=x+1
2640 if x>n then 2750
2650 go to 2610
2660 print"What range (low,high)";
2670 input l,h
2680 if l<h then 2700
2690 go to 2660
2700 for y=l to h
2710 let c(x)=y
2720 let x=x+1
2730 next y
2740 if x<=n then 2560
2750 rem sort in increasing order
2760 for x=n to 2 step -1
2770 for y=x-1 to 1 step -1
2780 if c(y)<c(x) then 2820
2790 let u=c(y)
2800 let c(y)=c(x)
2810 let c(x)=u
2820 next y
2830 next x
2840 print
2850 rem create normalized matrix of all samples
2860 for y=1 to 25
2870 let t=0
2880 for x=1 to n
2890 let t=t+m(y,c(x))
2900 next x
2910 if t<>0 then 3090
2920 let m(y,0)=1
2930 for x=1 to n
2940 let m(y,x)=0
2950 next x
2960 let x=1
2970 if s(x)=y then 3010
2980 let x=x+1
2990 if x<=s then 2970
3000 go to 3120
```

```
3010 reset #3:s(y)-1
3020 read #3:n$
3030 gosub 3050
3040 go to 3120
3050 print
3060 print"NOTE-NONE OF THE INDICATED COMPOUNDS ARE PRESENT IN ";n$
3070 print
3080 return
3090 for x=1 to n
3100 let m(y,x)=m(y,c(x))/t
3110 next x
3120 next y
3130 print
3140 print"date=";dat$,"matrix is ";f$
3150 print
3160 print"Reference Sample Number(May be other than inputted samples)=";
3170 input a
3180 reset #3:a-1
3190 read #3:n$
3200 if m(a,0)=0 then 3230
3210 gosub 3050
3220 go to 3710
3230 print
3240 print"DISSIMILARITY INDEX OF SAMPLE ";a;"    ";n$
3250 print
3260 for y=1 to s
3270 let t=0
3280 for x=1 to n
3290 let t=t+abs(m(s(y),x)-m(a,x))
3300 next x
3310 let f(y,2)=t/2
3320 next y
3330 print"Sample      Sample            Diss.    Delta"
3340 print"Number       Name             Index"
3350 print
3360 gosub 3390
3370 let f=g=0
3380 go to 3520
3390 rem sort in increasing order
3400 for x=1 to s-1
3410 for y=x+1 to s
3420 if f(x,2)<f(y,2) then 3490
3430 let t=f(x,2)
3440 let u=s(x)
3450 let f(x,2)=f(y,2)
3460 let s(x)=s(y)
3470 let f(y,2)=t
3480 let s(y)=u
3490 next y
3500 next x
3510 return
3520 for x=1 to s
3530 if s(x)=a then 3680
3540 if m(s(x),0)=0 then 3570
3550 let g=g+1
3560 go to 3680
3570 if m(s(x-1),0)=1 then 3600
3580 let d=f(x,2)-f(x-1,2)
3590 go to 3610
3600 let d=f(x,2)-f(x-1-g,2)
```

```
3610 reset #3:s(x)-1
3620 read #3:n$
3630 if f=1 then 3670
3640 let f=1
3650 print using"  -#    <#################### -.###",s(x),n$,f(x,2)
3660 go to 3680
3670 print using"  -#    <#################### -.###    -.###",s(x),n$,f(x,2),d
3680 next x
3690 print
3700 print
3710 print"Another Reference Sample for D.I.(y/n)";
3720 input p$
3730 if p$="y" then 3160
3740 if p$="n" then 3760
3750 go to 3710
3760 rem must recreate matrix m
3770 for x=1 to 25
3780 let m(x,0)=0
3790 next x
3800 go to 60
3810 rem Multiple numerators/denominators Ratios
3820 rem Fill numerator/denomerator matrix
3830 print"Maximum Numerators or Denominators is 9 each."
3840 for x=1 to r
3850 print"Number of numerators in ratio";x;"=";
3860 input n
3870 if n>9 then 3850
3880 let r(10,x)=n
3890 for y=1 to n
3900 print"   Numerator";y;" =";
3910 input r(y,x)
3920 next y
3930 print"Number of denominators in ratio";x;"=";
3940 input d
3950 if d>9 then 3930
3960 let r(20,x)=d
3970 for y=1 to d
3980 print"   Denominator";y;" =";
3990 input r(y+10,x)
4000 next y
4010 next x
4020 for x=1 to r
4030 print"Ratio";x
4040 print"          Numerators=";
4050 for y=1 to r(10,x)
4060 reset #2:r(y,x)-1
4070 read #2:n$
4080 if y=r(10,x) then 4110
4090 print n$;"+";
4100 go to 4120
4110 print n$
4120 next y
4130 print"          Denominators=";
4140 for y=1 to r(20,x)
4150 reset #2:r(y+10,x)-1
4160 read #2:n$
4170 if y=r(20,x) then 4200
4180 print n$;"+";
4190 go to 4210
4200 print n$
```

```
4210 next y
4220 next x
4230 rem fill ratio matrix
4240 for x=1 to r
4250 let n=r(10,x)
4260 for y=1 to n
4270 let n(y)=r(y,x)
4280 next y
4290 let d= r(20,x)
4300 for y=1 to d
4310 let d(y)=r(y+10,x)
4320 next y
4330 for y=1 to s
4340 let s=k=0
4350 for z=1 to n
4360 let s=s+m(s(y),n(z))
4370 next z
4380 for z=1 to d
4390 let k=k+m(s(y),d(z))
4400 next z
4410 rem test denominator for zero
4420 if k<>0 then 4450
4430 let r(s(y),x)=9999
4440 go to 4460
4450 let r(s(y),x)=s/k
4460 next y
4470 next x
4480 rem print table according to inputted sample order
4490 gosub 4660
4500 gosub 1240
4510 print"Print table with ratios sorted (y/n)";
4520 input p$
4530 if p$="y" then 4560
4540 if p$="n" then 120
4550 go to 4510
4560 rem print table according to one sorted ratio
4570 gosub 1160
4580 gosub 3390
4590 gosub 4660
4600 gosub 1240
4610 print"Another sort (y/n)";
4620 input p$
4630 if p$="n" then 120
4640 if p$="y" then 4570
4650 go to 4610
4660 rem multiple num/dem heading
4670 print
4680 print"date=";dat$,"matrix is ";f$
4690 print
4700 print"Sample     Sample              ";
4710 for x=1 to r
4720 print"   RATIO  ";
4730 next x
4740 print
4750 print"Number      Name               ";
4760 for x=1 to r
4770 print using"      -     ",x;
4780 next x
4790 print
4800 print
4810 return
4820 end
```