

MURRAP User Guide

by Michael D. Glascock (1 January 2021)

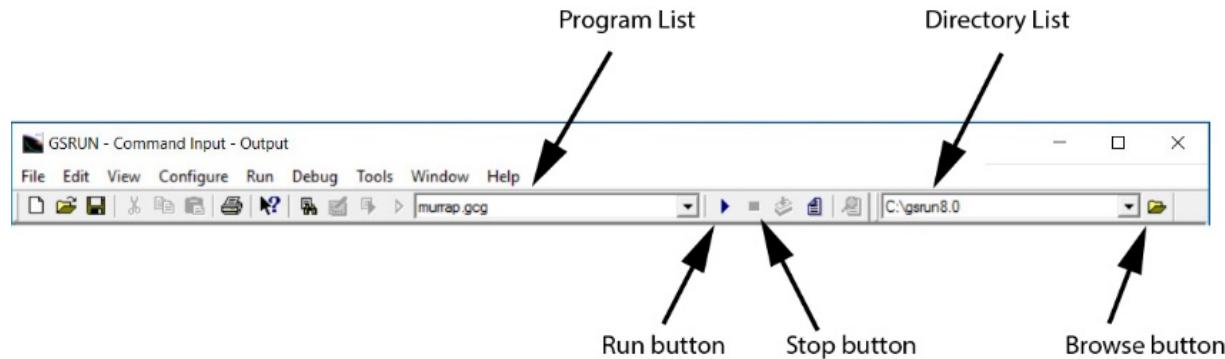
Introduction

These programs are based on the original FORTRAN routines written in the 1970s by Edward Sayre at the Brookhaven National Laboratory. For more information, see Sayre's article available online "Brookhaven Procedures for Statistical Analyses of Multivariate Archaeometric Data", BNL-21693. Available at <http://www.osti.gov/energycitations>. In the 1980s, the routines were translated into Gauss language by Hector Neff. Revisions to make the routines compatible with the Gauss Run-time GUI and additional programming (including help pages) were made by William Grimm in 2000. Since 2008, all changes to the routines were made by Michael Glascock. The freely distributed version of the program name is now called **MURRAP.GCG**.

The Gauss Run-time GUI (graphical user interface) is a stand-alone freeware product produced by Aptech, Inc. (www.aptech.com). A help file is included in the GUI for instructions on using it. Particular aspects of the operation of the GUI are included in the following instructions as needed.

Getting Started

You must have received this program by downloading it from the Archaeometry Lab at MURR webpages or directly from Mike Glascock at MURR to be reading this now. You have therefore already installed the Gauss Run-time as well as unzipped the MURR routines. The important things to know about the Gauss GUI are on the toolbar:



The Directory List: The directory or folder showing in this list is the one currently used by GAUSS to find and save data files for its routines. When you start GAUSS, this will be the directory containing the GAUSS program. If your data files are in another directory, you will need to use the **Browse** button to locate it. The directory list will retain your most recently used directories. We recommend using a separate directory for each project.

The Program List: This contains a list of Gauss routines that you have already run. You can select one of them and click on the arrow to the right to run it. If it is the first time you run GAUSS, the Program List may be blank. If you type: `run murrep.gcg` after the GAUSS command prompt (`>>`) in the Command Input window below the Toolbar and press the Enter key -- this creates a menu for the MURR routines in the Command Input Window. You should now have `murrep.gcg` in the Program List.

The Run Button When a program is visible in the Program List, clicking the Run button will run it.

The Stop Button When a program is running, the Stop button becomes active and can be clicked to stop the program (if the program is waiting for input you may then need to press the Enter or Return key). If you stop the program (or if it stops because it encountered an error) you can restart it by pressing the Run button while the program's name is visible in the Program List window.

Using the MURRAP Menu

Running `MURRAP.GCG` displays a menu of options consisting of the MURR statistical routines. Entering an option number after the "`==>`" prompt selects the option. A parentheses appearing before the prompt contains the default option which will be selected by pressing the **Enter key** without entering anything else. In most of the routines, entering "quit" or "0" will return you to the main MURRAP menu. To exit MURRAP while at the main menu, press the Enter key and you will be returned to the GAUSS prompt (`>>`), from which you can close the GUASS program by selecting "**Exit**" from the "File" menu. You can also click on the **Stop** button in the menu bar, but then you may also have to press the **Enter key** before Stop takes effect.

When you enter the name of a dataset or file, do not include the file extension, except when specifying a text output file, for which you will probably want to add the `.txt` extension.

Data Files

The data files used by this program are in a Gauss proprietary format. A routine has been included to import data from Microsoft Excel. Files imported from Excel must be formatted in a manner similar to the following:

ANID	Site Name	Ceramic Type	As	La	Sm	Cr
DMG0001	Sand Canyon	Mesa Verde Black on White	7.04	49.6	7.94	64.9
DMG0002	Sand Canyon	Mesa Verde Black on White	5.20	36.6	5.97	45.5
DMG0077	McElmo	McElmo Alluvial Clay	10.42	31.4	5.70	54.3
DMG0078	McElmo	McElmo Alluvial Clay	10.58	41.8	7.28	59.6

The first row should contain the sample name header (**ANID**) and the variable names used such as (Site name, Ceramic Type, As, La, Sm, Eu, etc.).

[Note: GAUSS is case-sensitive regarding elemental abbreviations, so CA is not treated as equivalent to Ca. You may also have to be sure spaces have been deleted in the Excel files because they may not be stripped during conversion]

Starting with the second row the spreadsheet cells **must** consist of the following:

Column A (**ANID**): Text column containing the sample names (not to exceed eight alphanumeric characters).

Columns B, C, etc.: Either text columns containing descriptions or Numeric columns containing elemental concentration data (the latter can only contain positive numbers between 0.0 and 1,000,000).

If you edited the Excel file (removed columns or rows), be sure to save it and **always close Excel** before importing the file into Gauss. The first blank row encountered will terminate the input process.

Making Selections

At various places through the MURRAP you are expected to select from a list of variables or element names. The element names are preceded by a number (e.g., 1 for Ca, 2 for Ba, etc.). You can select by entering the name or the number. In general, entering numbers is much easier.

When selecting multiple variables, you can enter:

1. "all" to select all of the variables or datasets in the list.
2. "1-5, 9-12" to select variables 1 thru 5 and 9 thru 12.
3. "not 6,8" to select all variables except numbers 6 and 8.

Saving Your Results

Some of the MURRAP routines offer the option of saving the results to a file, others do not. When results are displayed with no option to save to a file, use the "Save As" command under the "File" menu to save a text file containing what is in the command input-output window below the task bar. The contents of the window can also be copied and pasted as with text in a word processor. Tables displayed as text in the window can be pasted into an Excel cell and converted to rows and columns using Excel's "Text to Columns" option under its "Data" menu.

Quitting a Routine

Some of the MURRAP routines offer options for quitting and returning to the main menu. Typing "quit" or "0" will do the same in most routines. If an error occurs in a routine, you may need to click the Stop button and press the Enter key stop the routine. The "File" menu contains an "Exit" option, which may also be accomplished using the Ctrl and Q keyboard combination.

Missing Values

Routines are included for listing missing values ([option 3.3](#)) or replacing them ([option 2.5](#)), and some routines also offer the option to do this before performing an analysis.

These routine use the Mahalanobis distance metric to substitute approximations for the missing values in your dataset. It attempts to minimize the Mahalanobis distance between the sample with missing data and the centroid of the dataset. If the number of missing values in a dataset is excessive (e.g., greater than 50%), the replacement will be unreliable. Also, do not use this routine if you have more than one group represented in the dataset in which case the approximations will be skewed.

Log Transformation

The concentration data in the GAUSS datasets are automatically transformed to base-10 log values by each routine before any operations on the data are performed. This practice is common when analyzing compositional data obtained from archaeological specimens due to the advantages of logratios when interpreting geological compositional data. See Aitchinson (1999) for a more detailed explanation.

Graphics (TKF) Files

Plots produced by MURRAP will be displayed in one or more separate "TKF File Viewer" windows. In general, saving TKF files in TKF format is not recommended because they are not viewable by most graphics programs. The "Convert" menu of the TKF window gives several options for saving in a different graphic format: enhanced metafile, encapsulated postscript, HPGL plotter, or Windows bitmap. The Windows bitmap conversion produces a very low-resolution file — if you want to work with a better bitmap version, you can convert the TKF file to an enhanced metafile and paste it into another graphics program (e.g., Adobe Illustrator, Corel Draw, etc.) to save the enhanced metafile as a higher-resolution bitmap and then edit the plot further.

The TKF "Help" menu opens a text file containing a description of the viewer's features, including options for changing some colors in the TKF window itself ("View").

Another option is to save the file in PDF format using the "Print" function. The output PDF file is readable by most graphics programs and the resolution is generally better.

Contact Information

If you have questions not be answered by these help files, feel free to contact Mike Glascock at MURR.

Dr. Michael D. Glascock
Research Reactor Center
University of Missouri
Columbia, MO 65211
Email: GlascockM@missouri.edu
Laboratory Home Page: www.missouri.edu/~glascock/archlab.htm

1. Import/export/concatenate/extract datasets

Options:

```
-----  
0 Go back to the Main Menu  
1 Import data from Excel  
2 Export data to Excel  
3 Concatenate datasets  
4 Extract subgroups from a Master dataset  
-----
```

1.1. Import data from Excel

This routine allows you to convert an Excel file to a Gauss dataset. Refer to the [Data Files](#) section of the Introduction to this document to find out what the Excel data structure should look like (some routines require slightly different structures, so please refer to the specific routine you plan to use for any exceptions to these rules). When selected, this routine displays a numbered list of .xls or .xlsx files in the current directory, or displays "There are no Excel files in the current folder", in which case you must select the appropriate directory using the Directory List in the Taskbar.

- **Select file:** You will be asked to select a file from the list to be imported, by entering its number or file name as listed (without the extension .xls or .xlsx).
- **Select Worksheet:** When a file is selected, you will be asked to select a worksheet number in the file to import by entering its number. The single worksheet in a simple Excel file with only one spreadsheet is number 1. If you wish to go back to the main menu, enter "quit".

When the worksheet has been selected, the routine imports data from that worksheet as a set of variables as defined by the Excel file.

- **List variables with descriptive information:** You may list the variables with descriptive information, or if there are none, enter "none"r.
- **Create the GAUSS dataset:** Next enter the name of the dataset to be created (the default, in parentheses, will be the name of the Excel file which you imported). The routine will then create a .dat file of the name you enter, in the current directory. You will then be asked whether you want to import another Excel file, and may answer N or Y.

1.2. Export data to Excel

This routine allows you to convert a Gauss dataset to an Excel file of the same structure. When selected, it displays a numbered list of the dataset files (.dat) in the current directory, with number of samples in each file and total number of samples in the directory. This procedure may be faster if Excel is open.

- **Select a dataset:** Enter one of the datasets listed, by number or name.
- **Enter a filename:** Enter a name for the Excel file to be created. Pressing the Enter key without entering a name creates a file with the default name (which is in parentheses). If that filename already exists, you will be asked to either replace the existing file with the one you are exporting, or to store the exported data in the next available worksheet. Be patient, this procedure may take longer than you expect, even for small datasets. To return to the main menu without making one of these choices, enter "quit".

1.3. Concatenate datasets

This routine provides an easy way to put several Gauss datasets of the same structure together quickly. This operation could also be done using either Excel or the Dataset Manipulator ([option 4](#)), but this routine makes it much simpler. Make sure you have the same structure in all of the datasets you wish to concatenate.

- **Select the datasets to concatenate:** You may type ALL to concatenate all datasets in the current directory or type the number (from the list given) of the first of two datasets you wish to concatenate. After entering the first, you will be given a prompt to enter the second dataset.
- **Enter name of dataset to be created**

1.4. Extract subgroups from a Master dataset

A list of datasets in the working directory will be given, with the number of samples in each. Descriptive categories must be present in a dataset in order for subgroups to be extracted.

- **Select the Master dataset** from which subgroups will be extracted, entering either number or name of a descriptive categories from the given list. If here are no descriptive categories present in the dataset which you have entered, you will be informed, and will have to start over.
- **Select a category** to search for subgroups: If there are one or more descriptive categories present in the dataset which you have entered, the will be displayed in a numbered list, and you can enter name or number to search that category for subgroups. The groups present in that category will then be displayed in a numbered list.
- **Select group names** to use when extracting subgroups: Enter a name or number from the numbered list of groups present in the selected category.
- **Enter name of dataset to be created:** This will name the dataset containing the group selected.

2. Modify/sort/transform variables

Options:

- 0 Go back to the Main Menu
- 1 [View variables in datasets](#)
- 2 [Sort element variables according to atomic number](#)
- 3 [Remove variables from datasets](#)
- 4 [Rename variables in datasets](#)
- 5 [Replace missing values](#)
- 6 [Modify entries for descriptive variables](#)
- 7 [Adjust dataset for dilution effect from calcium temper](#)
- 8 [Add temper to a dataset of clays](#)
- 9 [Convert element concentrations into ratios](#)
- 10 [Un-log legacy GAUSS 5.0 data from base 10](#)

2.1. View variables in datasets

All of the datasets in the working directory will be displayed in a numbered list with the number of samples in each.

- **Select datasets in which to view variables:** All or a specific dataset may be selected by number or name. You may enter several datasets by number, separated by commas, at the prompt or enter "All". If entering anything but "All", you will need to press the Enter key a second time at the next prompt to display the variables. Entering another dataset at the second prompt does not include that dataset in the routine.

2.2. Sort element variables according to atomic number

All of the datasets in the working directory will be displayed in a numbered list with the number of samples in each.

- **Select datasets in which to sort variables:** All or a specific dataset may be selected by number or name. You may enter several datasets by number, separated by commas, at the prompt or enter "All". If entering anything but "All", you will need to press the Enter key a second time at the next prompt to sort the variables. Entering another dataset at the second prompt does not include that dataset in the routine. Elements will be sorted in atomic order followed by the descriptive variables in alphabetical order.
- **Enter name to be used for the sorted dataset.** You also have the option of replacing the original dataset file with the sorted one, using the same name, by entering no new name. Use routine 1 (View variables), above, to view the results.

2.3. Remove variables from datasets

All of the datasets in the working directory will be displayed in a numbered list with the number of samples in each. After one or more datasets are selected a list of variables in these datasets will be displayed: you may choose one or more variables to remove from ALL selected datasets (if different variables are to be removed from one or another dataset, repeat this option with each dataset).

- **Select datasets from which to remove variables:** All or a specific dataset may be selected by number or name. You may enter several datasets by number, separated by commas, at the prompt or enter "All". If entering anything but "All", you will need to press the Enter key a second time at the next prompt to replace the variables. Entering another dataset at the second prompt does not include that dataset in the routine. The current variable names in the selected datasets will be displayed.
- **Select variable(s) to be removed:** Enter one or more variables by name or number, separated by commas.
- **Enter name to be used for the modified dataset:** You also have the option of replacing the original dataset file with the sorted one, using the same name, by entering no new name. Use routine 1 (View variables), above, to view the results.

2.4. Rename variables in datasets

All of the datasets in the working directory will be displayed in a numbered list with the number of samples in each.

- **Select datasets in which to rename variables:** All or a specific dataset may be selected by number or name. You may enter several datasets by number, separated

by commas, at the prompt or enter "All". If entering anything but "All", you will need to press the Enter key a second time at the next prompt to replace the variables. Entering another dataset at the second prompt does not include that dataset in the routine. The current variable names in the selected datasets will be displayed.

- **Select variable(s) to be renamed:** Enter one or more variables by name or number, separated by commas. you will need to press the Enter key a second time at the next prompt to continue. Entering another variable at the second prompt does not include that variable in the routine.
- **Enter new name(s) for selected variable(s):** You will be asked to rename each of the selected variables, one at a time.
- **Enter name(s) of modified dataset(s):** You also have the option of replacing the original dataset file with the sorted one, using the same name, by entering no new name. Use routine 1 (View variables), above, to view the results.

2.5. Replace missing values

All of the datasets in the working directory will be displayed in a numbered list with the number of samples in each.

- **Select datasets to replace missing values or type ALL**
- **Enter name of modified dataset** The name of the selected dataset is the default; it will be replaced if no other name is entered.

2.6. Modify entries for descriptive variables

This routine allows you to modify descriptive variables in a dataset. You may choose the following options:

1. Modify a descriptive variable in a single sample
2. Modify a descriptive variable in a series of samples
3. Modify several descriptive variables in a single sample

2.7. Adjust dataset for dilution effect from calcium temper

This option performs a dilution correction to data for limestone- or shell-tempered pottery by removing calcium. After the correction, the element Sr will also be removed. WARNING: if Ca is greater than 40%, negative numbers will occur. if Ca values are < 1%, no adjustments will be made.

You should identify the type of calcium temper present in your pottery based on the following options:

1. Fresh-water shell
2. Salt-water shell
3. Limestone

After selecting the type of temper present, you have the option of correcting for the elements Mn and Na that are often present in shell-tempered pottery. Based on a publication by Cogswell, et. al. 1998, correction assumes that fresh-water shell has Mn = 578 ppm & Na = 1488 ppm. Salt-water shell and limestone have different Mn and Na concentrations depending upon the particular region of the world. The program allow you to substitute different concentrations for Mn and Na, if known.

If this correction results in negative values for Mn or Na, they are replaced by zeros, and you will be given the option of removing the element from further calculations.

2.8. Add temper to a dataset of clays

This option allows you to simulate tempered pottery by diluting individual samples in a clay dataset using values derived from a separate temper dataset. Be sure you have the same variables in the same order in both datasets before using this procedure. The routine computes an average chemical composition for the temper dataset and uses this average to dilute samples in the clay dataset.

You will be asked to select a temper dataset, select a clay dataset, enter a number for the percentage of temper (between 0 and 100) to be added to each clay sample, and enter the name of the mixed tempered clay dataset to be created.

For example, a dataset named MIX25 can be created by mixing 25% of dataset TEMPER and 75% of dataset CLAY by entering the responses:

```
Select a temper dataset =====> TEMPER
Select a clay dataset =====> CLAY
Enter number between 0 and 100 for percentage of temper ===> 25
Enter name of mixed clay&temper dataset to be created ===> MIX25
```

2.9. Convert element concentrations into ratios

This option allows you to convert element concentrations in one or more datasets to ratios of each element to a selected element in the dataset. Before using this routine, you may want to check for missing values ([Option 3.3](#)) and make replacements ([Option 2.5](#)), especially if the element you wish to use as the denominator may be missing values in some samples.

After selecting the datasets, you will select one element to be used as a denominator in converting to ratios in all datasets selected. For each modified dataset you may enter a new name, (the default is to overwrite the original dataset).

2.10. Un-log legacy GAUSS 5.0 data from base 10

Older datasets which were converted to base-10 logarithms should be un-logged before being used in the latest version of murrap.gcg. The latest version automatically converts ppm concentrations to base-10 logarithms where appropriate during each routine.

3. List sample ANIDs/concentrations/descriptions

Options:

- 0 Go back to Main Menu
- 1 [List sample ANIDs in each dataset](#)
- 2 [List sample concentrations & descriptions](#)

3.1. List sample ANIDS in each dataset

All of the datasets in the working directory will be displayed in a numbered list. After selecting one or more datasets, the ANIDS (in the ANID column) for each sample in the chosen datasets will be listed.

3.2. List sample concentrations & descriptions

All of the datasets in the working directory will be displayed in a numbered list. After selecting one or more datasets, a list of elements in these datasets is displayed, from which you may select one or more. You may choose to list missing value replacements or not. Rounded concentration values for the selected elements in each selected dataset will be displayed in tabular form as shown below.

Values for selected variables in dataset ABC are:

ANID	Mn	Fe	Rb	Sr	Y	Zr
FMNV01	543.87	16209	154.91	11.400	54.209	390.39
FMNV02	563.43	14842	144.76	--	55.917	372.77
FMNV03	701.00<	15726	159.31	12.800	53.509	388.05
FMNV04	573.13	17430	156.19	12.000	61.254	417.90
FMNV05	561.03	16270	150.60	--	54.423	389.01
FMNV06	546.68	16103	157.29	12.000	63.054	416.53
FMNV07	479.03	16644	149.35	11.500	+DEN<	393.04
FMNV08	562.03	16450	163.05	12.100	63.021	411.58
Average	566.27	16209	154.43	8.975	50.673	397.41
	+/-	+/-	+/-	+/-	+/-	+/-
1-Sigma	61.82	742	5.90	5.556	20.865	16.13

< indicates the concentration deviates more than 2-sigma from the mean

* indicates the concentration is a calculated replacement value

Note: -- indicates missing values and +DEN indicates presence of a non-numeric value.

4. Manipulate datasets

This displays datasets in the working directory and gives the option of starting the Dataset Manipulator (MANIP.EXE), a separate program which allows one to transfer samples to and from GAUSS datasets. A warning will be given if the datasets in the working directory are not consistent: i.e. do not have the same variables in the same order. However, one may still press the ENTER key to start the Dataset Manipulator; for example, if one wants to transfer a subset of samples from one dataset to a new dataset, the consistency of other datasets does not matter. (The Data Manipulator may also be started independent of MURRAP by opening MANIP.EXE in the GSRUN folder.) When the Dataset Manipulator opens it will NOT show files from the current working directory of the GAUSS runtime. You must change the working directory using the Data Manipulator's "Change" button. There is a help file, DMHelp.chm, in the GAUSS folder. It may not open automatically when you click the Dataset Manipulator's "Help" button.

5. Summary statistics

After selecting one or more datasets in the working directory, one is given the option of printing the summary statistics to the GAUSS window or exporting them to an Excel file. Exporting to Excel will be much faster if Excel is already open. The Excel file will be saved in the working directory. The Excel file does not automatically open after export. Completion of the exporting is indicated in the GAUSS window.

An example of the output from this routine is shown below:

Element	Normal Distribution				Lognormal Distribution				Min	Max	N/18
	Mean	SD	%SD	Fit	Mean	SD	Fit				
Na	10136.69	1266.77	12.5	72.2	10059.65	1.14	72.2	7419.86	12539.61	18	
Al	72745.66	4984.87	6.9	61.1	72583.51	1.07	61.1	63896.98	81465.54	18	
K	16886.53	2129.07	12.6	66.7	16767.18	1.13	66.7	14120.10	21952.64	18	
Ca	9773.55	1634.12	16.7	61.1	9649.56	1.18	61.1	7700.40	12669.84	18	
Sc	13.42	1.15	8.6	83.3	13.37	1.09	83.3	10.46	15.33	18	
Ti	4005.16	333.20	8.3	72.2	3992.13	1.09	72.2	3471.52	4683.85	18	
V	117.18	15.95	13.6	77.8	116.14	1.15	77.8	84.91	148.84	18	
Cr	328.53	91.58	27.9	83.3	317.76	1.30	77.8	193.55	600.93	18	
Mn	830.81	169.37	20.4	66.7	811.77	1.26	77.8	459.38	1133.16	18	
Fe	39112.89	4676.45	12.0	77.8	38827.74	1.14	83.3	26214.43	49367.23	18	
Co	20.25	3.96	19.6	66.7	19.91	1.20	66.7	14.69	31.20	18	
Ni	91.43	30.90	33.8	83.3	87.48	1.34	77.8	53.91	174.93	18	
Zn	73.38	9.89	13.5	77.8	72.74	1.15	77.8	54.06	96.79	18	
As	19.13	10.64	55.6	66.7	16.52	1.78	66.7	5.55	45.82	18	
Rb	87.32	9.90	11.3	77.8	86.79	1.12	77.8	67.12	107.89	18	
Sr	162.63	67.37	41.4	77.8	151.70	1.45	66.7	97.91	297.07	18	
Zr	129.32	16.79	13.0	72.2	128.27	1.14	72.2	99.24	161.47	18	
Sb	0.64	0.30	47.5	83.3	0.59	1.46	77.8	0.40	1.44	18	
Cs	4.49	0.65	14.4	66.7	4.44	1.16	66.7	3.37	5.47	18	
Ba	665.21	140.15	21.1	50.0	650.92	1.24	50.0	433.05	851.50	18	
La	27.49	1.95	7.1	61.1	27.42	1.07	61.1	23.85	30.91	18	
Ce	58.58	3.65	6.2	61.1	58.47	1.07	61.1	50.89	63.87	18	
Nd	23.62	1.60	6.8	66.7	23.57	1.07	66.7	20.21	26.11	18	
Sm	5.03	0.28	5.6	72.2	5.02	1.06	77.8	4.40	5.36	18	
Eu	1.12	0.08	7.3	72.2	1.12	1.08	72.2	0.93	1.25	18	
Tb	0.71	0.06	9.1	66.7	0.71	1.09	66.7	0.63	0.87	18	
Dy	4.11	0.28	6.8	72.2	4.10	1.07	72.2	3.54	4.56	18	
Yb	2.47	0.18	7.4	72.2	2.46	1.08	72.2	2.15	2.87	18	
Lu	0.35	0.02	6.3	66.7	0.35	1.06	66.7	0.32	0.39	18	
Hf	5.75	0.84	14.7	83.3	5.69	1.17	83.3	3.83	7.10	18	
Ta	0.84	0.06	7.1	77.8	0.83	1.07	77.8	0.73	1.00	18	
Th	9.54	0.93	9.7	72.2	9.50	1.10	72.2	7.88	11.69	18	
U	2.32	0.54	23.2	72.2	2.27	1.23	72.2	1.77	3.63	18	

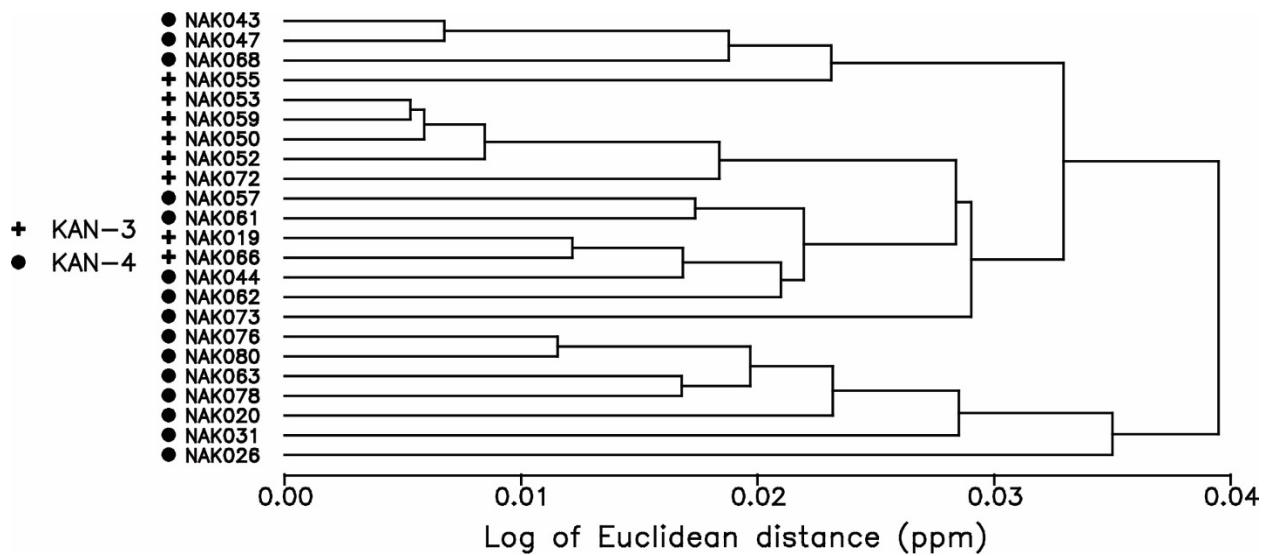
SD = Standard deviation

Fit = Percentage of samples within 1 SD

6. Hierarchical cluster analysis

This routine computes the squared mean Euclidean distance (SMED) between all samples and constructs a dendrogram using the average linkage algorithm. The routine does not automatically replace missing values in a dataset, but considers only non-zero values in the dataset. If you want the routine to consider replacements for missing values, so be sure to replace before you run this routine otherwise the results may be unsatisfactory.

You may select one or more datasets in the working directory. Each dataset will be given a separate symbol and color in the resulting dendrogram. All samples in a single dataset will have the same color. After selecting the elements to be included in the cluster analysis, the dendrogram is displayed in a separate "[TKF File Viewer](#)" window with has options for saving the file in other formats. A dendrogram is shown below:



Be reminded that the interpretation of cluster analysis is because it fails to consider correlations between elements. Cluster analysis is a valuable tool, but should NEVER be consider the final answer for most archaeometric investigations.

7. Principal components analysis

Principal components analysis (PCA) is a statistical procedure that uses orthogonal transformation to convert a set of observations of potentially correlated variables into a set of uncorrelated variables can principal components (PCs). The transformation is defined in such a way that the first PC explains the largest possible variance in the dataset, and subsequent PCs explain the highest amount of remaining variance under the constraint that they be orthogonal to all preceding components. PCA is useful as an exploratory tool to reveal the internal structure of the data. Because the earliest PCs explain a greater portion of the variance, PCA is often used as a dimension-reduction technique.

This routine performs a PCA on a set of one or more Gauss datasets using the variance-covariance matrix. After choosing the datasets to include in the analysis (the datasets must be consistent in structure — same variables in same order), you may choose the elements that will be included.

In order for PCA to work properly, the number of samples must be equal to or greater than the number of elements considered. The percentage of variance explained by each principal component is displayed in a table along with the cumulative variance. You have the option of viewing the PC matrix and the option of saving it as an Excel file (this file will be saved in the working directory). You can also save the PC transformation matrix (*.fmt), which can be used later for creating principal component plots and for calculating probabilities for group membership using Mahalanobis distance calculations (this matrix file is also saved in the working directory). You can also save the PC scores in an Excel file. The name of the Excel file is assigned by the program and will be saved in the working directory.

A typical output from PCA is shown below:

PERCENT OF VARIANCE EXPLAINED BY THE PRINCIPAL COMPONENTS

PC	%Variance	Cumulative	PC	%Variance	Cumulative
1	50.71	50.71	18	0.19	99.14
2	16.01	66.71	19	0.17	99.31
3	8.92	75.63	20	0.12	99.43
4	6.21	81.83	21	0.11	99.54
5	4.21	86.04	22	0.09	99.62
6	2.67	88.71	23	0.07	99.69
7	2.25	90.96	24	0.06	99.75
8	2.00	92.96	25	0.06	99.81
9	1.32	94.29	26	0.05	99.85
10	0.96	95.24	27	0.03	99.89
11	0.85	96.10	28	0.03	99.91
12	0.81	96.90	29	0.02	99.94
13	0.64	97.55	30	0.02	99.96
14	0.51	98.06	31	0.02	99.98
15	0.37	98.42	32	0.02	99.99
16	0.28	98.71	33	0.01	100.00
17	0.25	98.95			

PCA RESULTS BASED ON VARIANCE-COVARIANCE MATRIX

The first 7 PCs explain 91.0% of the variance in this assemblage.

Element	PC1	PC2	PC3	PC4	PC5	PC6	PC7
	-50.71%	-16.01%	-8.92%	-6.21%	-4.21%	-2.67%	-2.25%
As	-1.18E-01	-3.48E-01	-7.23E-01	2.26E-01	4.75E-01	-1.12E-01	4.73E-02
Mn	-3.11E-01	4.90E-01	-2.10E-01	-1.24E-01	7.63E-02	3.99E-01	3.43E-01
Ca	-8.62E-02	1.42E-01	-4.28E-01	-1.64E-01	-4.27E-01	2.60E-01	-3.96E-01
Ni	-3.75E-01	1.75E-01	1.48E-01	2.27E-03	2.04E-01	-2.28E-01	-5.46E-01
Sr	-7.62E-02	4.83E-02	-2.60E-01	-3.36E-03	-5.15E-01	-4.13E-01	-1.19E-01
Cr	-5.56E-01	-6.95E-02	2.18E-01	-5.95E-02	1.91E-01	-1.94E-01	-1.01E-01
Sb	-7.95E-02	-1.13E-02	-7.29E-02	5.49E-01	-1.91E-01	-3.70E-02	1.22E-01
Cs	-1.20E-01	-1.97E-02	8.14E-02	3.99E-01	-2.28E-01	6.50E-02	2.12E-01
Na	3.79E-02	2.97E-01	-6.94E-02	-1.15E-01	-3.52E-02	-2.73E-01	2.15E-01
U	4.20E-02	9.51E-02	4.11E-02	3.83E-01	-8.74E-02	-8.01E-02	-2.02E-01
Hf	5.00E-02	-3.79E-02	1.33E-02	8.62E-02	1.18E-01	3.48E-01	-2.45E-01
Th	2.52E-01	2.15E-01	2.35E-03	1.01E-01	1.44E-01	-2.36E-01	-2.39E-02
Co	-3.33E-01	2.28E-01	-5.23E-02	-3.27E-02	3.44E-02	-1.40E-02	1.51E-01
Zr	3.49E-02	5.13E-03	1.44E-02	1.13E-01	5.58E-02	3.17E-01	-2.08E-01
Ba	3.86E-02	1.68E-02	-2.44E-01	-1.44E-01	-7.86E-03	-1.02E-01	-1.62E-01
K	1.87E-01	1.53E-01	-9.97E-02	-7.20E-02	8.14E-02	-1.44E-01	6.25E-02
Rb	1.79E-01	1.54E-01	1.19E-02	-1.25E-02	1.14E-01	-1.61E-01	7.32E-02
Ta	1.11E-01	6.79E-02	7.12E-02	2.30E-01	3.57E-02	6.59E-02	-1.25E-01
Ti	-1.19E-01	-1.08E-02	5.92E-02	2.22E-01	-8.89E-02	1.06E-01	-3.90E-02
Nd	1.04E-01	2.01E-01	-4.96E-02	8.83E-02	1.14E-01	-3.57E-02	-1.05E-01
V	-2.16E-01	-3.97E-02	-1.21E-03	1.16E-01	-6.48E-02	2.60E-02	1.38E-01
Sm	9.24E-02	2.04E-01	-3.36E-02	6.80E-02	1.14E-01	-2.25E-02	-8.67E-02
La	1.08E-01	1.91E-01	-2.17E-02	1.22E-01	5.65E-02	-4.55E-03	-6.33E-02
Zn	-3.39E-02	1.85E-01	-4.59E-02	5.61E-02	3.89E-02	-1.48E-01	4.02E-02
Ce	1.03E-01	1.92E-01	-2.43E-02	6.25E-02	7.10E-02	4.28E-03	-6.51E-02
Tb	5.61E-02	1.97E-01	-2.39E-02	5.80E-02	1.02E-01	1.77E-02	-4.20E-02
Dy	4.92E-02	1.69E-01	-1.05E-02	8.26E-02	9.72E-02	5.20E-02	-8.01E-02
Yb	3.84E-02	1.35E-01	-2.91E-03	9.91E-02	1.03E-01	9.33E-02	-5.55E-02
Eu	-5.32E-02	1.75E-01	-3.76E-02	1.15E-01	-2.17E-02	-4.47E-03	1.47E-02
Fe	-1.34E-01	3.45E-02	2.73E-02	1.13E-01	-2.22E-02	-8.22E-02	9.19E-02
Lu	3.01E-02	1.09E-01	-1.57E-04	1.38E-01	7.30E-02	6.58E-02	-6.90E-02
Sc	-9.64E-02	6.02E-02	5.46E-02	6.46E-02	2.89E-02	-7.92E-02	3.57E-02
Al	3.66E-02	8.03E-02	-1.15E-03	7.82E-02	-1.04E-02	-1.02E-01	2.75E-02
--							
Eigenvalue	0.3986	0.12583	0.07009	0.04878	0.03307	0.02097	0.01771

8. Canonical discriminant analysis

CDA is a dimension-reduction technique similar to principal components analysis. Given two or more groups of observations with measurements on several variables, canonical discriminant analysis derives a linear combination of the variables that has the highest possible multiple correlation with the groups. This maximal multiple correlation is called the first canonical correlation. The coefficients of the linear combination are the canonical coefficients or canonical weights. The variable defined by the linear combination is the first canonical variable or canonical component. The second canonical correlation is obtained by finding the linear combination uncorrelated with the first canonical variable that has the highest possible multiple correlation with the groups. The process of extracting canonical variables is repeated until the number of canonical variables equals the number of original variables or the number of classes minus one, whichever is smaller.

The minimum requirements to perform CDA are:

1. Two or more groups
2. At least two observations per group
3. The number of variables must be less than the number of samples by two.
4. No variables may be a linear combinations of another variable

This routine performs a canonical discriminant analysis (CDA) on two or more groups. Each of the groups must be in a separate dataset — these datasets must contain the same elements in the same order. If missing values would be a problem in the analysis they should be replaced before using this routine. After choosing two or more datasets, you may select variables to include in the analysis. If any variable is missing values in all samples in a group, it will automatically be removed from the analysis. The routine automatically limits the number of variables you can choose, based on the limitations inherent in the CDA method. A larger number of samples in the groups allows the use of more variables.

A table of the separation by the canonical discriminants (CDs) is produced, giving the percent separation for each CD and the cumulative percent separation. You can then select the number of CDs to include in the summary of results (the default is all) and can print this summary to the screen or save it in an Excel file. This summary of results is a table listing the contribution of each element (factor loadings) to each canonical discriminant function. Values are also given for Wilke's Lambda, approximate F, and p. You may save a CD transformation matrix which can later be used to produce a CD plot, or re-used in this routine to display the results again.

For example, assume we have four datasets that have a sufficient number of samples to perform CDA. The output might look something like this:

SEPARATION BY THE CANONICAL DISCRIMINANTS		
CD	%Separation	Cumulative
--	-----	-----
1	74.70	74.70
2	23.06	97.76
3	2.24	100.00

CANONICAL DISCRIMINANT ANALYSIS BASED ON 4 GROUPS:

Elemental contributions to the canonical discriminant functions.

Variable	Magnitude	CD1 (74.70%)	CD2 (23.06%)	CD3 (2.24%)
Fe	1.34	-1.24E0	-2.19E-1	4.59E-1
Sc	1.27	1.03E0	7.58E-3	-7.33E-1
Eu	1.18	-8.69E-1	6.91E-1	3.97E-1
Al	1.16	7.69E-1	7.53E-1	-4.29E-1
Sm	0.82	8.08E-1	1.10E-1	-1.10E-1
Cs	0.63	-3.48E-1	-4.46E-1	2.81E-1
Cr	0.60	-4.76E-1	3.47E-1	-1.15E-1
Rb	0.57	3.18E-2	4.75E-1	-3.05E-1
Sb	0.52	4.89E-1	3.08E-2	-1.71E-1
Hf	0.50	-4.89E-1	-9.93E-2	7.90E-2
Th	0.50	4.17E-1	-9.12E-2	2.52E-1
Lu	0.45	-3.93E-1	-1.59E-1	-1.40E-1
Na	0.39	-2.00E-1	2.97E-1	1.52E-1
Nd	0.37	1.20E-2	-2.69E-1	-2.60E-1
U	0.36	-2.16E-1	-2.50E-3	-2.92E-1
Ti	0.35	2.32E-3	-3.16E-1	1.58E-1
Zr	0.34	2.96E-1	-6.39E-2	1.59E-1
Tb	0.28	2.78E-1	1.70E-2	-4.74E-2
Co	0.27	8.73E-3	-1.43E-1	2.27E-1
Ba	0.26	-1.27E-1	-1.44E-1	1.81E-1
Ta	0.26	-1.41E-1	1.53E-1	1.48E-1
Dy	0.25	-1.06E-1	-9.13E-2	2.04E-1
Zn	0.24	-1.65E-2	1.58E-1	1.77E-1
Ca	0.22	2.03E-1	2.26E-2	-7.04E-2
Ce	0.21	-1.94E-1	2.10E-2	8.79E-2
Sr	0.21	-1.86E-1	-9.69E-2	1.66E-2
K	0.21	-1.48E-2	9.74E-2	1.81E-1
Ni	0.18	-1.57E-1	8.04E-2	1.84E-2
La	0.16	-1.19E-1	-1.10E-1	3.10E-2
Yb	0.16	1.65E-2	-1.51E-1	4.14E-2
Mn	0.11	-1.63E-2	4.05E-2	-1.00E-1
V	0.09	1.56E-3	-4.35E-2	-8.44E-2
As	0.08	2.81E-2	1.82E-2	6.79E-2

Wilk's lambda: 3.8593888e-006

Approx. F:
50.341833
p-value:
4.6368128e-047

Wilk's lambda is a measure of the percent variance in dependent variables not explained by differences in the independent variables. The scale ranges from 0 to 1 and a value near zero is ideal.

9. Euclidean distance search

This routine searches one or more reference dataset[s] for the closest samples to each of the unknown samples in a datasets, as measured by squared mean Euclidean distance. The individual calculated distance values are the average of the squared differences between the listed concentrations for all elements in the sample.

The concentrations values in the datasets are automatically transformed to base-10 log values during this calculation. Datasets chosen for this routine must have the same elements to avoid an error during the analysis. After choosing the unknown dataset containing one or more unknown samples, you may choose one or more reference datasets to search for samples compositionally closest to the unknowns. Elements completely missing values in any of the datasets are automatically removed. You can choose some or all of the remaining elements to be included in the calculations.

You may choose to list up to 20 [the default is 10] results for each of the unknown samples. You will not get an error if the number of samples you choose is smaller than the number of reference samples. The routine just lists as many as possible up to the indicated number.

You have the choice of listing the results by individual sample or in a summary matrix, and can either save these results in a text file (without displaying it on the screen) or display the results on the screen (in which case you can save them in a text file using the "Save As" option under the "File" menu).

The calculated distance values have no units. In our experience, the relative importance of individual distance values in identifying related samples is approximately as follows:

Distance values: <0.010 are excellent
 >0.010 and <0.015 are very good
 >0.015 and <0.020 are good
 >0.020 and <0.025 are fair
 >0.025 are poor

A sample output is shown below:

```
RESULTS OF SQUARED-MEAN EUCLIDEAN DISTANCE SEARCH
=====
Using the following transformation matrix: PCA-KAN.fmt
Results are based on the following variables:
PC1 PC2 PC3 PC4 PC5 PC6 PC7

Reference samples closest to: NAK019
ANID      Distance   Chem Group
-----
NAK079    0.0099    KAN-5
NAK071    0.0111    KAN-5
NAK048    0.0121    KAN-5
NAK045    0.0123    KAN-5
NAK067    0.0123    KAN-5
NAK054    0.0131    KAN-5
NAK077    0.0133    KAN-5
NAK023    0.0136    KAN-5
NAK064    0.0143    KAN-5
NAK014    0.0144    KAN-5
-----
```

10. Group membership probabilities

This routine produces a table of group membership probabilities based on the Mahalanobis distance of each sample from the centroid of the group. It assumes that you have previously assembled one or more reference groups of samples into separate datasets. The calculations assume that each reference dataset contains samples from only one reference group and that each reference group is contained in only one reference dataset. Other datasets can contain unrelated samples that you wish to compare with the reference groups, and you can compare the reference groups with each other. All datasets must contain the same elements in the same order, and it would be best to take care of missing values before using this routine if missing values would be a concern in your case. Concentration values are automatically converted to base-10 log values during the calculations.

Computation of Mahalanobis distance requires that your groups contain at least two more samples than variables (the more the better). When performing this calculation with reference groups containing a small number of samples, you may find it more useful to use principal components rather than elements as variables, because the first few principal components usually comprises much of the variance in the group. In this case you should first perform a principal component analysis on any reference group you will be using and save the transformation matrix in the same directory you will be using for the group membership calculations. If there are any transformation matrices in the working directory, you will be given the option to choose one to use in this calculation — otherwise it is assumed that you will be using elements as variables.

You will first choose one or more reference groups from a list of datasets in the working directory. If you choose more than one reference group, individual samples will be compared to all groups. Group membership for reference groups will be "jackknifed" which means that each sample will be removed from the host group before the distance and probability calculations are made. This means that the sample's values will not contribute to the group's mean or to its distribution. Samples located near the edge of a group are likely to have considerably reduced probabilities of membership to that group when using jackknifed calculations.

Next you may choose one or more datasets containing samples that are not part of a reference group — if you are choosing none, be sure to type "none" at the prompt; there is no default entry. You may also choose the same datasets that are used for reference groups here — if you do so the routine will also list them with non-jackknifed values for each sample's distance and membership probability.

Next you will choose which elements or principal components to use in the calculations; the routine automatically limits your choice based on the number of samples in the reference groups.

The result is one or more tables listing group membership probabilities (in percent) for each sample for each of the reference groups, and for each unknown sample for each of the reference groups. A column is included which suggesting the best group for each sample based on its highest membership probability over 0.001%. These tables may be saved as a text file by using the "Save As" option under the "File" menu.

You are finally given the following options for recalculation of the probabilities:

- 0 Go back to the Main Menu
- 1 Recalculate membership probabilities
- 2 Sort samples based on probability
- 3 Change variables used in classification
- 4 Assume very large groups
- 5 List Mahalanobis distance values on screen
- 6 Save results to an Excel file

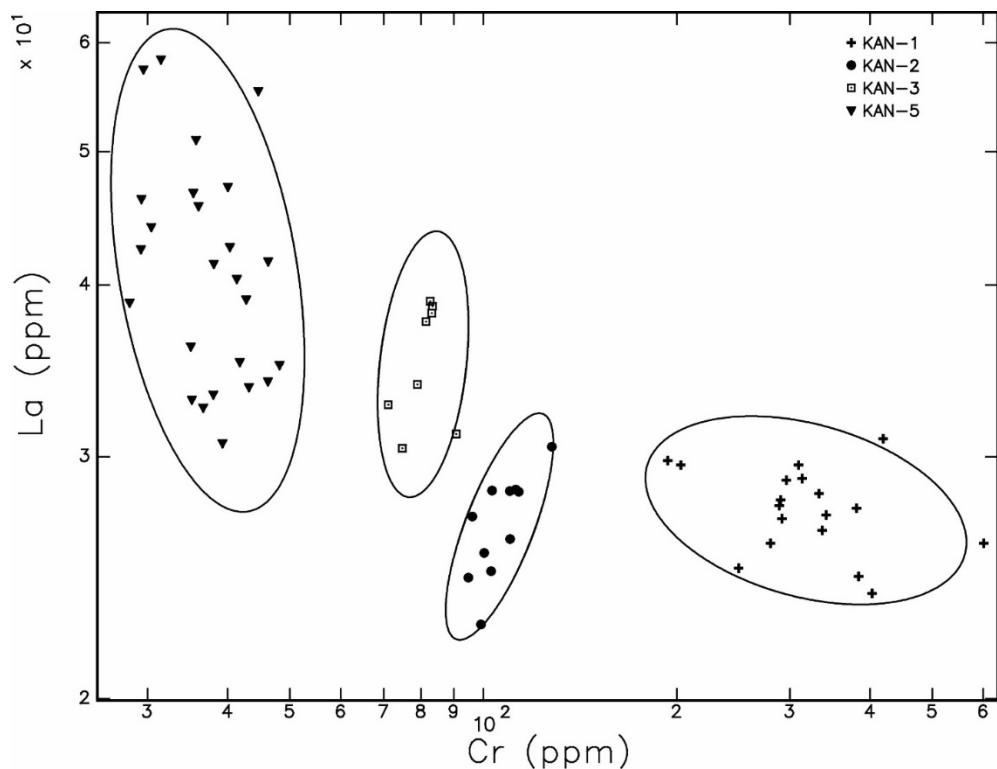
In particular, option 3 is useful for changing the elements or principal components used in the calculations, which may give very different results. This is recursive. When you make the desired change you are brought back to this list of options and must then choose "1" to see the recalculated results based on your changes. Option 5 will list the Mahalanobis distances from which the probabilities are derived. A sample output is shown below:

```
GROUP CLASSIFICATION USING MAHALANOBIS DISTANCE
=====
Using the following transformation matrix: PCA-KAN.fmt
Results are based on the following variables:
PC1 PC2 PC3 PC4 PC5 PC6
The first 6 PCs explain 88.7% of the variance
Best Group is based on highest membership probability > 0.001%

Membership probabilities(%) for samples in group: KAN-2
Probabilities calculated after removing each sample from group.
ANID      KAN-2      KAN-3  Best Group
-----
NAK003    28.262    0.141  KAN-2
NAK008    55.611    0.182  KAN-2
NAK012    37.079    0.205  KAN-2
NAK017    64.586    0.234  KAN-2
NAK032    2.566     0.231  KAN-2
NAK034    94.004    0.147  KAN-2
NAK035    33.858    0.141  KAN-2
NAK036    90.285    0.126  KAN-2
NAK037    20.442    0.163  KAN-2
NAK041    4.567     0.127  KAN-2
NAK042    98.689    0.139  KAN-2
-----
Membership probabilities(%) for samples in group: KAN-3
Probabilities calculated after removing each sample from group.
ANID      KAN-2      KAN-3  Best Group
-----
NAK019    0.908     20.283  KAN-3
NAK050    1.223     94.888  KAN-3
NAK052    1.408     29.805  KAN-3
NAK053    1.706     28.568  KAN-3
NAK055    0.380     10.227  KAN-3
NAK059    1.304     97.021  KAN-3
NAK066    1.375     11.843  KAN-3
NAK072    2.788     13.529  KAN-3
-----
```

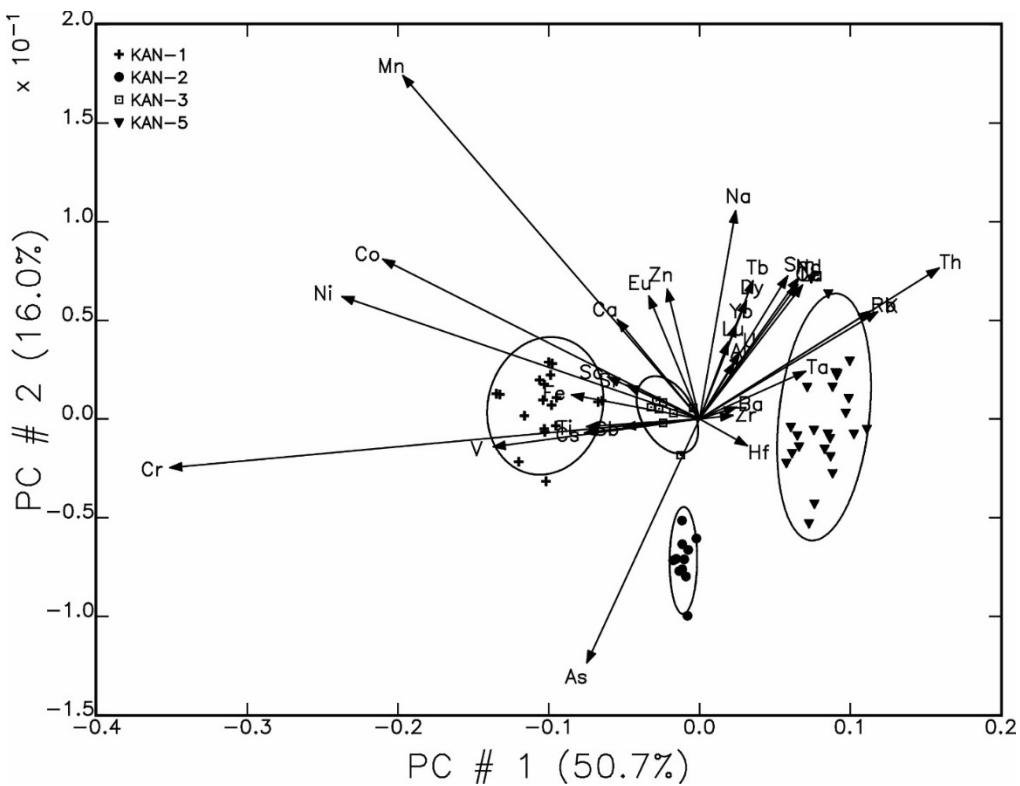
11. 2-D Compositional plots

This routine allows you to produce scatterplots (i.e. bivariate plots) of elements or PCs in Gauss datasets, with the option of plotting points, confidence intervals, labels, etc. You need to have all of the groups you want to plot in separate datasets, and all of the datasets must have the same variables in the same order. Each dataset will be given a separate color and symbol in the resulting plot, and the calculation of confidence ellipses assume a dataset is a valid group, so collect your samples into datasets accordingly. But you do not have to calculate confidence intervals for datasets containing unrelated samples if you wish only to display their points. A dataset can contain only a single sample if you wish it to have a unique color and symbol on the plot. A simple scatterplot for is shown below.



The figure shows a log-log scatterplot for elements Cr and La on samples of pottery from a site located in Turkey. The sample groups are surrounded by 90% confidence ellipses indicating that approximately 90% of the samples will be inside the ellipse. If desired, the confidence interval can be adjusted to another value.

If a principal component analysis (PCA) or canonical discriminant analysis (CDA) has been performed previously, its transformation matrix can also be used to produce what is known in the literature as a "biplot" (Neff 1994). An RQ-mode biplot contains both data points and vectors. The vectors display each element's contribution (both direction and relative magnitude) to the two principal components used in the plot. These vectors also display some information about the relationships of these elements one another. An RQ-mode biplot for the Turkish pottery used for illustration here is shown below.



Principal components 1 and 2 were used to create this plot. By definition, the PCs are orthogonal to one another. PC #1 explains 50.7% of the overall variance in the data and PC #2 explains 16.0% of the variance. The vectors indicate the direction and magnitude of increasing concentrations for each element. Confidence ellipses indicate the default 90% confidence interval.

You are asked to choose one or more datasets to display in the plot. Next you will select what components to display for each dataset. You can choose a different set of components for each dataset (the default is 3):

Available Plot Components

1 Plot Datapoints	5 Plot, ellipse, & label points
2 Ellipse only	6 Plot, ellipse & label outliers
3 Plot & ellipse	7 Plot Nothing
4 Plot & label points	

In this list of options "Plot" refers to datapoints: eg: "Plot & ellipse" displays the datapoints and the ellipse for a given dataset.

After selecting the component option for each dataset, you can specify the confidence interval to use for all datasets for which you have chosen to display one. Enter a number representing the percent but without the percent sign (the default is 90).

Next you will be given a list of variables in the datasets, either elements or principal components, and asked which to use for the X-axis and for the Y-axis. You can specify more than one variable for each axis: in this case a number of plots will be produced, each with only one of the variables for each axis.

You have the option of replacing missing values, if any, in all of the groups to be plotted (the default is yes). You should not use this option if you have any datasets containing unrelated samples that you are not treating as a group but have collected only for the purpose of displaying points, because the calculations for replacing missing values assume the samples in a given dataset are a valid group.

A plot will be displayed in a separate "[TKF File Viewer](#)" window. After the TKF plots have been created, the following options can be used to change various features of the plots:

```
0 Exit
1 Plot Again
```

Options for modifying plots are as follows:

```
-----
2 Change axis limits          7 Change plot elements or PCs
3 Change plot components      8 Change axis labels
4 Change plot symbols          9 Change axis from log to linear
5 Change plot colors          10 Change plot size
6 Change legend position
-----
```

The procedure for making these changes is recursive: when you have selected an option and specified the new values, you are returned to this list of options and must enter "1" (Plot Again) in order to produce a new TKF window with the changed plot. The TKF windows containing the previous plots are not altered or closed, so if you do make changes a few times you may want to close the TKF windows containing the previous plots, keeping in mind that they are not automatically saved. However, since the all the TKF windows are in the same screen position you can cycle through them easily by clicking on one after another in the taskbar, useful for comparing plots of different variables for the same groups.

Most of these options are obvious, and the default values are the current values, so you can choose each option in turn to see what is offered without fear of losing the current appearance of the plot. At any rate the current plot window is not altered when a change is made.

A few notes:

2. Change axis limits:

Pay close attention to the current limits (minimum and maximum values in ppm displayed on the X and Y axis, respectively) and make changes accordingly. Enter the cartesian coordinates in this order, with a comma between each value:

minimum X-value, maximum X-value, minimum Y-value, maximum Y-value

The value zero (0) is not allowed.

3. Change plot components:

This also allows you to change the size of the confidence interval used to draw ellipses.

7. Change plot elements:

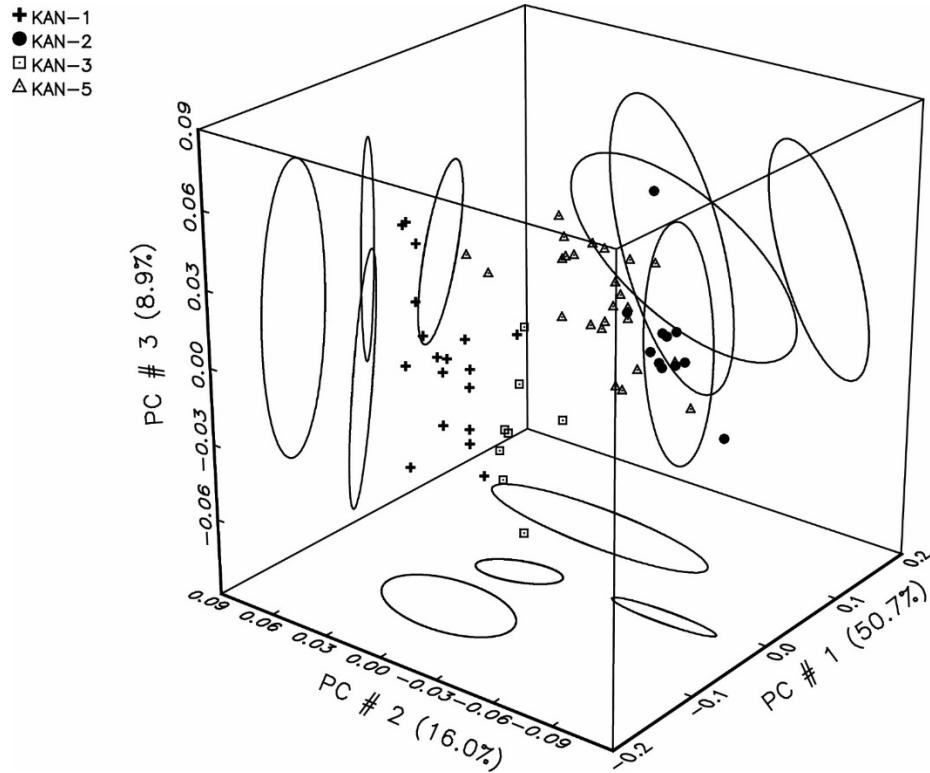
This refers to the variables used on the axes of the plots, either elements or principal components.

9. Change axis from log to linear:

The default is log values on both axes.

12. 3-D Compositional plots

This routine allows you to create scatterplots in three dimensions (trivariate plots). 3D plots are not commonly used in publication since they tend to be difficult to read. This routine is mainly intended for use as a data exploration tool. Here is an output example:



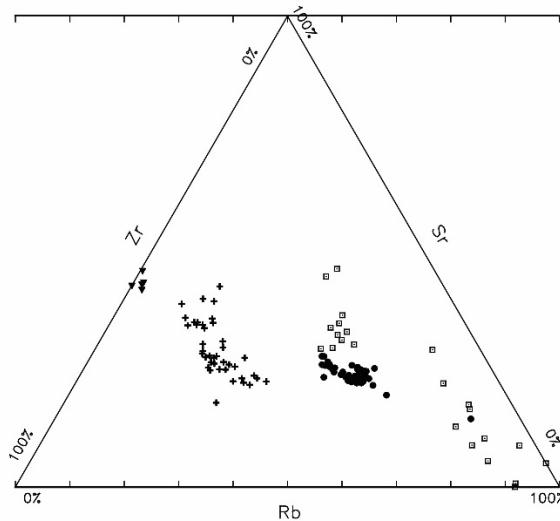
The example shows a simple trivariate plot showing three groups of data. The routine works in much the same way as the 2D scatterplot routine, except that you must select three variables. As with 2-D plots, 3-D plots include a legend: each group is given a different symbol and color. If you have a previously-created Principal Component transformation matrix in the working directory, you may choose to use it, in which case your variables will be principal components. If there is no PC transformation matrix in the working directory, then you will be using elements as variables. For more information, please refer to the help entry for [2D scatterplots](#).

13. Ternary diagrams

This routine allows you to produce a ternary scatterplot from Gauss datasets. This is a way to display the relations between datapoints and groups with respect to three variables; it may be easier to portray these relations in a ternary plot than in a 3-D plot. Each dataset will be given a separate color and symbol in the resulting plot, so collect your samples into datasets accordingly. A dataset can contain only a single sample if you wish it to have a unique color and symbol in the plot. You may choose to have the routine replace missing variables in all or none of the datasets; if you wish to replace missing variables in some, but not all, of the datasets, do so before using this routine.

You will first be asked to choose one or more datasets to display in the plot. Next you will select three elements, one for each of the three axes of the ternary diagram. You will also have the option of replacing missing variables in all or none of the datasets.

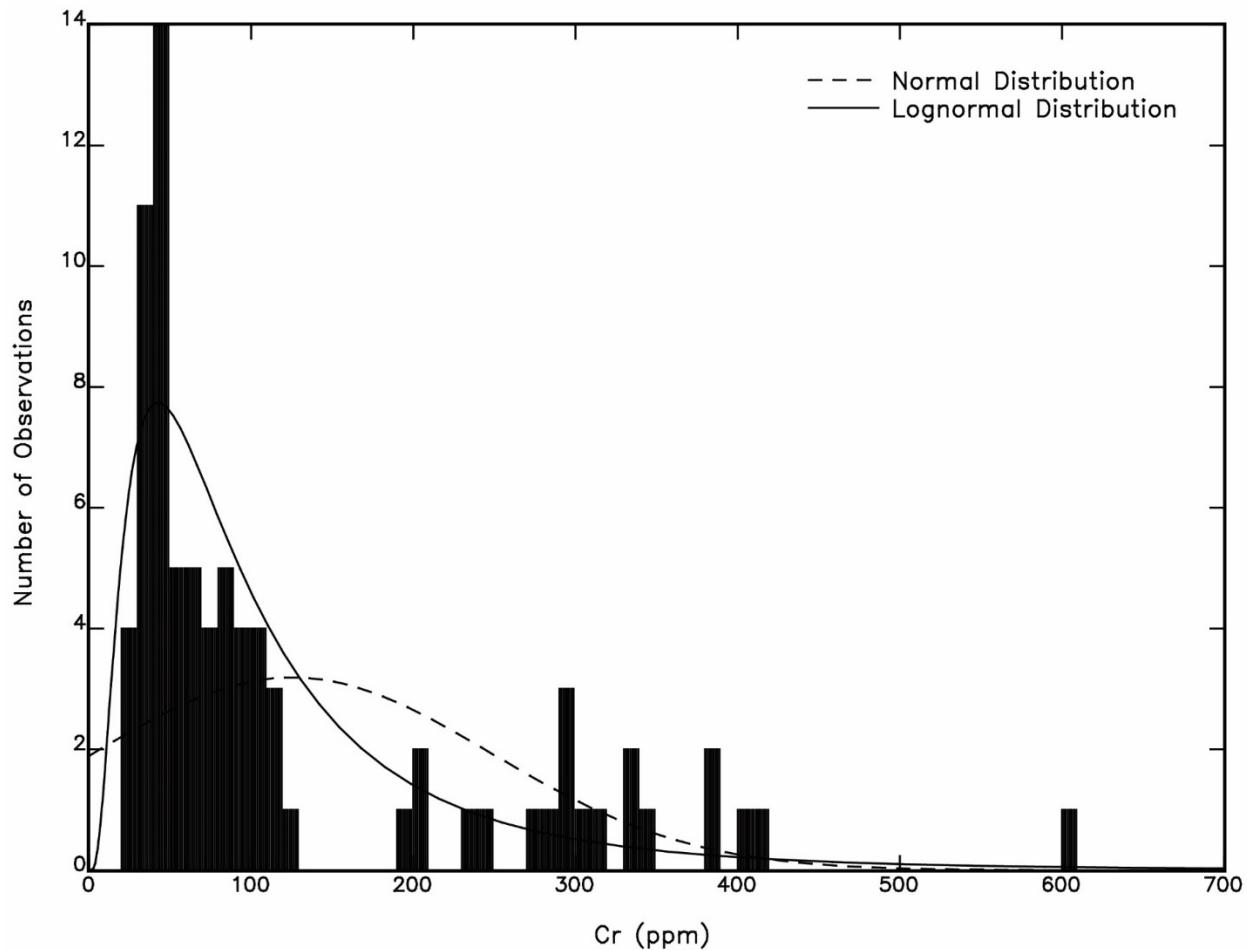
The ternary diagram produced by MURRAP will be displayed in a separate "[TKF File Viewer](#)" window. In general, Ternary diagrams are hard to interpret and have limited application. See the example for several related obsidian sources in Argentina below:



14. Distribution plots

This routine allows you to create a distribution plot displaying a histogram for a selected element from data in one or more datasets.

You may choose more than one element — a separate plot will be produced for each element. You have the option to display best-fit curves (default is yes), to specify the bin size (default is auto), and to select the color of the best-fit curves. The plot will be displayed in a separate "[TKF File Viewer](#)" window and may be saved in other formats. The routine itself does not offer options for making changes to the TKF file. Here is an example of what this routine will produce:



15. Compositional profile plots

This routine allows you to create a compositional profile plot for individual samples or groups of samples, displaying concentrations of selected elements in the form of a profile. Samples which are to be treated as a single group in the plot should be in a single dataset: the routine will plot the mean concentration values of the group's elements, with standard deviations indicated if desired. Several groups can be plotted, each group in its own dataset. The plot for each will be given a separate symbol and color. Put each individual sample in its own dataset if you want each to have a unique symbol and color on the plot.

You will first select one or more datasets containing samples which are to be treated as individual samples in the plot (you may enter NONE), and will then select one or more datasets containing groups you wish to plot. You have the option of replacing missing values in individual samples (default is no).

After selecting the elements to include, you have the option to plot these elements in order of decreasing concentration (default is no).

The plot will be displayed in a separate "[TKF File Viewer](#)" window and may be saved in other formats. The routine will present the following options for making changes to the TKF file:

- 0 Exit
- 1 Plot again

Options for modifying the profile plot are as follows:

- 2 Data distribution
- 3 Axis limits
- 4 Plot symbols
- 5 Plot colors
- 6 Axis label
- 7 Figure size
- 8 Axis scale (log/lin)
- 9 Legend position
- 10 Plot elements
- 11 Group components

Each option will have its defaults, so you can choose each to see what they offer and press the Enter key to get back to the option list without making changes. You must enter "1" (Plot again) in order for any changes you've made to be effected. When you plot again, a new TKF file with changes is produced in a new TKF window. The previous TKF file remains open and unchanged.

17. Best Relative Fit (Mommsen)

This routine applies a corrective 'best relative fit factor', also called a 'dilution factor', to each sample of a group of samples in a dataset, using an adaptation of Mommsen's method (Mommsen et al. 1988). After selecting datasets, you may select the elements to use in the calculations.

A table is produced for each dataset which contains four columns: 1) sample IDs (ANIDs), 2) dilution factors for each sample, 3) the spread for each sample, and 4) an "Evaluate" column indicating which elements should be evaluated based on a large value for their dilution factor or spread. To save this table in a text file, use the "Save As" option under the "File" menu. If you have used only one dataset, you have also the option of extracting samples from a group based on the dilution factors, to produce a new subgroup (default is No).

18. Total Variation Matrix (Aitchison/Buxeda)

This routine calculates a "total variation matrix" which can be used to assess the degree of variation in single dataset and the relative contribution of each element to that variation. The elements contributing most and least to the dataset's variation are identified, and ratios of the other elements to the latter element can be calculated. The method used is based on Aitchison's log-ratio technique and recent publications by Buxeda i Garrigos 1999 and Buxeda i Garrigos and Kilikoglou 2003.

After selecting a single dataset for analysis, and elements to include, a variation matrix is produced in the GSRUN window, as well as a table giving the relative variations of elements in the dataset, the total variation, and the elements contributing most and least to the dataset's variation.

You may save this data to a text file using the "Save As" option in the GSRUN "File" menu. The structure of the tables also makes it easy to copy and paste them directly from the GSRUN window into an Excel spreadsheet cell and use Excel's "Text to Columns" command under its "Data" menu to recover the table's structure in the spreadsheet.

References

Aitchison, J.A. 1990. Relative variation diagrams for describing patterns of compositional variability. *Mathematical Geology* 22(4): 487-511.

Aitchison, J.A. 1999. Logratios and the natural laws in compositional data analysis. *Mathematical Geology* 31(5): 563-589.

Buxeda i Garrigós, J. 1999. Alteration and contamination of archaeological ceramics: the perturbation problem. *Journal of Archaeological Science* 26(3): 295-303.

Buxeda i Garrigós, J., and V. Kilikoglou 2003. Total variation as a measure of variability in chemical data sets, in *Patterns and process: a Festschrift in honor of Dr. Edward V. Sayre* (ed. L. van Zelst). Washington, DC: Smithsonian Center for Materials Research and Education, pp. 185-198.

Cogswell, J.W, H. Neff, and M.D. Glascock 1998. Analysis of shell-tempered pottery replicates: implications for provenance studies. *American Antiquity* 63(1): 63-72.

Mommsen, H., Kreuser, A. & Weber, J. 1988. A method for grouping pottery by chemical composition. *Archaeometry* 30(1): 47-57.

Neff, H. 1994. RQ-mode principal components analysis of ceramic compositional data. *Archaeometry* 36(1): 115-130.

Sayre, E.V. 1975. Brookhaven procedures for Statistical Analyses of Multivariate Archaeometric Data. Brookhaven National Laboratory report, BNL-21693.