

PetroPlot: A plotting and data management tool set for Microsoft Excel

Yongjun Su and Charles H. Langmuir

*Department of Earth and Environmental Sciences, Columbia University, Palisades, New York, 10964, USA
(ysu@ldeo.columbia.edu)*

Paul D. Asimow

Division of Geological and Planetary Sciences, California Institute of Technology, M/C 170-25, Pasadena, California, 91125, USA

[1] PetroPlot is a 4000-line software code written in Visual Basic for the spreadsheet program Excel that automates plotting and data management tasks for large amount of data. The major plotting functions include: automation of large numbers of multiseried XY plots; normalized diagrams (e.g., spider diagrams); replotting of any complex formatted diagram with multiple series for any other axis parameters; addition of customized labels for individual data points; and labeling flexible log scale axes. Other functions include: assignment of groups for samples based on multiple customized criteria; removal of nonnumeric values; calculation of averages/standard deviations; calculation of correlation matrices; deletion of nonconsecutive rows; and compilation of multiple rows of data for a single sample to single rows appropriate for plotting. A cubic spline function permits curve fitting to complex time series, and comparison of data to the fits. For users of Excel, PetroPlot increases efficiency of data manipulation and visualization by orders of magnitude and allows exploration of large data sets that would not be possible making plots individually. The source codes are open to all users.

Components: 2993 words, 12 figures, 5 auxiliary material.

Keywords: Excel; plot; data-management; tool.

Index Terms: 1094 Geochemistry: Instruments and techniques; 3694 Mineralogy and Petrology: Instruments and techniques; 9815 General or Miscellaneous: Notices and announcements; 9820 General or Miscellaneous: Techniques applicable in three or more fields.

Received 4 February 2002; **Revised** 5 June 2002; **Accepted** 27 June 2002; **Published** 26 March 2003.

Su, Y., C. H. Langmuir, and P. D. Asimow, PetroPlot: A plotting and data management tool set for Microsoft Excel, *Geochem. Geophys. Geosyst.*, 4(3), 1030, doi:10.1029/2002GC000323, 2003.

1. Introduction

[2] Interpretation of geochemical data requires visualization of increasingly large data sets in multiple dimensions. Data are commonly archived in spreadsheet format, and therefore efficient visualization tools within this environment are necessary.

[3] Microsoft Excel is a spreadsheet program that is widely used for data storage and manipulation. It allows data to be organized, plotted and easily accessed in workbook format. However, the Excel program is not tuned for many aspects of scientific research. For example, many earth scientists need to view multiple sets of data on the same diagram. Excel permits this by dividing data into “series.”

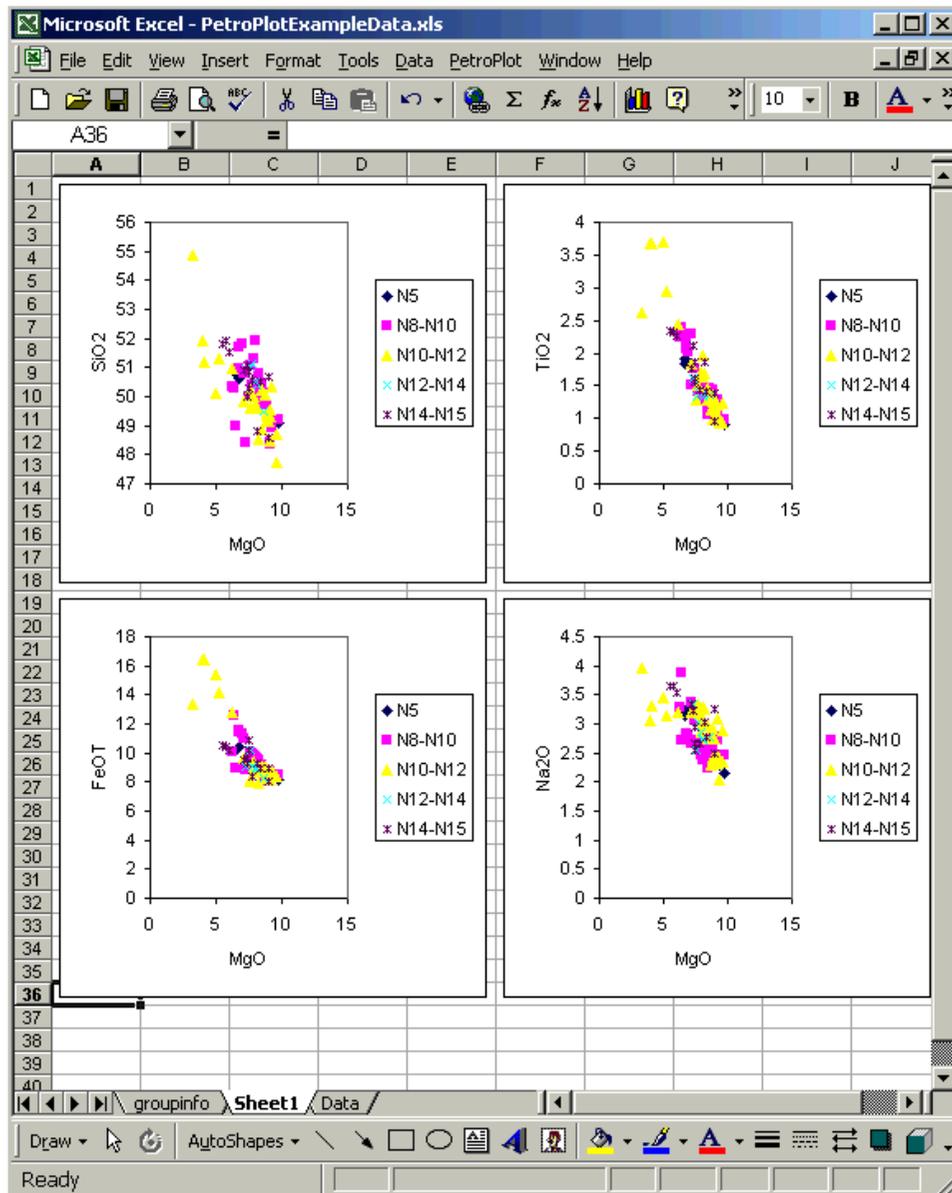


Figure 1. Example of plots made by the “MultiXY plots” function. While four diagrams are shown, hundreds of diagrams can be plotted if desired.

However, in order to make a multiseries XY plot, it is necessary to manually select the data range for each series in the “Chart Wizard” tool; or create a single-series chart using the wizard tool, and then repeatedly copy data for new series and paste to the chart. One could also copy the first “series” from the Formula bar, click on the chart, and paste the same “series” onto the formula bar again, and just change the row or column references. All these steps require extensive manual

labor for an XY plot with many series. Investigators may wish to examine several tens or hundreds of such multiseries XY plots, requiring hours of repetitive keyboard work.

[4] The labor involved with such data exploration has become increasingly onerous with the development of sampling and analytical technology as the amount of data grows exponentially. On-line databases have been developed to serve data over the

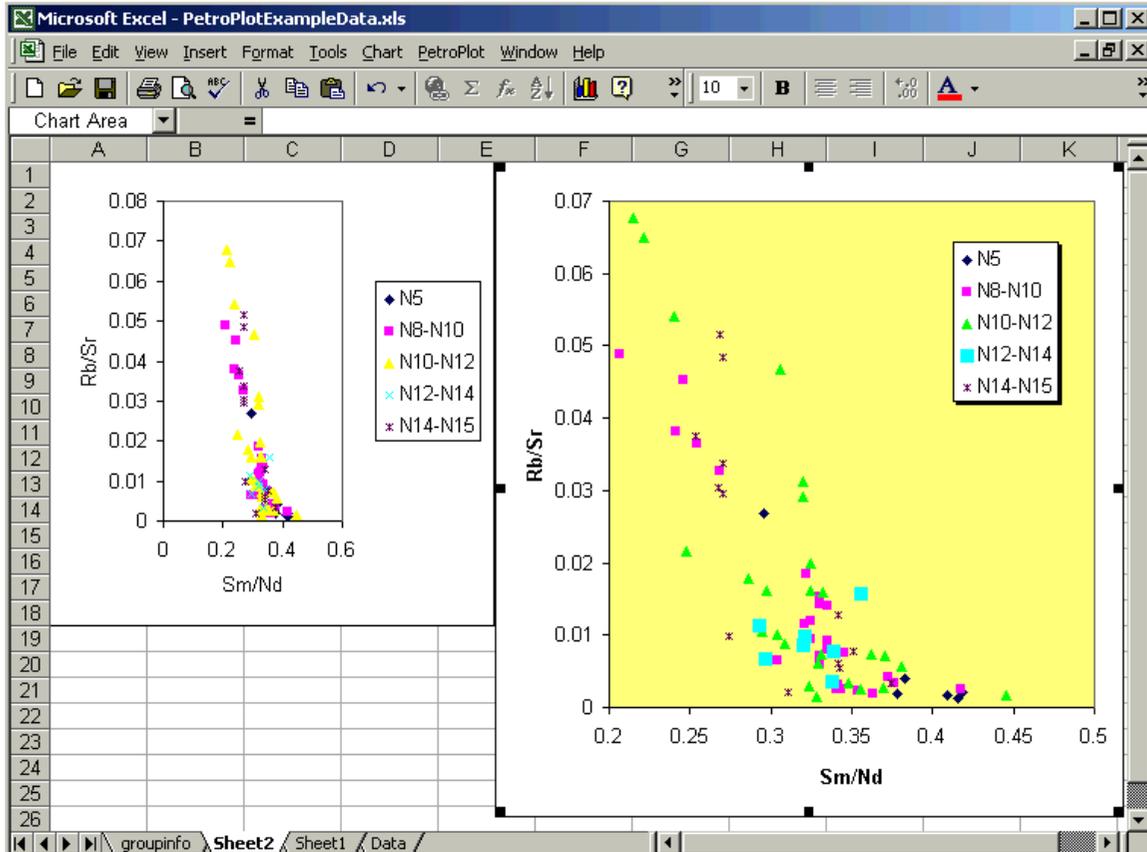


Figure 2. Comparison of Excel-generated chart format with customized chart format. In most cases, the Excel-generated chart format is not suitable for publication or lecture presentation. The user can create a well-formatted chart, use it as blueprint, copy it elsewhere and then use PetroPlot “Change XY” function to produce other plots with the same formatting. See Figure 3 for further information.

Web, making large amounts of data readily available to individual investigators (e.g., PETDB and GEOROC [Y. Su, K. Lehnert, and C. Langmuir, PETDB on World Wide Web: A comprehensive method for accessing petrologic data over the Internet, manuscript in preparation, 2002; Lehnert *et al.*, 1999; Nohl and Sarbas, 1999]). A single file may contain thousands of rows and hundreds of columns, and may not be in a form suitable for plotting. For example, in a file that is organized by analytical method, two measurements for a single sample or time may be in two rows, because these two elements are analyzed by different methods. In order to compare the values on a scatterplot, the two values for the same sample must be aligned in a single row. To manually prepare the data can take hours.

[5] Previous plotting tools [e.g., Christie and Langmuir, 1994] attempted to address these prob-

lems. While these earlier macros can still be made to function in the newer generation of Excel, they were not developed under the Visual Basic environment, and did not take advantage of the increased speed and flexibility of Visual Basic code.

[6] This paper presents new tools for plotting and data management: The PetroPlot program is designed using Visual Basic as an Excel add-in program; it can be applied on any system with Excel 97 or higher (both PC and Macintosh). A graphical user interface (GUI) is utilized to guide users through application steps; the usage of PetroPlot is simple and straightforward. Detailed instructions have been included as a tutorial package in an appendix that can also be downloaded together with the program. The source codes of PetroPlot are protected but may be freely copied or

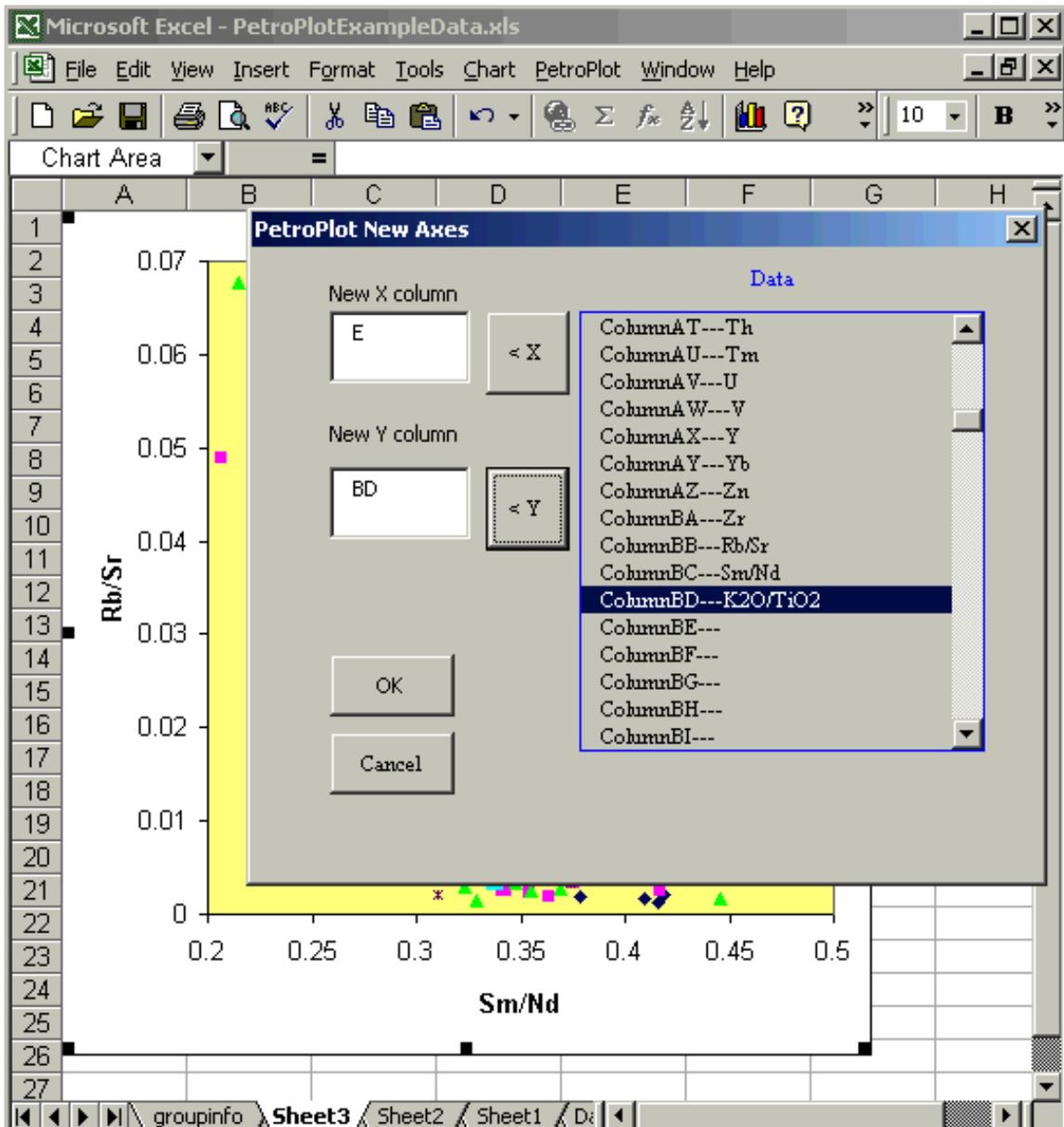


Figure 3. Example of PetroPlot “ChangeXY” function. (a) “ChangeXY” dialog and (b) “ChangeXY” result.

modified in order that individual users can customize their own needs and add new functions. The Programming environment is the Visual Basic Editor that comes with Excel. No stand-alone VB environment is needed. Because the code is open source, we plan ourselves and invite others to publish future updates as technical briefs, making use of the forward and backward linking environment of *Geochemistry, Geophysics, Geosystems (G-Cubed)*.

2. PetroPlot Functions

2.1. XY Plotting Tools

2.1.1. Multi XY Plots

[7] Normally each group, or “Series” in Excel terminology, requires a separate line of code in manual charting, which in practice limits the number of groups that can be investigated on a single chart. The PetroPlot multiplot function makes a

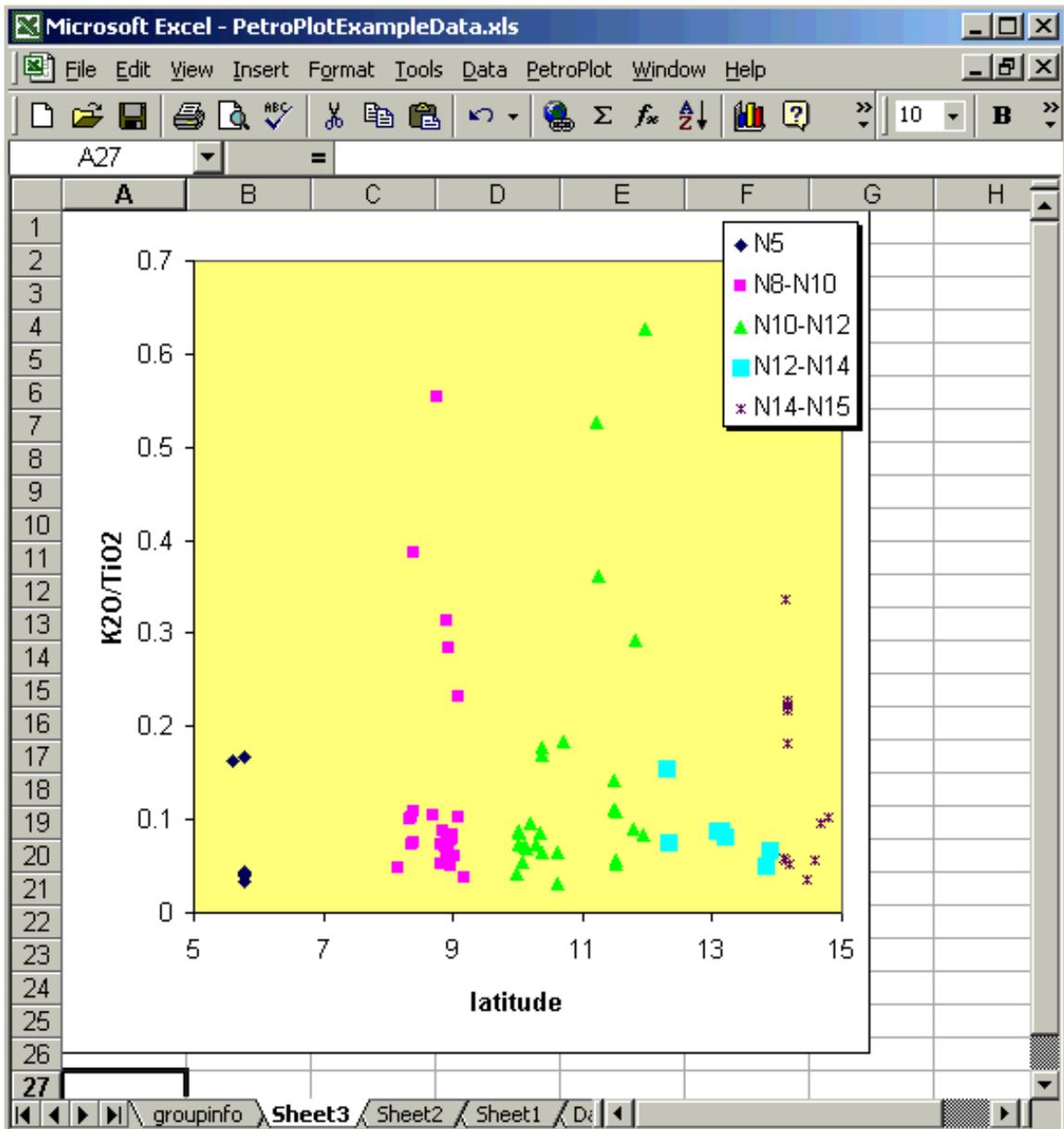


Figure 3. (continued)

large number of XY plots at one time. Each plot allows multiple series to be plotted using discrete symbols. For example, to investigate variations in the Th/Ba ratio of volcanoes from different arcs, many hundreds of samples from some 20-arc suites may be plotted with a few mouse clicks. Or, multiple time series from different locations could be plotted on the same diagram efficiently. Figure 1 shows examples of plots of multiple elements versus MgO. Hundreds of plots can be automati-

cally generated within 10 min and it is possible to view data and address questions that cannot be dealt with otherwise.

2.1.2. ChangeXY

[8] This function changes axes for all series in a single XY plot, while preserving the format of a chart. It is useful for generating multiple high-quality plots with complex formatting, especially when there are a large number of series.

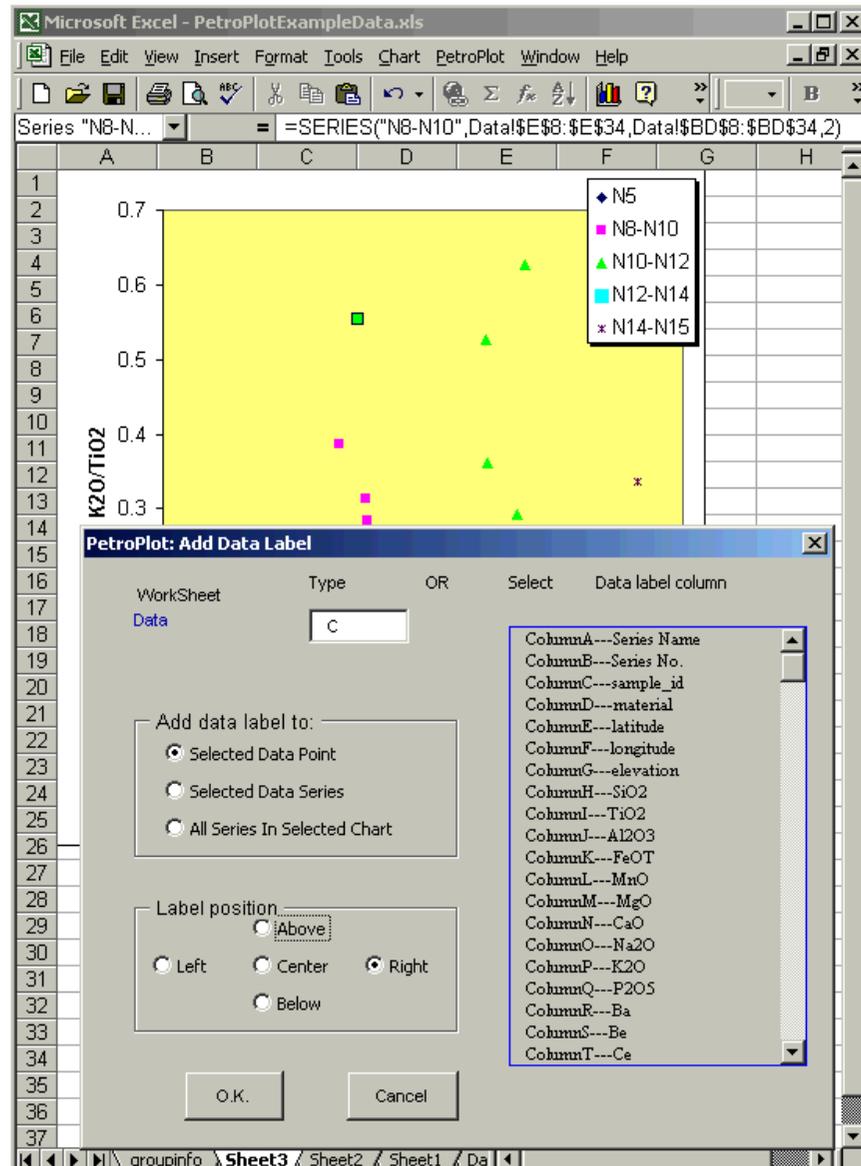


Figure 4. Example of PetroPlot Add custom data label function. (a) PetroPlot dialog. (b) Add label result. Excel does not provide the ability to label points with data other than (x,y) values. This PetroPlot function allows any data, such as sample name, location, or other variables, to be used as labels.

[9] For example, normal formatting of XY charts follows Excel’s default settings, which might be unsatisfactory. In Figure 1, series 1 uses blue diamonds, series 2 uses pink squares, and series 3 uses yellow triangles, etc. Users may want to change the colors or the shapes. A suggested solution is to make a single XY plot with multiple series using PetroPlot “Multi XY Plots” function; then manually modify its format (Figure 2). This plot can then be used as a template, and PetroPlot

“ChangeXY” function can be used to generate new plots (Figures 3a and 3b). Note, however that the scales of changed axes may need to be adjusted.

2.1.3. Add Data Label

[10] It can be useful to add labels to the individual points on a graph. Labels added using Excel’s menus are limited to either *x* or *y* values. Customized labels such as sample names, locations, and references that are included in the data table, are

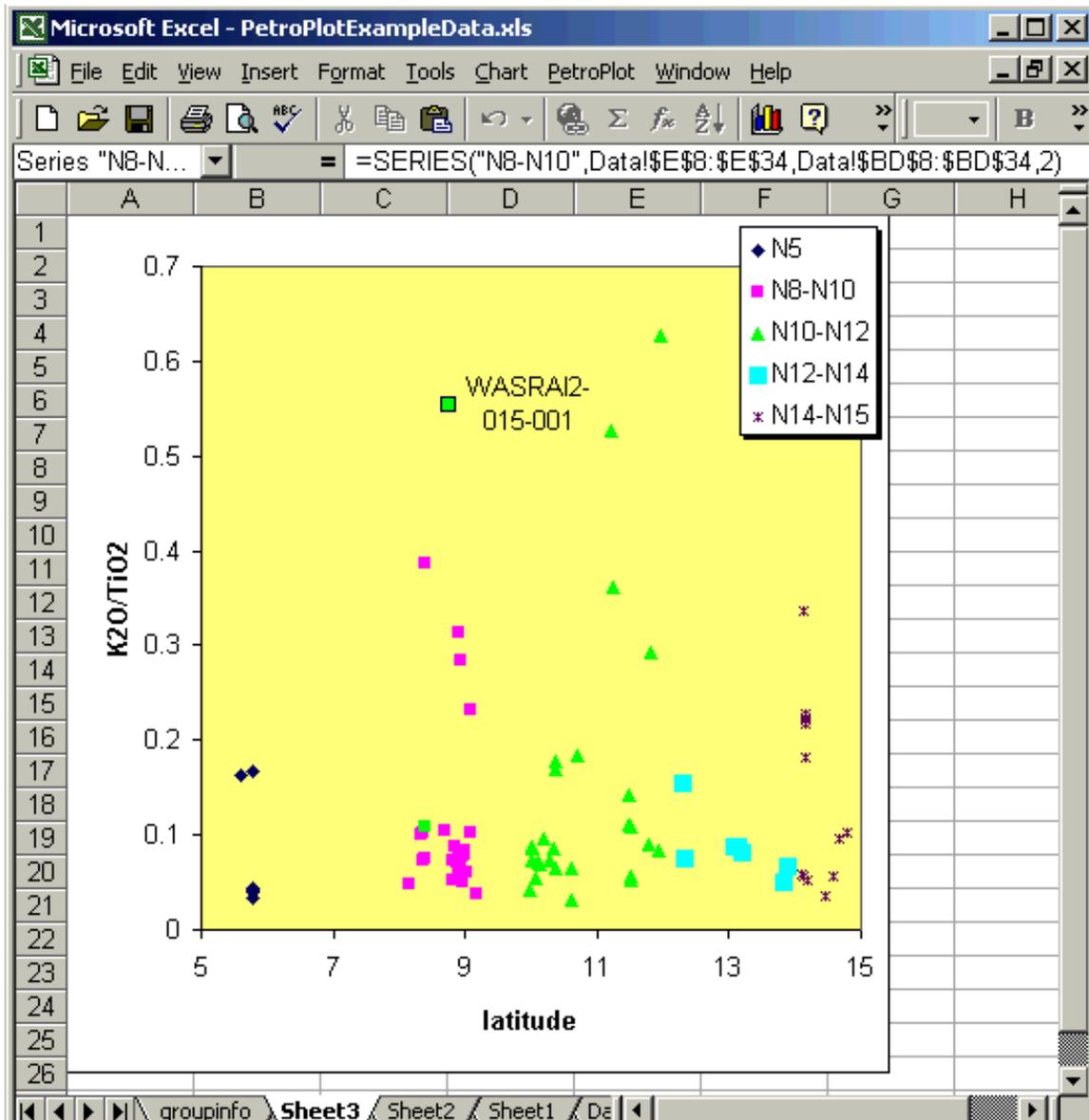


Figure 4. (continued)

much more valuable information on a plot. Adding such labels manually for plots with many points is too cumbersome to be practical. With PetroPlot, any kind of information listed in a column of worksheet can be attached as a label, and adding customized labels is as easy as to add Excel default labels (Figure 4).

2.2. Spider Diagram Tool

[11] Spider diagrams are plots of normalized data where data for many elements are normalized to a reference such as primitive mantle. Such plots are

commonly used to compare samples and to investigate trace element systematics in geochemistry. In order to make a spider diagram, the data have to be organized such that elements are in a specific order, and chemical values are divided by the standard values. PetroPlot "SpiderDiagram" function automates the entire process (Figure 5).

2.3. Log Scale Labels

[12] In Excel log scale plots, the default minimum and maximum values on axes are fixed to the integer powers of ten, i.e., 0.001, 0.01, etc. It is

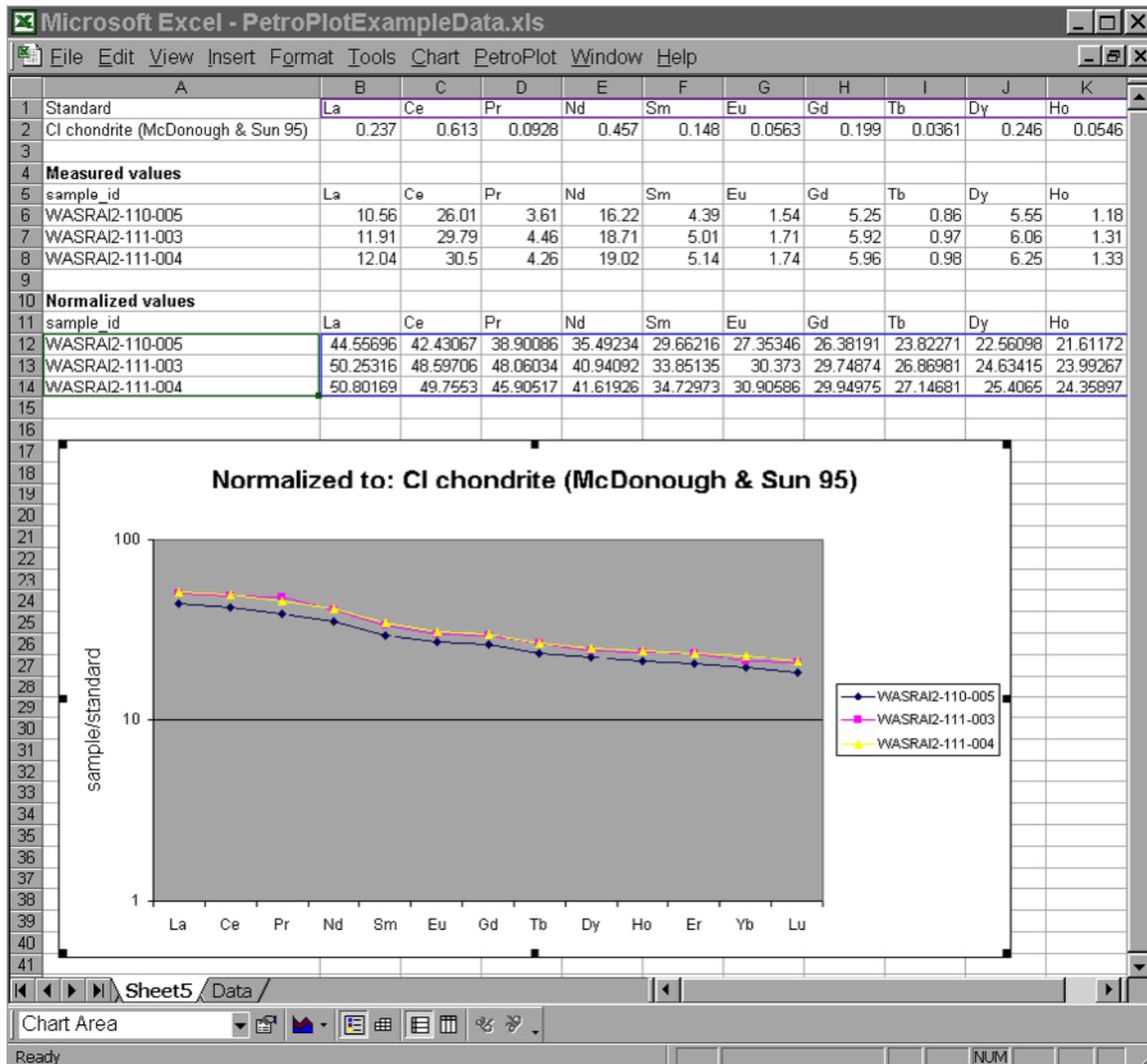


Figure 5. Example result of PetroPlot “SpiderDiagram” function. The plot is automatically updated once original data are changed.

often desirable to set the maximum and minimum on an axis to be any value, e.g., 0.2, 0.6, etc. The PetroPlot “Label Log Scale” function adds flexible log-scale labels to a plot where log values have been calculated on a separate column on the worksheet (Figure 6). It works not only for XY Scatter charts, but also line charts (Figure 7) or bar charts, etc.

2.4. Data Management Utilities

2.4.1. Clear Values

[13] Many scientific fields recalculate data as ratios (e.g., La/Sm, R/R_A) or through other conversions

(e.g., Na₈ [Klein and Langmuir, 1987]). An Excel calculation may return zeros or nonnumeric values (e.g., “#VALUE!” or “#DIV/0!”). Aggregate functions such as average or standard deviation cannot be applied when the selection contains nonnumeric cells. To manually delete these bad cells can be time consuming if the data volume is large. The PetroPlot function automates the cleaning process by making the meaningless cells empty.

2.4.2. Correl Matrix

[14] This function calculates a correlation matrix for a selected data area. The output is sent to a

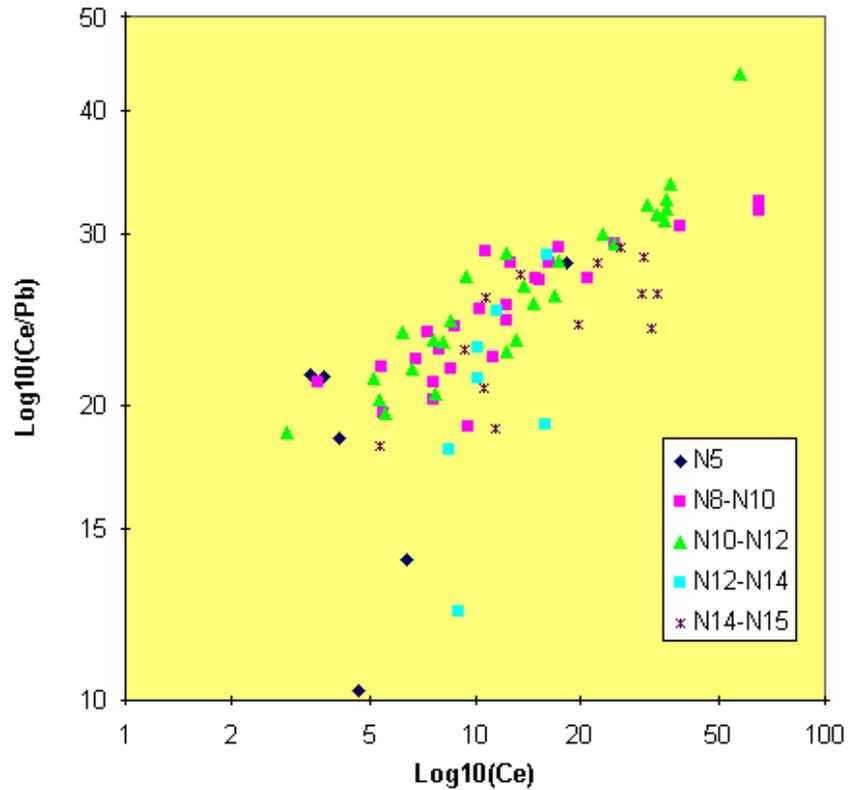


Figure 6. Label Log Scale example on an XY Scatter chart.

new worksheet (Figure 8). The result is similar to the correlation function in Excel Add-In “Data Analyses” tool package. The advantage of this PetroPlot function is that it outputs

formulas instead of calculated values. If the source data are modified, the correlation matrix generated by PetroPlot will be automatically updated.

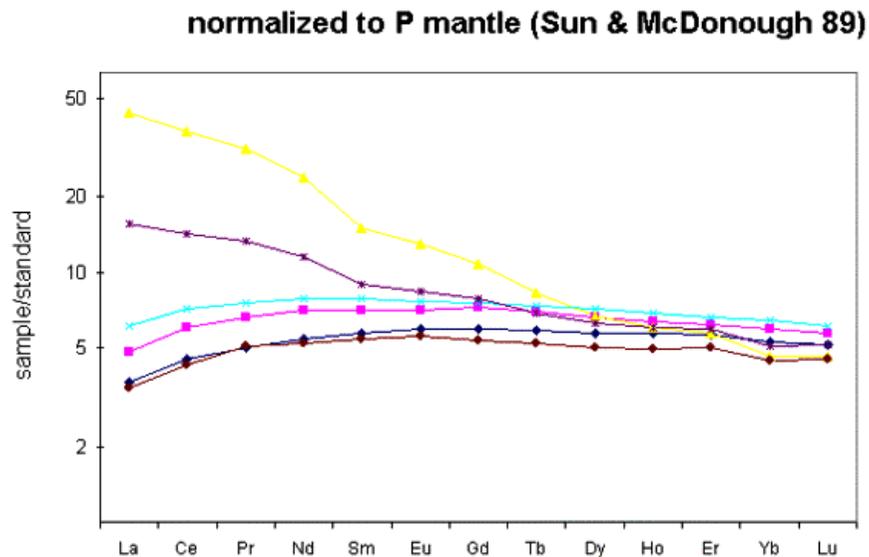


Figure 7. Label Log Scale example on a line chart.

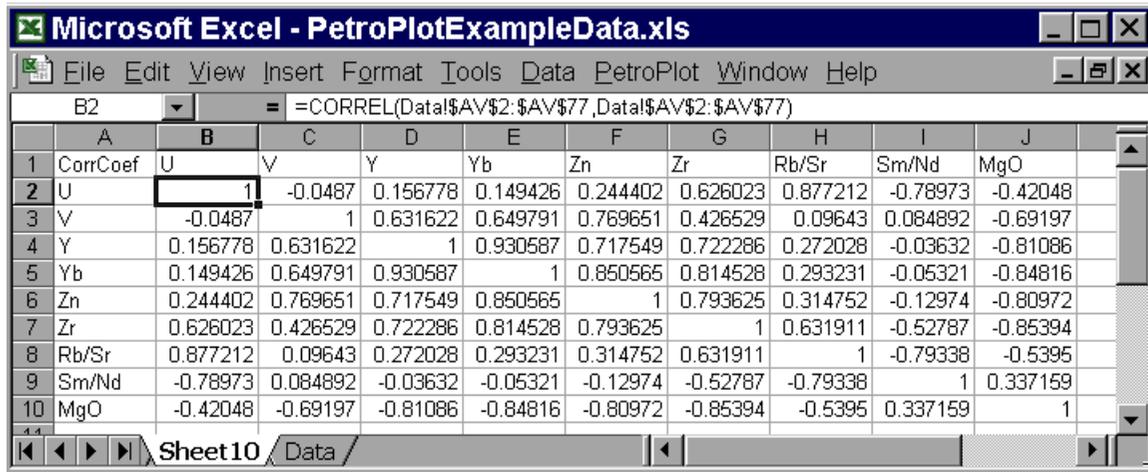


Figure 8. Example result of the PetroPlot “Correl Matrix” function.

2.4.3. Delete Sample

[15] This function can delete nonconsecutive rows based on customized criteria (Figure 9). Calculation formulas are exported to a new worksheet. The averages will be automatically updated once source data are modified.

2.4.4. Multiple Average

[16] This function calculates multiple averages or other aggregates functions such as “STDEV,” “MIN,” “MAX,” and “SUM” based on a grouping classification that is defined in a separate column. For example, all samples with the same group name could be automatically averaged (Figure 10). Calculation formulas are exported to a new worksheet. The averages will be automatically updated once source data are modified.

2.4.5. Grouping

[17] In petrological studies, samples are often grouped by a variety of criteria, such as locations, materials, chemical values, calculated indexes (e.g., Na_8 , La/Sm), analytical methods etc. The PetroPlot function reads grouping information from a grouping worksheet, and assigns grouping number/ID to individual samples in the data sheet based on user-specified criteria.

2.4.6. Management of Database Files

[18] Data files downloaded from PETDB (<http://petdb.ldeo.columbia.edu>) and GEOROC (<http://georoc.mpch-mainz.gwdg.de/>) are sometimes

listed as individual rows sorted by publication or analytical method. A single sample then has data that occupy more than one row. To manually combine all rows of one sample into a single row is very time-consuming, and can even take longer time than typing the data manually. The PetroPlot function automates this task by picking the first value of an element for one sample into a single row in a new sheet. If multiple values exist for that element, the result cell will be marked with an indicative color.

2.4.7. Cubic Spline Interpolation

[19] These functions (Spline and Splint) implement the cubic spline algorithm for interpolating to any x-value given a series of (x,y) pairs. The interpolating function is made up of a sequence of cubic polynomial curves that meet at the given data points with continuous first and second derivatives. The result is significantly better than linear interpolation for relatively smooth data [see *Press et al.*, 1992].

3. How to Use PetroPlot

[20] Detailed PetroPlot tutorials are provided as an appendix in the same package with the PetroPlot software. The tutorials present step-by-step examples of the installation and each of the functions discussed above. These tutorials can be updated using the *G-Cubed* forward and backward linking environment.

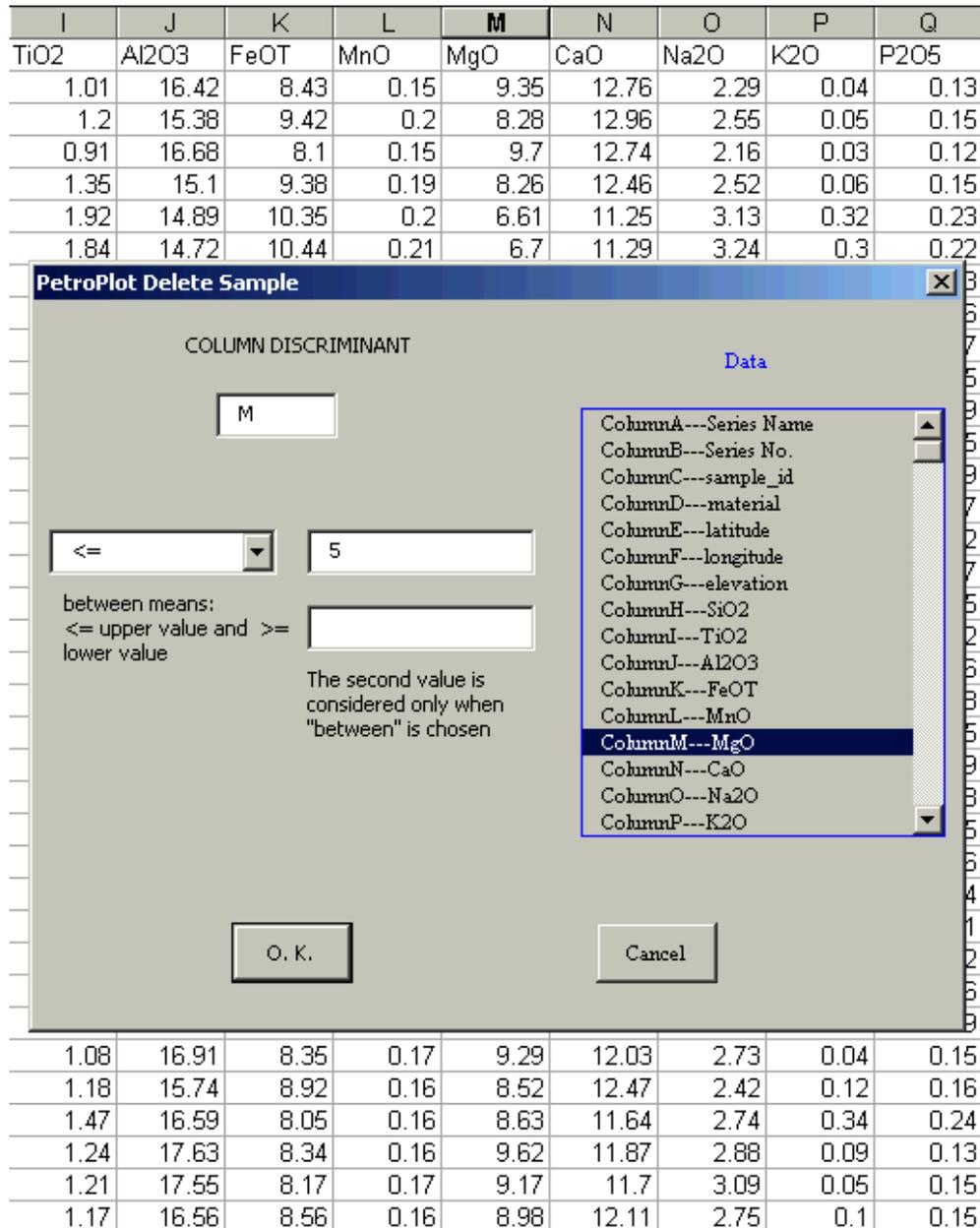


Figure 9. PetroPlot “Delete Sample” dialog. Samples can be deleted based on a discriminant function for a given variable.

[21] The “PetroPlot” menu buttons are straightforward, and the dialogs will guide users through information-entry steps. Additional rules below will help users to work more efficiently with PetroPlot functions.

1. A data sheet should be organized with rows corresponding to sample/analysis and columns representing categories (location, element, element

ratio, etc). The first row should contain only header information (i.e., element names etc).

2. All sheet-mode functions except “Compile Rows” work on selected areas in the data sheet. In general, whole sheet, whole column or whole row selections should be avoided. The program works most efficiently if only the cells with data to be worked on are selected.

	A	B	G	H	I	J	K	L
1	Series Name	Series No.	elevation	SiO2	TiO2	Al2O3	FeOT	MnO
3	N5	1	-1938	49.87	1.2	15.38	9.42	0.2
4	N5	1	-1788	49.12	0.91	16.68	8.1	0.15
5	N5	1	-1788	50.27	1.35	15.1	9.38	0.19
6	N5	1	-1788	50.72	1.92	14.89	10.35	0.2
7	N5	1	-2320	50.65	1.84	14.72	10.44	0.21
8	N8-N10	2	-2087	50.45	1.05	14.96	9.15	0.19
9	N8-N10	2	-3800	50.81	1.51	14.18	10.97	0.23
10	N8-N10	2	-3100	50.01	1.48	15.22	9.6	0.19
11	N8-N10	2	-2731	50.29	2.28	13.66	12.55	0.23
12	N8-N10							0.18
13	N8-N10							0.169
14	N8-N10							0.18
15	N8-N10							0.2
16	N8-N10							0.23
17	N8-N10							0.19
18	N8-N10							0.24
19	N8-N10							0.19
20	N8-N10							0.19
21	N8-N10							0.2
22	N8-N10							0.21
23	N8-N10							0.17
24	N8-N10							0.19
25	N8-N10							0.15
26	N8-N10							0.21
27	N8-N10							0.21
28	N8-N10							0.17
29	N8-N10							0.22
30	N8-N10							0.19
31	N8-N10	2	-2726	49.73	1.44	16.35	8.6	0.17
32	N8-N10	2	-1892	48.94	1.08	16.91	8.35	0.17
33	N8-N10	2	-2345	49.99	1.18	15.74	8.92	0.16
34	N8-N10	2	-3025	49.8	1.47	16.59	8.05	0.16
35	N10-N12	3	-2640	47.76	1.24	17.63	8.34	0.16
36	N10-N12	3	-2493	48.48	1.21	17.55	8.17	0.17
37	N10-N12	3	-2440	49.15	1.17	16.56	8.56	0.16
38	N10-N12	3	-2440	49.31	1.25	16.34	8.63	0.16
39	N10-N12	3	-2380	50.22	1.28	15.22	9.09	0.18
40	N10-N12	3	-2150	49.42	0.99	16.96	8.3	0.13
41	N10-N12	3	-2600	49.52	1.03	16.33	8.57	0.17
42	N10-N12	3	-3220	51.94	3.68	11.7	16.45	0.31

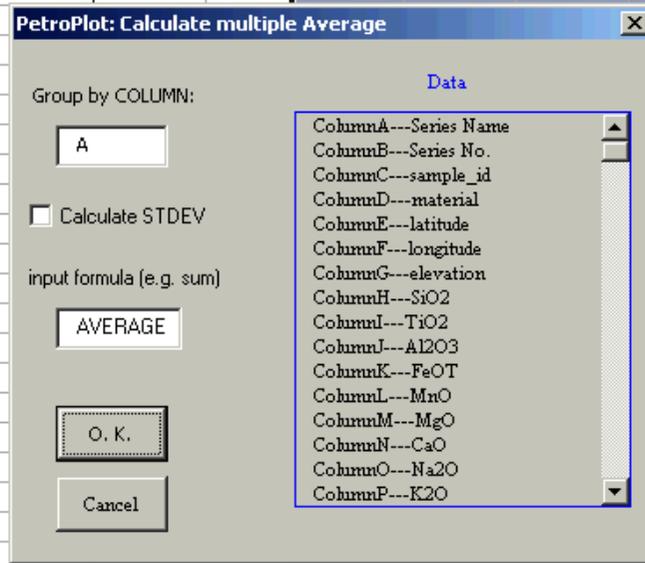


Figure 10. PetroPlot “Multiple Average” dialog. This function will sort the data according to the selected column, and then calculate an average for each unique identifier.

3. For “Multi XY Plots”, it is recommended that Series Names are stored in column A, and Series Numbers are stored in column B.

4. To use PetroPlot “ChangeXY” function, data sources for all series in one chart must come from a single worksheet. Series with data from other files will NOT be updated to the new axes.

5. “Spider Diagram” function requires that “PetroPlotSpider.xls” is open.

6. Sample names should include a text string, and not be only numbers.

7. For the “Grouping” function, at least one group info sheet is needed. There are no naming restrictions for group info sheets. A “-1” sign should be put in column A after the last grouping record. No “-1” is needed in the header row.

8. To use the “Compile Rows” function, users should presort the data sheet based on analytical

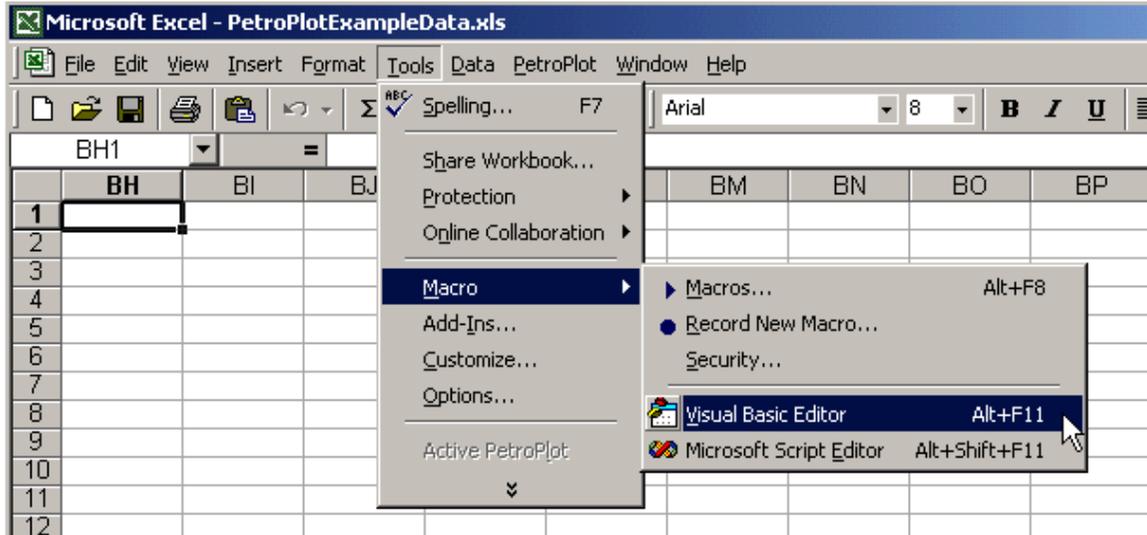


Figure 11. Open the "Visual Basic Editor" in Excel.

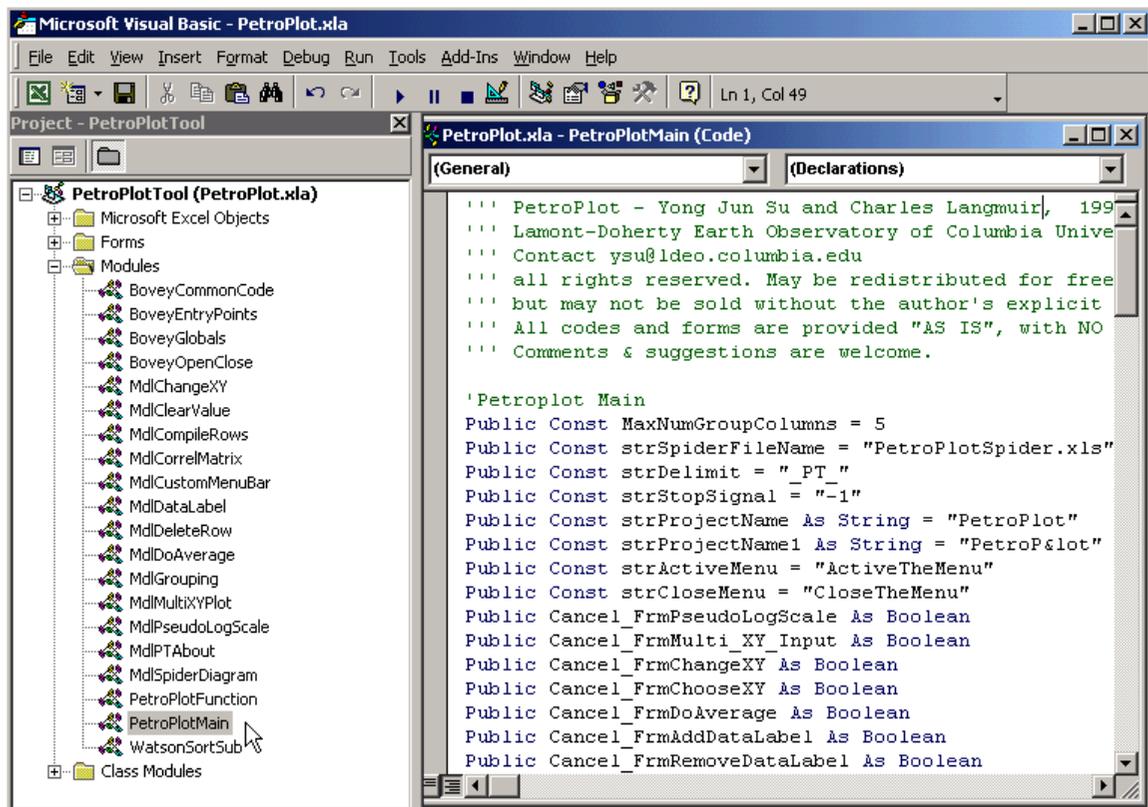


Figure 12. Visual Basic Editor with PetroPlot modules and forms.

quality. Make sure that better data appear first for a single sample.

4. Code Design

[22] PetroPlot is written in Visual Basic language. The source code of approximately 4000 lines is stored in the PetroPlot.xla file and is open to the public. The code can be viewed in the “Visual Basic Editor” in Excel (Figures 11 and 12).

[23] PetroPlot.xla contains modules and forms (Figure 12). The modules and forms are organized and named by functionality. The “PetroMain” defines public constants and variables. “PetroFunction” contains public functions in the PetroPlot project. “SortSub” is a quick sorting utility downloaded from the Internet. The remaining modules are named as “Mdlxxxx”. All forms are named as “Frmxxxx”.

[24] The code and a tutorial documenting installation and usage can be accessed by downloading the auxiliary material from the *Geochemistry, Geophysics, Geosystems* website. Interested readers are able to look at individual modules and forms once they have access to this program.

5. Summary

[25] PetroPlot contains both plotting and data management tools. These tools are designed to automate the manual work inside Excel. PetroPlot can be applied on most platforms. The source codes may be viewed and modified using Visual Basic Editor in Excel. PetroPlot has been successfully tested by the LDEO petrology group for two years.

Combined with on-line databases which enable access to large volume of data, PetroPlot can help earth scientists to effectively manage today’s larger volume of sample-based data.

Acknowledgments

[26] Colleagues in the LDEO petrology group have been loyal beta-testers for PetroPlot. Numerous feedback and suggestions from them have substantially improved the program. Comments from Dennis Geist, Bill White, Yaoling Niu, Peter Michael, and David Christie have greatly improved both functions and documentation of this software, and lead to the creation of the tutorial.

References

- Christie, D. M., and C. H. Langmuir, Automated XY plots from Microsoft Excel, *Comp. Geosci.*, 20, 47–52, 1994.
- Klein, E. M., and C. H. Langmuir, Global correlations of ocean ridge basalt chemistry with axial depth and crustal thickness, *J. Geophys. Res.*, 92, 8081–8089, 1987.
- Lehnert, K., Y. Su, C. H. Langmuir, B. Sarbas, and U. Nohl, A global geochemical database structure for rocks, *Geochem. Geophys. Geosyst.*, 1, Paper number 1999GC000026, 2000.
- McDonough, W. F., and S.-S. Sun, The composition of the Earth, *Chem. Geol.*, 120, 228 pp., 1995.
- Niu, Y., and R. Batiza, Trace element evidence from seamounts for recycled oceanic crust in the eastern Pacific mantle, *Earth Planet. Sci. Lett.*, 148, 471–483, 1997.
- Nohl, U., and B. Sarbas, GEOROC, the MPI Geochemical Rock Database: Introduction of the Web Interface, *Eos. Trans., AGU*, 80(46), Fall Meet. Suppl, F1184, 1999.
- Press, W. H., et al., *Numerical Recipes*, 2nd ed., pp. 113–116, Cambridge Univ. Press, New York, 1992.
- Sun, S.-S., and W. F. McDonough, Chemical and isotopic systematics of oceanic basalts: implications for mantle composition and processes, in *Magmatism in Ocean Basins*, edited by A. D. Saunders and M. J. Norry, *Geol. Soc. Spec. Publ. London*, 42, 313–345, 1989.