JOURNAL OF PETROLOGY | VOLUME 47 | NUMBER 6 | PAGES 1255–1259 | 2006 | doi:10.1093/petrology/egl013

# TECHNICAL NOTE
# Interpretation of Whole-rock Geochemical Data in Igneous Geochemistry: Introducing Geochemical Data Toolkit (*GCDkit*)

**VOJTĚCH JANOUŠEK[1,2]*, C. M. FARROW[3] AND VOJTĚCH ERBAN[1]**

[1]CZECH GEOLOGICAL SURVEY, KLÁROV 3/131, 118 21 PRAGUE 1, CZECH REPUBLIC

[2]INSTITUTE OF PETROLOGY AND STRUCTURAL GEOLOGY, CHARLES UNIVERSITY, ALBERTOV 6, 128 43 PRAGUE 2, CZECH REPUBLIC

[3]UNIVERSITY OF GLASGOW COMPUTING SERVICE, GLASGOW G12 8QQ, UK

**RECEIVED NOVEMBER 3, 2004; ACCEPTED FEBRUARY 16, 2006**
**ADVANCE ACCESS PUBLICATION MARCH 22, 2006**

*Geochemical Data Toolkit (*GCDkit*) is a program for handling and recalculation of geochemical data from igneous and metamorphic rocks. It is built using the Windows version of R, which provides a flexible and comprehensive language and environment for data analysis and graphics.* GCDkit *was designed to eliminate routine and tedious operations involving large collections of whole-rock data and, at the same time, provide access to the wealth of statistical functions built into R. Data management tools include import and export of data files in a number of formats, data editing, searching, grouping and generation of subsets. Included are a variety of calculation and normative schemes, for instance CIPW and Mesonorm, as are the common geochemical graphs (e.g. binary and ternary graphs, Harker plots, spider plots, and several dozens of classification and geotectonic discrimination diagrams). The graphical output is publication ready but can be further retouched if required. The system can be further expanded by means of plug-in modules that provide specialist applications.* GCDkit *is available as Free Software under the terms of the Free Software Foundation's GNU General Public License and can be downloaded from http://www.gla.ac.uk/gcdkit. The product is actively maintained and updated to provide additional functionality; Unix/Linux and Mac OS versions are being developed.*

KEY WORDS: *igneous rocks; geochemistry; norms; graphs; software; MS Windows; R language*

## INTRODUCTION

### The challenge: geochemical recalculations and plotting

The need to quantify aspects of the world in which we live is inherent in contemporary scientific thinking. Such an approach leads inevitably to a flood of numerical data, which have to be analysed by an appropriate software tool.

The interpretation of whole-rock geochemical data is not different, as it requires complex and time-consuming calculations, which on PC compatible computers are mainly performed using dedicated, stand-alone programs (e.g. *NewPet*—Clarke *et al.*, 1994; *IgPet*—Carr, 1995; *MinPet*—Richard, 1995; *PetroGraph*—Petrelli *et al.*, 2005). An alternative preferred by many researchers are spreadsheets, sometimes aided by special macros (Sidder, 1994; Su *et al.*, 2003). Most of these otherwise useful programs are, however, not well suited to management and interpretation of larger datasets. In addition, the usability of many of these tools is hindered by the fact that it is often difficult to determine exactly which algorithm has been employed, because documentation is not detailed enough and it is not a common practice to make the source code available. Even if it is, any modifications or additions to the original program can

*Corresponding author. Telephone +420 251085308. Fax: +420 251818748. E-mail: janousek@cgu.cz

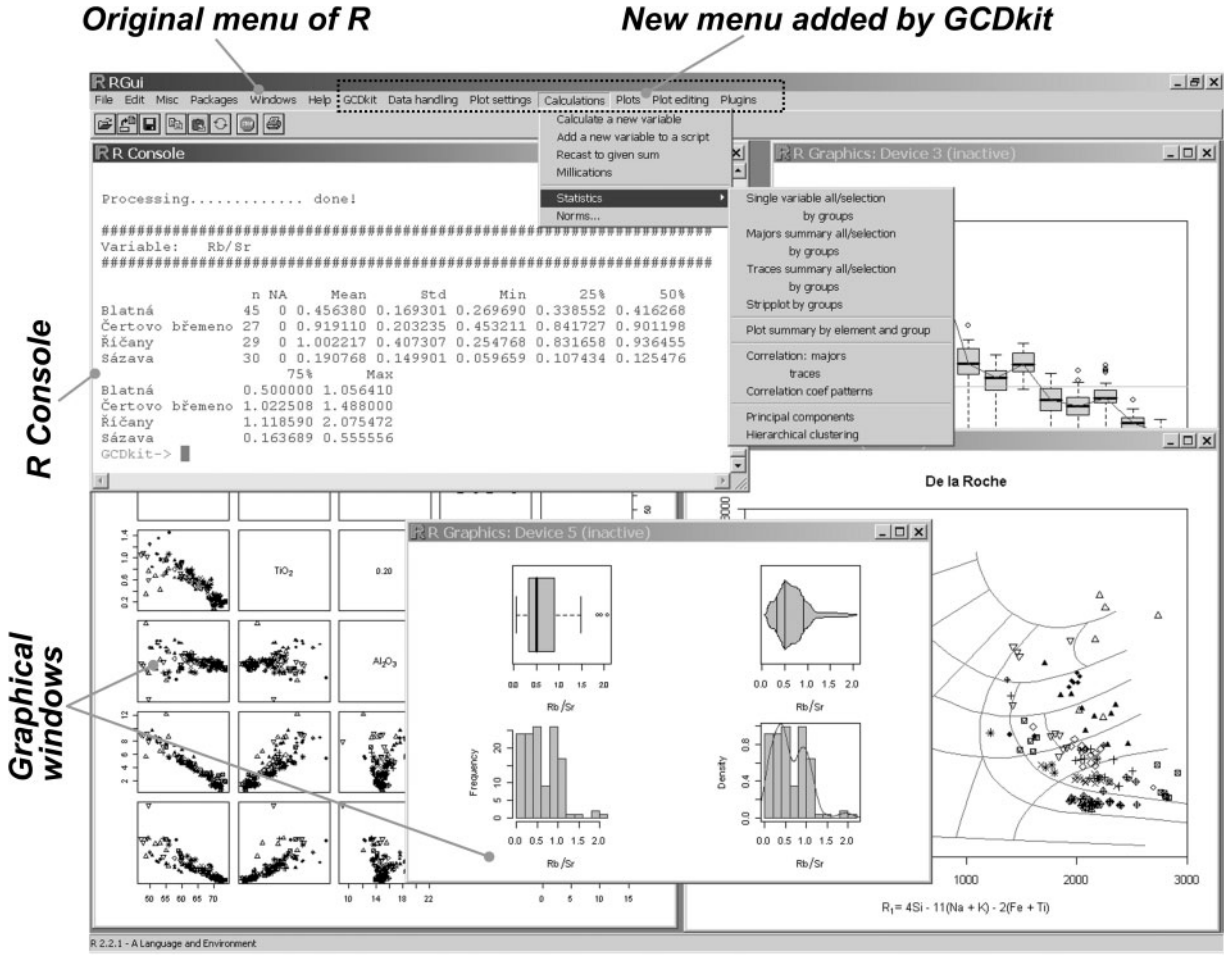


**Fig. 1.** Typical screen display of R with embedded *GCDkit*.

be tedious, or impossible for legal reasons. Concerning spreadsheets, they have low efficiency for repeated tasks, poor quality of graphic output, as well as limited protection of the primary data. For more complicated calculations the worksheet soon becomes too complex and prone to errors.

## The solution: Geochemical Data Toolkit (*GCDkit*)

To address the major drawbacks of the existing geochemical software, we have developed a new package called *Geochemical Data Toolkit for R* (*GCDkit*). *GCDkit* is a program for management and recalculation of whole-rock geochemical data for igneous and metamorphic rocks. It is built using the open-source (freeware) R language, version for Windows, which in itself provides a flexible and comprehensive environment for efficient data processing, visualization and statistical analysis (Ihaka & Gentleman, 1996; Grunsky, 2002; R Development Core Team, 2005).

*GCDkit* provides:

(1) a graphical user interface aiding the user with pull-down menus and dialogue boxes (Fig. 1);
(2) user-friendly front-end to the powerful statistical and graphical functions of R;
(3) core routines for effortless import, modification, searching, subsetting, classification and output of the geochemical data;
(4) calculation and plotting tools designed specifically for igneous geochemists;
(5) flexible and high-level graphical functions with an output into widely used graphical formats;
(6) no licensing problems, as it is distributed free via the WWW.

# PROGRAM DESCRIPTION

## Data management

The main area of application for *GCDkit* is the interpretation of larger datasets. Accordingly, great emphasis has

been put on data management tools (loading and saving data files, data editing, together with searching and generation of subsets).

*GCDkit* can read plain text files, including the variants represented by the *NewPet* (*.ROC) and *PetroGraph* (*.PEG) files as well as the outputs from WWW-based databases such as *GEOROC* (http://georoc.mpch-mainz.gwdg.de/georoc) and *PETDB* (http://www.petdb.org). Data can be imported from Microsoft Excel (*.XLS), Access (*.MDB), or dBase (*.DBF) formats, the latter being the native format for *IgPet* (Carr, 1995) or *MinPet* (Richard, 1995) packages. As an alternative, the data can be pasted, via the clipboard, from any Windows-based software, including popular spreadsheets. In any case the data are freeform, which means that the columns can be given in an arbitrary order and missing values are allowed.

Grouping of data is a key part of data management, controlling the output of calculations and the graphical routines. Analyses may be grouped on the basis of various criteria including sample attributes, so-called *labels* (locality, rock type, etc.), numerical variable, cluster analysis or selected classification diagram (e.g. total alkali–silica). For greatest flexibility queries may be logically combined, allowing, for example, selection based on rock type and contents of particular elements. Interactive selection from classification diagrams is also supported.

## Calculations

New variables can be calculated by mathematical expressions including constants, brackets, arithmetic operators and the whole spectrum of R functions. Statistical functions include descriptive statistics, box-and-whiskers and correlation plots or multivariate methods. Also implemented are a variety of calculation schemes used in igneous geochemistry, for instance CIPW norm, Catanorm, Niggli's cationic values, multicationic parameters of De La Roche *et al.* (1980) and Debon & Le Fort (1983), and improved Mesonorm for granitoid rocks (Mielke & Winkler, 1979). Given major-element analyses of a rock and its main mineral constituents, the best approximation of the modal composition, using either unconstrained or constrained least-squares methods (Albarède, 1995), may be calculated.

The results of all calculations can be appended to the current dataset for further processing (e.g. CIPW-normative albite, quartz and orthoclase to be plotted on a Ab–Qz–Or ternary diagram) or copied to a clipboard, saved or exported to HTML, DBF, XLS, MDB or text files.

## Plotting

The main strength of *GCDkit* is arguably the wealth of built-in publication-quality plots that can be exported into a number of data formats (including PostScript, WMF, JPG, PNG and BMP). The available graphs involve user-defined binary, ternary and multiple binary diagrams (such as Harker plots) as well as a wide palette of classification and geotectonic discrimination diagrams. Three variables can be displayed in pseudo three-dimensional plots, or on binary diagrams in which the size and colour of the plotted symbols correspond to the third variable (Fig. 2a).

All plots are publication ready but most can be customized with the included plot editing facilities. The system permits zooming and scaling of the diagrams, editing of the text, font, size and colour of the titles and axis labels, and of the colour, size and symbol for the data points, together with colour, type and width of the lines. Moreover, there is a function for interactive labelling of individual analyses (typically by sample names but other labels can be specified). The graph templates can be used also as a basis for classification. Put simply, the classification algorithm looks for the name of the polygon within the diagram, into which the rock analysis falls according to its $x$–$y$ coordinates.

*GCDkit* offers normalization of trace-element data by any chosen standard and the ability to generate normalized multielement diagrams, otherwise called 'spider plots'. One special feature are the 'spider boxplots', in which no individual patterns are drawn. Instead, the statistical distribution of each element is portrayed by a box-and-whiskers plot of normalized concentrations (Fig. 2b). The standard compositions used for normalization can be readily edited and new schemes added. For visualization of larger datasets there are contour and frequency plots (Fig. 2c).

## Expandability

The open architecture and simplicity mean that users will readily be able to further enhance the capabilities of the system, by modifying existing functions or creating new ones. The system incorporates a plug-in mechanism whereby new calculation or plotting code is automatically loaded and made accessible through the menu system.

Included with the current version are plug-in modules to calculate saturation temperatures of accessory minerals (Janoušek, 2006) as well as for interpretation of Sr–Nd isotopic data (Fig. 2d).

# CONCLUDING REMARKS

*GCDkit* will run on any IBM PC compatible computer with Windows 98/ME/2000/NT/XP (but latest versions are strongly recommended). The program with corresponding documentation is available free of charge at http://www.gla.ac.uk/gcdkit and is easy to install.

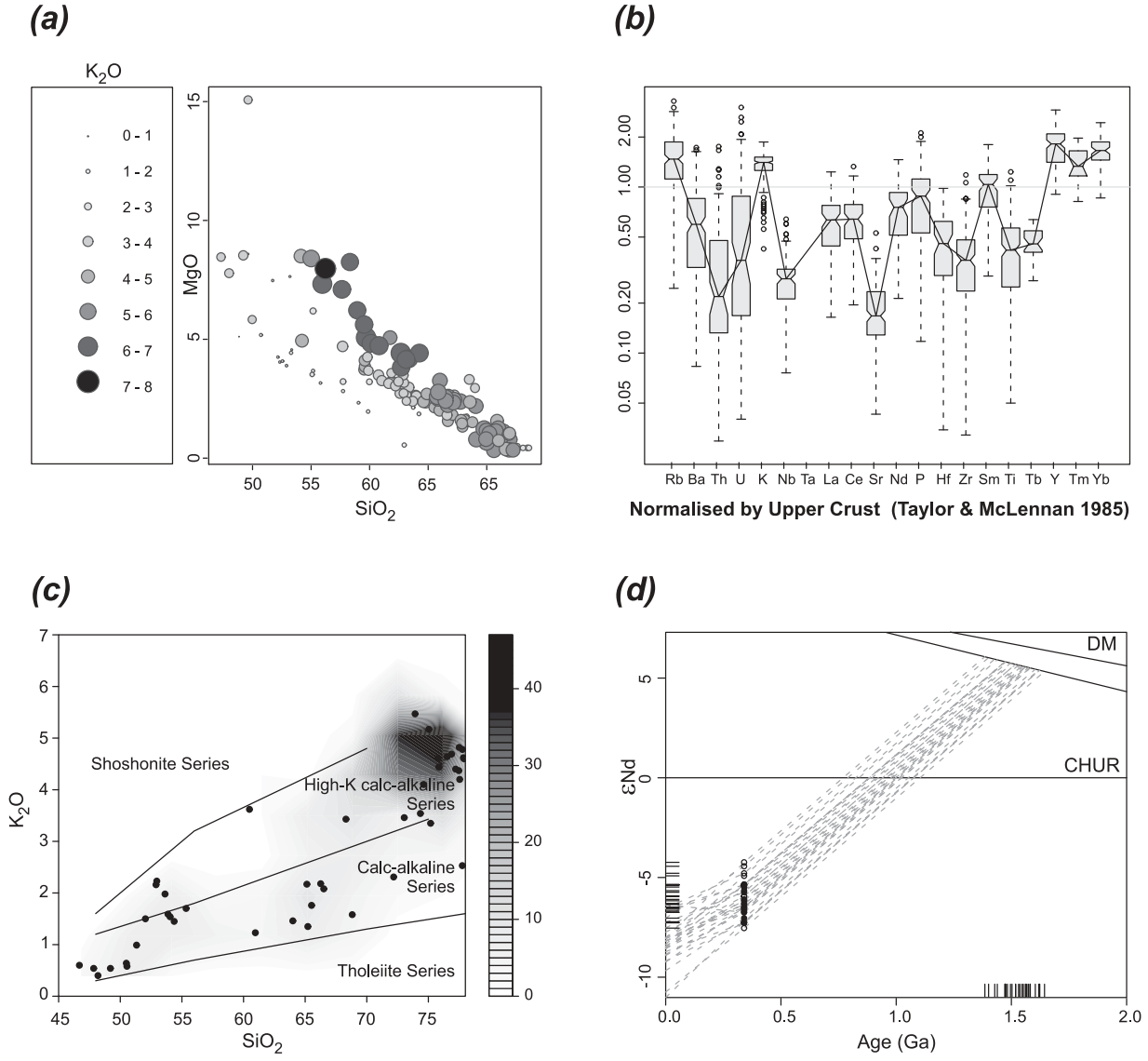


**Fig. 2.** Examples of diagrams generated by *GCDkit*. (a) $SiO_2$–MgO plot for granitoids of the Variscan Central Bohemian Pluton, Czech Republic, with size and grey shade of the plotting symbols corresponding to the $K_2O$ contents of individual samples. The data are from Janoušek *et al.* (2000, and unpublished). (b) Average upper crust normalized (Taylor & McLennan, 1985) 'spider boxplots' for felsic Variscan granulites from the Moldanubian Unit, Bohemian Massif (Janoušek *et al.*, 2004). (See text for explanation.) (c) $SiO_2$–$K_2O$ plot with the discrimination boundaries between the tholeiitic, calc-alkaline, high-K calc-alkaline and shoshonitic rocks of Peccerillo & Taylor (1976). Contoured grey fields are for all available analyses from Moldanubian granulites and data points are for the Lišov Massif, Southern Bohemia (Janoušek *et al.*, 2006). (d) Two-stage Nd isotope development diagram for felsic Moldanubian granulites. DM, Depleted Mantle evolution lines after Goldstein *et al.* (1984) and Liew & Hofmann (1988); CHUR, Chondritic Uniform Reservoir. The extra tick marks on axes show two-stage Nd model ages (*x*-axis) and initial $\varepsilon_{Nd}$ values (*y*-axis).

Our mission is to develop and release a platform-independent system, for Windows, Unix/Linux and Macintosh (System X). In the shorter-term perspective, new plug-ins for direct and inverse modelling of major petrogenetic processes in igneous geochemistry are under development.

The whole system, which is modular and straightforward to modify, provides a practical solution to the needs of petrologists and geochemists. If current modules do not meet requirements, or new techniques evolve, it is hoped that the open source availability of the code will encourage its onward development by users. In this context any feedback (bug reports, suggestions for further development, and pieces of contributed code or even ready plug-ins) would be highly appreciated.

## ACKNOWLEDGEMENTS

We are indebted to Jakub Šmíd for much of the ternary plotting function, the Core Team and the whole community for their sterling efforts in continuous developing of R—*GCDkit* is often merely an interface to their functions. Thanks are specifically due to Friedrich Leisch (Technische Universität Wien) and Duncan Murdoch (University of Western Ontario) for their invaluable advice. Last but not least, we thank beta testers (in particular Pavel Hanžl and Bernhard Humer) for their courage and patience, and the students in Prague, Brno and Salzburg for asking nasty questions. This paper and the program itself has benefited from extensive testing, detailed comments and words of encouragement of J.-F. Moyen. Also, the reviews by R. Giere, D. Geist and T. Barry have helped us to improve aspects of the software. Most of the coding of *GCDkit* originated during the research stay of V.J. at the Institute of Mineralogy, University of Salzburg, in the framework of the FWF Project 15133-GEO (to Fritz Finger). Support from the Czech Grant Agency (Postdoctoral Grant 205/97/P113) and the Czech Geological Survey (Internal Project 3314) is gratefully acknowledged.

## REFERENCES

Albarède, F. (1995). *Introduction to Geochemical Modeling*. Cambridge: Cambridge University Press, 543 pp.

Carr, M. (1995). *Program IgPet*. Somerset, NJ: Terra Softa.

Clarke, D., Mengel, F., Coish, R. A. & Kosinowski, M. H. F. (1994). *NewPet for DOS, version 94.01.07*. St. John's: Department of Earth Sciences, Memorial University of Newfoundland.

De La Roche, H., Leterrier, J., Grandclaude, P. & Marchal, M. (1980). A classification of volcanic and plutonic rocks using $R_1R_2$-diagram and major element analyses—its relationships with current nomenclature. *Chemical Geology* **29**, 183–210.

Debon, F. & Le Fort, P. (1983). A chemical–mineralogical classification of common plutonic rocks and associations. *Transactions of the Royal Society of Edinburgh, Earth Sciences* **73**, 135–149.

Goldstein, S. L., O'Nions, R. K. & Hamilton, P. J. (1984). A Sm–Nd isotopic study of atmospheric dusts and particulates from major river systems. *Earth and Planetary Science Letters* **70**, 221–236.

Grunsky, E. C. (2002). R: a data analysis and statistical programming environment—an emerging tool for the geosciences. *Computers & Geosciences* **28**, 1219–1222.

Ihaka, R. & Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**, 299–344.

Janoušek, V. (2006). *Saturnin*, R language script for application of accessory-mineral saturation models in igneous geochemistry. *Geologica Carpathica* **57**, (in press).

Janoušek, V., Bowes, D. R., Rogers, G., Farrow, C. M. & Jelínek, E. (2000). Modelling diverse processes in the petrogenesis of a composite batholith: the Central Bohemian Pluton, Central European Hercynides. *Journal of Petrology* **41**, 511–543.

Janoušek, V., Finger, F., Roberts, M. P., Frýda, J., Pin, C. & Dolejš, D. (2004). Deciphering petrogenesis of deeply buried granites: whole-rock geochemical constraints on the origin of largely undepleted felsic granulites from the Moldanubian Zone of the Bohemian Massif. *Transactions of the Royal Society of Edinburgh, Earth Sciences* **95**, 141–159.

Janoušek, V., Gerdes, A., Vrána, S., Finger, F., Erban, V., Friedl, G. & Braithwaite, C. J. R. (2006). Low-pressure granulites of the Lišov Massif, southern Bohemia: Viséan metamorphism of late Devonian plutonic arc rocks. *Journal of Petrology* **47**, 705–744.

Liew, T. C. & Hofmann, A. W. (1988). Precambrian crustal components, plutonic associations, plate environment of the Hercynian Fold Belt of Central Europe: indications from a Nd and Sr isotopic study. *Contributions to Mineralogy and Petrology* **98**, 129–138.

Mielke, P. & Winkler, H. G. F. (1979). Eine bessere Berechnung der Mesonorm für granitische Gesteine. *Neues Jahrbuch für Mineralogie, Monatshefte* 471–480.

Peccerillo, A. & Taylor, S. R. (1976). Geochemistry of Eocene calc-alkaline volcanic rocks from the Kastamonu area, Northern Turkey. *Contributions to Mineralogy and Petrology* **58**, 63–81.

Petrelli, M., Poli, G., Perugini, D. & Peccerillo, A. (2005). PetroGraph: a new software to visualize, model, and present geochemical data in igneous petrology. *Geochemistry, Geophysics, Geosystems* **6**, 1–15.

R Development Core Team (2005). R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. http://www.R-project.org.

Richard, L. R. (1995). *MinPet: Mineralogical and Petrological Data Processing System, Version 2.02*. Québec: MinPet Geological Software.

Sidder, G. B. (1994). Petro.calc.plot, Microsoft Excel macros to aid petrologic interpretation. *Computers & Geosciences* **20**, 1041–1061.

Su, Y. J., Langmuir, C. H. & Asimow, P. D. (2003). PetroPlot: a plotting and data management tool set for Microsoft Excel. *Geochemistry, Geophysics, Geosystems* **4**, 1–14.

Taylor, S. R. & McLennan, S. M. (1985). *The Continental Crust: its Composition and Evolution*. Blackwell: Oxford, 312 pp.