Geochemical Data Toolkit (GCDkit): a key for magmatic geochemists to the treasury of data analysis, statistics and graphics in R

Vojtěch ERBAN[1], Vojtěch JANOUŠEK[1,2], Colin FARROW[3], Jakub ŠMÍD[1]

[1]Czech Geological Survey, Klárov 3, 118 21 Praha, Czech Republic

[2]Institut für Mineralogie, Universität Salzburg, Hellbrunnerstraße 34, A-5020 Salzburg, Austria

[3]Computing Service Dept., University of Glasgow, G12 8QQ Glasgow, Scotland

The need to quantify aspects of the world in which we live is inherent in contemporary scientific thinking. Magmatic geochemistry is no exception, inasmuch as there is often no first-hand experience with the studied phenomena. Such an approach leads inevitably to a flood of numeric data, which have to be analysed by an appropriate and potent software tool.

In our view this has to have several features: (1) perform commonly used petrochemical re-calculations and straightforward data handling, (2) offer flexible and high-level graphical functions with an output into a widely used vector format, (3) allow additional calculations and statistical analysis when necessary, (4) avoid any licensing problems. As there is strong demand for such software, several programs have been released, such as MinCalc (Melín & Kunst 1992), Newpet (Clarke 1993), MinPet 2 (Richard 1995), IgPet 2000 (Carr & Dehn 2003), Norman (Janoušek 2000) or various MS Excel plug-ins. However, none of these products meets all the above mentioned requirements. Geochemical Data Toolkit for R (GCDkit) is our response to this challenge.

GCDkit is a tool for whole-rock magmatic geochemistry built using R, an environment for data analysis and graphics (Ihaka and Gentleman, 1996). R is available publicly under the terms of the Free Software Foundation's GNU General Public License, and can be downloaded from *http://www.r-project.org*. The user communicates mainly via the command line, but includes functions for facilitating Windows-like interaction (GUI = graphical user interface) and these are exploited by GCDkit.

GCDkit not only provides a user interface to the powerful functions built in R, but also adds on specialised tools for handling and recalculation of whole-rock major- and trace-element analyses and Sr–Nd isotopic data from igneous rocks. All functions are accessible via pull-down menus, as well as in the interactive regime.

When a tab-delimited data set is loaded into the system, all metadata (i.e. non-numeric strings) are stored in a data frame '*labels*' whereas numerical data are allocated into a matrix '*WR*'. Additionally, all missing data are replaced with the *NA* (= not available) values, and some additional parameters are calculated (anhydrous major-element data and millications), and, if necessary, columns with some commonly used geochemical indexes are appended to the matrix '*WR*' (mg#, FeO_t, A/CNK etc). Subsequently, operations available to the user can be split into several categories:

- **Data handling**, including generation of subsets, sorting into groups on the basis of various criteria, direct editing of the data or adding new variables (columns). The data searches and creation of subsets is performed using conditions employing regular expressions and Boolean logic. Important feature of data handling in the GCDkit is grouping. All analyses may be sorted into unique groups, which are subsequently utilised by some statistical and plotting functions (and as a factor by power users). Groups can be assigned on the basis of labels (locality, rock type …), a value of a chosen numerical variable, position in classification diagram (e.g., TAS) or cluster analysis. The grouping information is stored in a vector *'groups'* until the next assignment is performed.
- **Basic descriptive statistics**, including histograms, box-and-whiskers diagrams, correlation plots, principal components or cluster analysis. Each of the implemented functions is also equipped with a powerful data filter to specify the data that are to be processed.
- **Plotting functions**, including user defined binary or ternary plots, coplots, as well as Harker diagrams, spiderplots or variable visualisations of Sr–Nd isotopic data. A wide palette of classification and geotectonic discrimination diagrams is available, and new templates can be

added in a rather simple way. All plots are publication ready, however, if need be, the graphic output may be saved in Postscript or Windows Metafile (WMF) and further embellished in any vector-based drawing program.

- **Figaro** – a set of interactive tools for editing and annotation of all classification and some geotectonic diagrams (zooming, editing axis labels, adding comments or legend, identifying points and many others)
- **Recalculation algorithms**, such as CIPW, Catanorm, Granite Mesonorm, Niggli's numbers, as well as variable saturation calculations (apatite, zircon and monazite).

To sum up, the GCDkit is software written by geochemists for geochemists that grew continually in response to authors' everyday needs. Thus, it was designed to eliminate routine and tedious operations involving large collections of whole-rock data and, at the same time, to make accessible a wealth of statistical functions built into R. Nevertheless the main aim was to free the igneous geochemists for inventive thinking.

This project was supported by the grants FWF 15133–GEO and GAČR 205/97/P113 to VJ.

CARR M. and DEHN J. IGPET petrologic software. Accessed February 14, 2003, at URL *http://www.v-home.alaska.edu/~jdehn/vpetro/igpet.htm*.

CLARKE D., 1993. NewPet for DOS. Accessed November 15, 1999, at URL ftp://www.esd.mun.ca/pub/geoprogs/np940107.exe.

IHAKA R. and GENTLEMAN R., 1996. R: A language for data analysis and graphics*: Journal of Computational and Graphical Statistics*, 5: 299–314.

JANOUŠEK V., 2000. Norman, a QuickBasic programme for petrochemical re-calculation of whole-rock major-element analyses on IBM PC. *J. Czech Geol. Soc.,* 46: 9–13.

MELÍN M. and KUNST M., 1992. MinCalc development kit 2.1. Geol. Ůstav Akademie věd, Praha.

RICHARD L.R., 1995. MinPet: Mineralogical and petrological data processing system, version 2.02. MinPet Geological Software, Québec, Canada.