


Graphical and Numerical Analysis of Spatial Data with a GUI and R

Rudolf Dutter


Vienna University of Technology, Austria

Abstract: This paper is on the basis of DAS+R¹, a package of  still under development using a graphical user interface which should ease the application of more or less sophisticated methods. It is heavily relying on Reimann, Filzmoser, Garrett, and Dutter (2008), for which it can represent a companion. The basis of the graphical user interface comes from the R Commander (see Fox, 2004). It uses Tcl/Tk programming tools. The emphasis is on the analysis of spatially depending uni- or multivariate data, particularly on problems of geochemical data.

We describe in short the important features of data analysis in this field with realized functionalities with its clickable icons in DAS+R. Finally, we give a short illustration on a practical example.

Keywords: Data Analysis, Spatial Data, Interactive Software.

1 Introduction


This paper is on the basis of DAS+R, and formally a short description of it, a package of  still under development using a graphical user interface which should ease the application of more or less sophisticated methods.

The basis of the graphical user interface comes from the R Commander (see Fox, 2004). It uses Tcl/Tk programming tools (see Welch and Jones, 2003). The emphasis is on the analysis of spatially depending uni- or multivariate data, particularly on problems of geochemical data.

Three special properties of DAS+R should be stressed:

- Interactive definition of data subsets (numerically or graphically) together with set operations. Usage of these subsets in almost all graphics and computations.
- Intensive use of possible relations between the geographical information with the values of the data in the statistical and graphical analysis.
- The strong requirement of fast reproducibility and repeatability with small variations of the analysis.

The paper consists of a section with a short abstract of the features contained in DAS+R including illustrations produced with the data published in Reimann et al. (1998). Further, some illustrations on spatial analysis including kriging are presented.

The main window of the GUI is shown in Figure 1. It consists of (from top to bottom) a menu bar, a toolbar, a script window, an output window, and a messages window. Most of the -commands are entered via a menu by clicking in the menu bar. In Figure 1 the entry Advanced is clicked such that further submenu buttons show up.

¹<http://www.statistik.tuwien.ac.at/StatDA/DASplusR/>

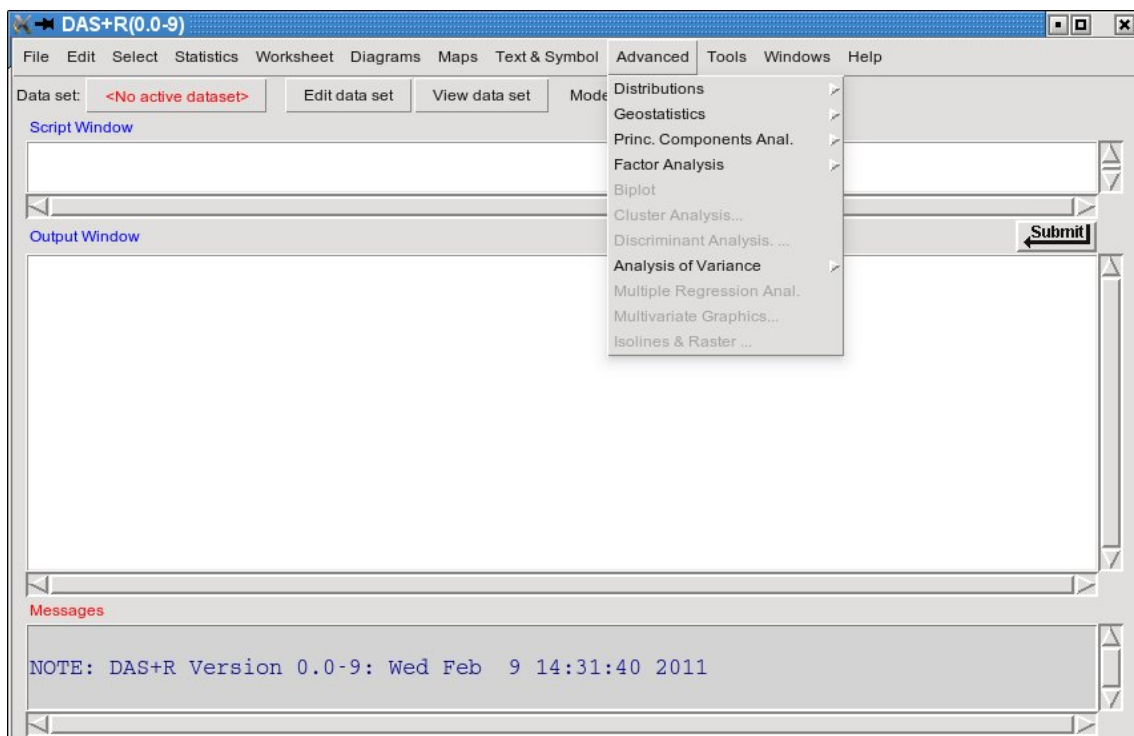


Figure 1: Main graphical user interface of DAS+R



2 Important Features in the Implementation of Data Analysis

Here we describe in short the contents of the main entries in the pull down menu in Figure 1.

Useful Data Structures and Entering Data

The following discussion on user oriented data structures is concerned about the entry 'file' in the pull down menu of DAS+R and is by no means complete or exhaustive but shows some useful features of data coming from people working in applied fields like geosciences.

Data to be analyzed may be entered in different ways:

- By numbers and/or characters to fill a table. (This case will not be very frequent because usually, the data will be prepared beforehand on some storage median in some sort of format.
- From a spreadsheet, usually in a csv-format.
- A data set is already available in the -workspace.
- A data set previously stored in a .R-dump file.
- A data set stored in another package of .

- A data set stored in a file with a known ‘foreign’ format like SPSS, Minitab, STATA, etc.

We only comment on data from spreadsheets: Besides getting numbers or character strings for ‘data values’ in a sort of matrix of values, it is very interesting (and usual) to attribute to each column (variable) a unit, the kind of generation of the data value, and some other comments for each variable. This can be made available by using some common specific keywords in the rows of the spreadsheet. An illustration is shown in Figure 2.

	A	E	F	G	H	I	J	K	L	M	N	O
1	HEADER						KOLA PROJECT, regional sampling 1995 (Finland (FIN), Norway (NOR))					
2	COMMENT DATASET						C-Horizon of Podsol profiles, air dried, fraction <2 mm, nylon screen					
3	SAMPLE IDENTIFIER											
4	COORDINATES											
5	COMMENT VARIABLES							all <DL				
6	EXTRACTION						Aqua Regi	TOTAL	Aqua Regi	TOTAL	TOTAL	Aqua Regi
7	METHOD						GF-AAS	INAA	ICP-AES	XRF	XRF	GF-AAS
8	UDL											
9	LDL						0	5	10	300		0.1
10	UNIT	m a.s.l.			cm		mg/kg	mg/kg	mg/kg	mg/kg	wt.-%	mg/kg
11	VARIABLE	ELEV	COUN	ASP	TOPC	LITO	Ag	Ag_INAA	Al	Al_XRF	Al2O3	As
12		135	FIN	NW	35	20	0.01	2.5	10200	75100	14.19	0.3
13		140	RUS	SW	52	4	0.01	2.5	3540	74400	14.06	0.2
14		255	FIN	N	52	31	0.02	2.5	16100	80600	15.24	0.4
15		240	RUS	NE	40	20	0.02	2.5	12000	79300	15	0.3
16		80	NOR	N	50	10	0.02	2.5	9850	72800	13.76	2.2

Figure 2: Excerpt of a spreadsheet with specific keywords of data

Editing

The entry ‘Edit’ in the menu bar allows specification of editing of the data set, the variables and cases, and all kind of parameters. It also allows to add other data sets by appending or linking (which means binding in respect to columns in this case) to the ‘active data set’. Furthermore, subsets can be generated (numerically or graphically) and/or edited.

We illustrate this graphically by a simple definition of a subset. Figure 3 shows on the left side the widget for specifying necessary attributes like subset name, type of definition, eventually description, etc.

The right side of Figure 3 shows this graphical definition by clicking at the points of a polygon line which encloses the subset.

Note that subsets also can be edited, renamed, deleted, selected, sorted, combined by set operations, or used as a new, restricted active data set.

Selecting

From a (possibly huge) data set, variables, cases and/or subsets can be selected on which will be specially concentrated in the further analysis.

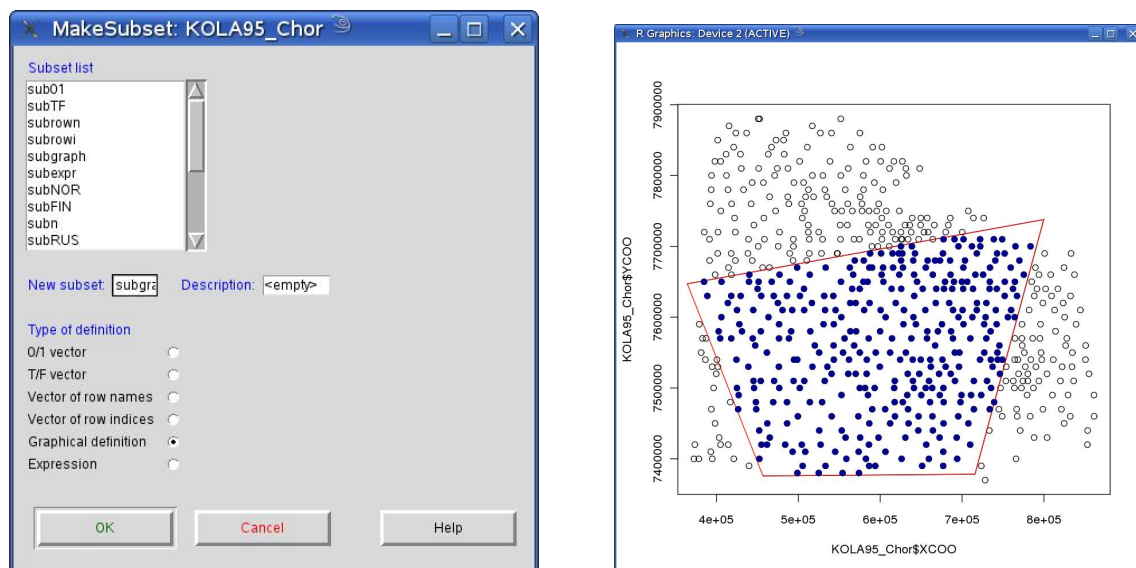



Figure 3: Graphical generation of a data subset

Statistics

Different statistics, e.g. all kind of summaries of the data and simple tests, are available: Tests on normality, contingency tables, tests on mean(s), variance(s), analysis of variance, simple linear regression.

Worksheet

The main idea is that some graphics should be nicely plotted on a ‘worksheet’ of size to be specified. A ‘worksheet’ is splitted into ‘frames’ in which a graphical output of  can be placed. These frames may be of any size and can be placed anywhere in the worksheet (also in an overlapping manner) (see Figure 4 for a simple final result).

Diagrams (Graphics)

Different standard graphics can be produced: histograms, density plots, cdf’s, boxplots, xy-plots, ternary,

An illustration is given in Figure 5 with 3 widgets for plotting density traces: The deepest one shows the specification of two variables (Al and Ca) and grouping of the data by the variable (factor) COUN with 3 levels. The next higher widget shows the default specifications for all the combinations of the 2 variables and the 3 group levels FIN, NOR and RUS. Finally, a graphics window with the 6 requested density traces is shown together with a legend and a question if its position should be changed (interactively).

Maps

Background maps can be imported and used, data plotted in form of maps via different ‘symbols’, proportional dots, surface maps using simple interpolation or even kriging.

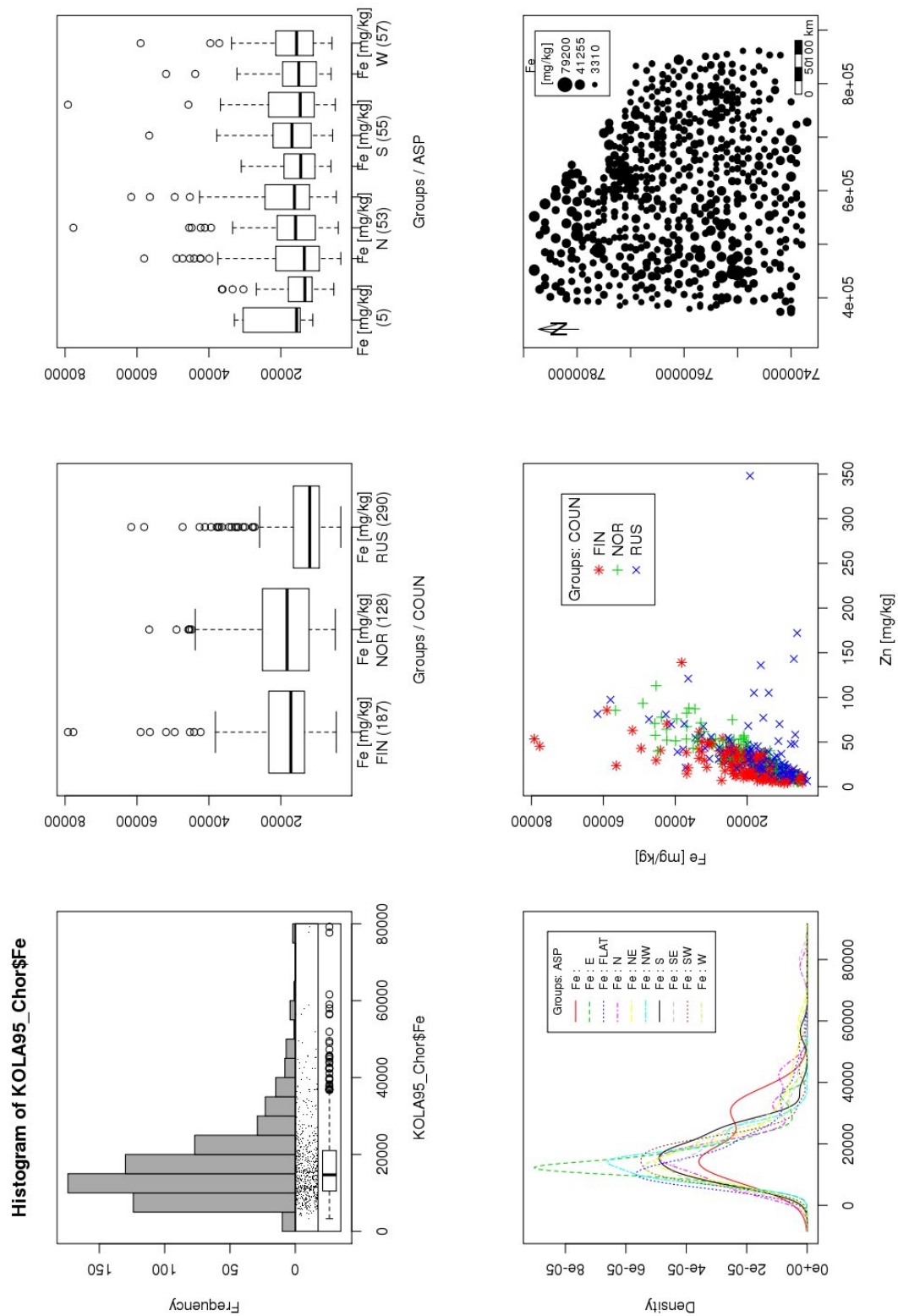


Figure 4: Different graphics placed on a worksheet

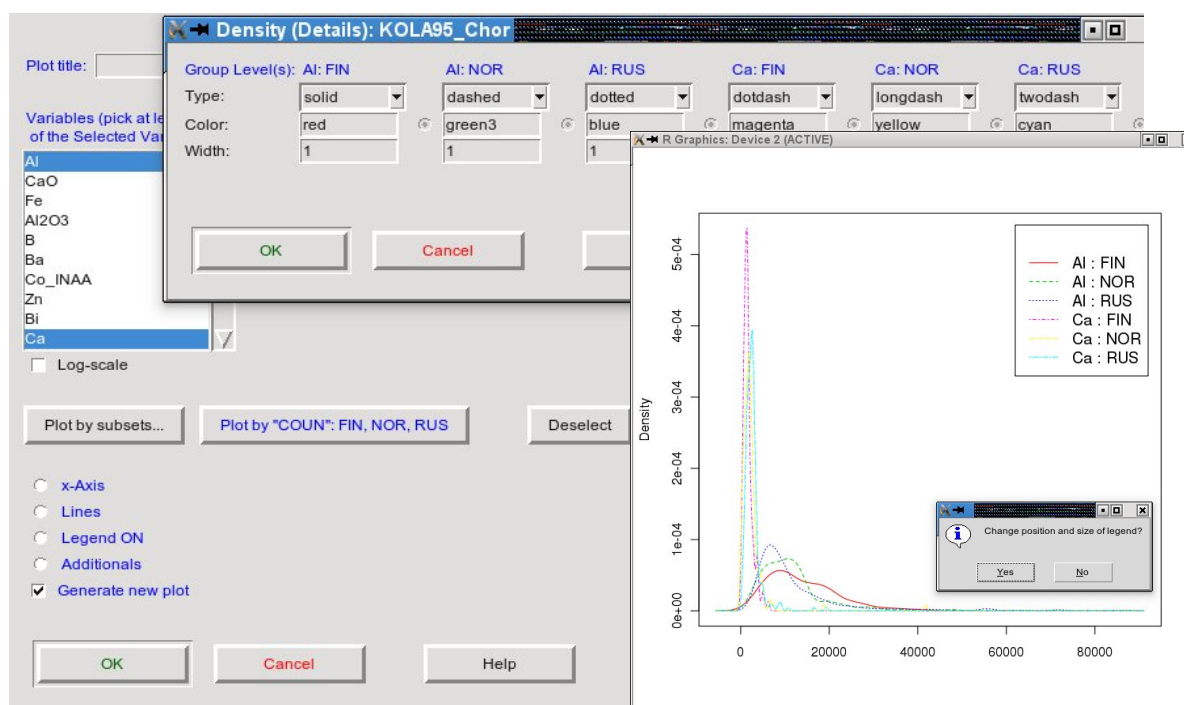


Figure 5: Widgets for density plots

Text and Symbols

Text and/or symbols may be added to a graphic.

Advanced

The entry 'Advanced' opens the following menus.

- Distributions (random numbers!)
- Geostatistics (spatial data!)
- Princ. Components Anal.
- Factor Analysis
- Biplot
- Cluster Analysis
- Discriminant Analysis
- Analysis of Variance
- Multiple Regression Anal.
- Multivariate Graphics
- Isolines & Raster

We only comment on the entry Geostatistics in the next section.

Tools

Here the user gets some auxiliary tools as for loading available packages, showing useful objects like colorsets or the available symbolsets, etc.


Windows

This is the usual icon for arranging the windows, setting up the toolbar, etc.

Help

The usual button for calling some help about the function assembly. However, it should be noted that every function which is visible to the user, in our case also each graphical widget, has a help button which enables to get aid at the specific places.

3 Geostatistics: Example with the Kola Data


Here the functions of the -package `geoR` are internally called and used. The analysis could be done by the following simple clicking sequences:

From the active data set, first a `geoR`-object has to be created:

Advanced → Geostatistics → Create `geoR` Object

(see Figure 6(a)). Then the structural analysis of the random field may be performed by computing empirical variograms with the subsequent estimation of theoretical models:

Advanced → Geostatistics → Empirical Variogram Estimation

(see Figure 6(b) for the graphical result). The estimation of a theoretical model can be done by a numerical procedure or interactively controlled by the -function `eyefit` which allows continuously changing the parameters by moving a slider and visualizing the result simultaneously:

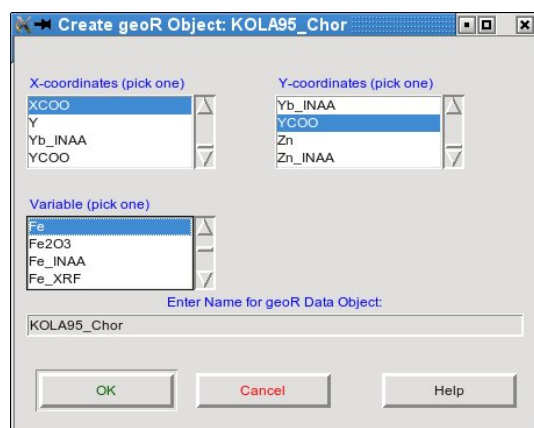
Advanced → Geostatistics → Variogram Function Estimation → Variogram Function - Eye Fit

(see Figure 6(c)). An exponential model with indicated sill, range and nugget effect has been selected. The estimated model is shown simultaneously in Figure 6(d).

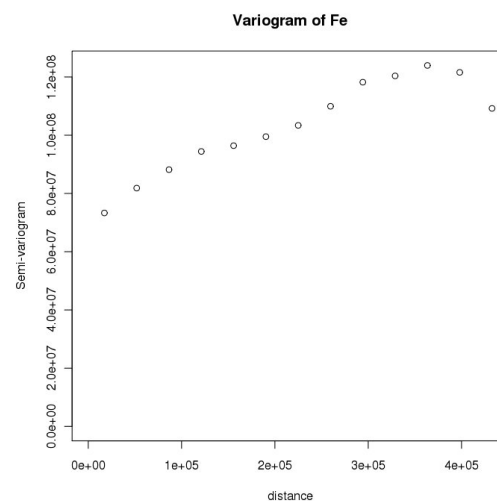
Using the estimated variogram model, a kriging procedure may be invoked for inter- and extrapolation by clicking

Advanced → Geostatistics → Conventional Kriging

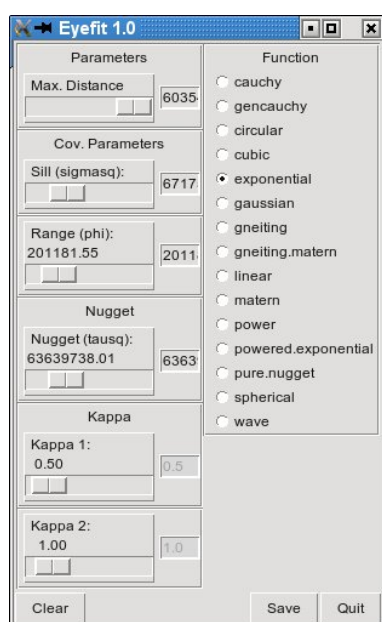
A widget like in Figure 7 appears where only the entry for `Raster size` has to be filled in. Pressing the OK-button will show e.g. the kriged values of Fe in form of a raster plot with isolines as in Figure 8.



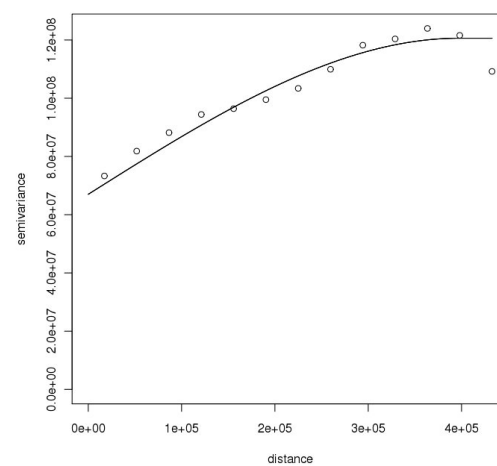
(a)



(b)



(c)



(d)

Figure 6: Empirical variogram and estimated model



Figure 7: Specification of Ordinary Kriging

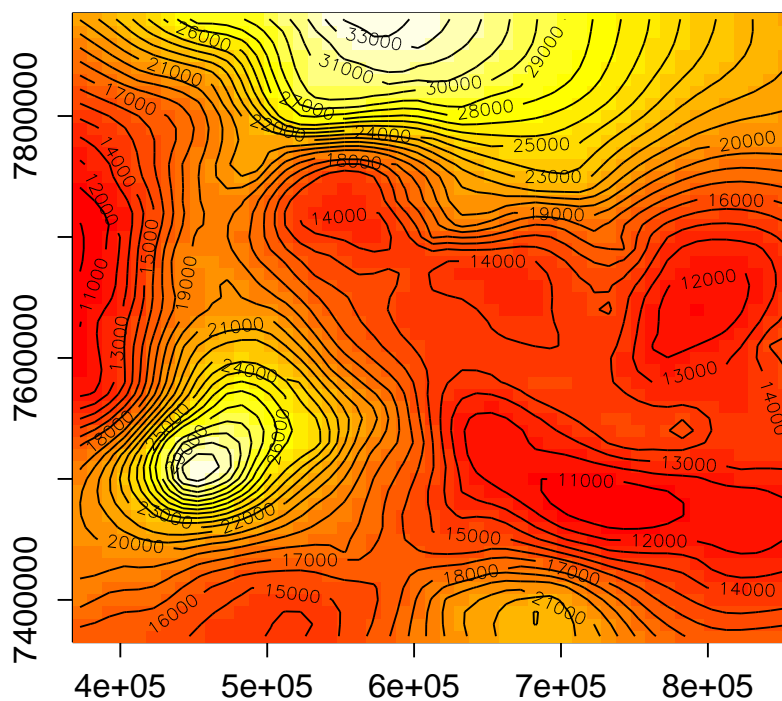


Figure 8: Kriged values of Fe

References

- Fox, J. (2004). Getting started with the R commander: A basic-statistics graphical user interface to R. *J. Statist. Software*, 14(9), 1-42. Available from <http://socserv.socsci.mcmaster.ca/jfox/Misc/Rcmdr/index.html>
- Reimann, C., Äyräs, M., Chekushin, V., Bogatyrev, I., Boyd, R., Caritat, P. de, et al. (1998). *Environmental Geochemical Atlas of the Central Barents Region*. P.O. Box 3006, Lade, N-7002 Trondheim, Norway: Geological Survey of Norway (NGU).
- Reimann, C., Filzmoser, P., Garrett, R., and Dutter, R. (2008). *Statistical Data Analysis Explained. Applied Environmental Statistics with R*. New York: Wiley & Sons.
- Welch, B., and Jones, K. (2003). *Practical Programming in Tcl and Tk*. New York: Prentice Hall PTR.

Author's address:

R. Dutter
Department of Statistics
Vienna University of Technology
Wiedner Hauptstrasse 7 (Gold. Lamm)
A-1040 Vienna, Austria
R.Dutter@TUWien.ac.at