

SimSurvey – a tool for (geo-)statistical analyses with R on the web

Mario Gellrich¹, Rudolf Gubler², Andreas Papritz¹, Andreas Schönborn³,
Rainer Schulin¹

¹Institute of Terrestrial Ecosystems, ETH Zürich (<http://www.ites.ethz.ch>)

²Terraplan Gubler, Lucern (gubler@terraplan.ch)

³Armadillo Media GmbH, Udligenswil (<http://www.armadillo-media.ch>)

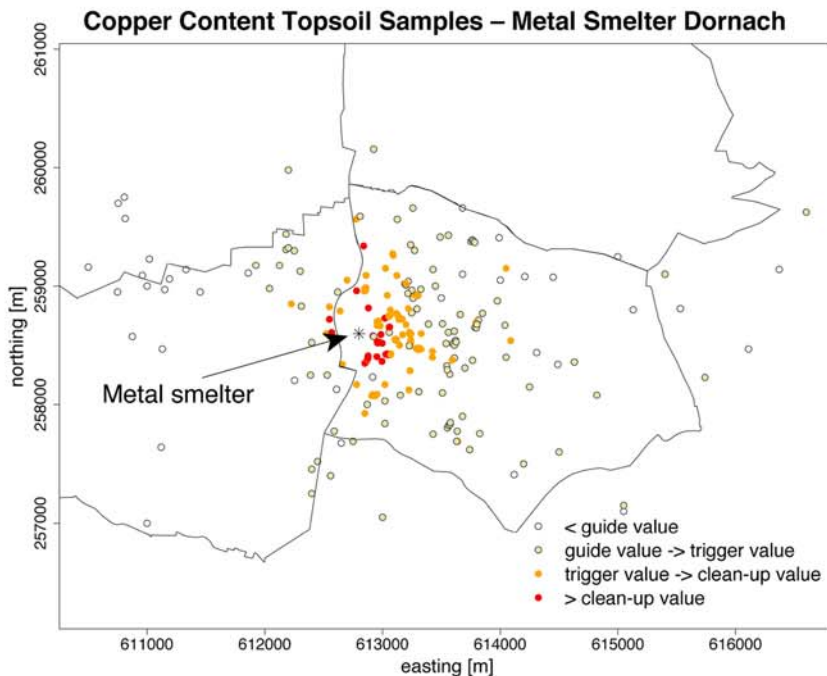
Outline

1. Background & Problem
2. What is *SimSurvey*?
3. Used Software
4. The GUI
5. Example analysis

1. Background & Problem

1. Background & Problem

Background: the assessment of soil pollution is a task that environmental scientists and engineers may face in their daily work. An important part of such an assessment is the spatial delineation of the polluted zone.



Geostatistical
Analyses



1. Background & Problem

Problem: courses in geostatistics are part of the education of environmental scientists and engineers. But geostatistics is rather difficult to teach. Apart from the mathematics, the lack of powerful, flexible, but easy-to-use software increases the difficulties.

Universal Kriging

- UK predictor

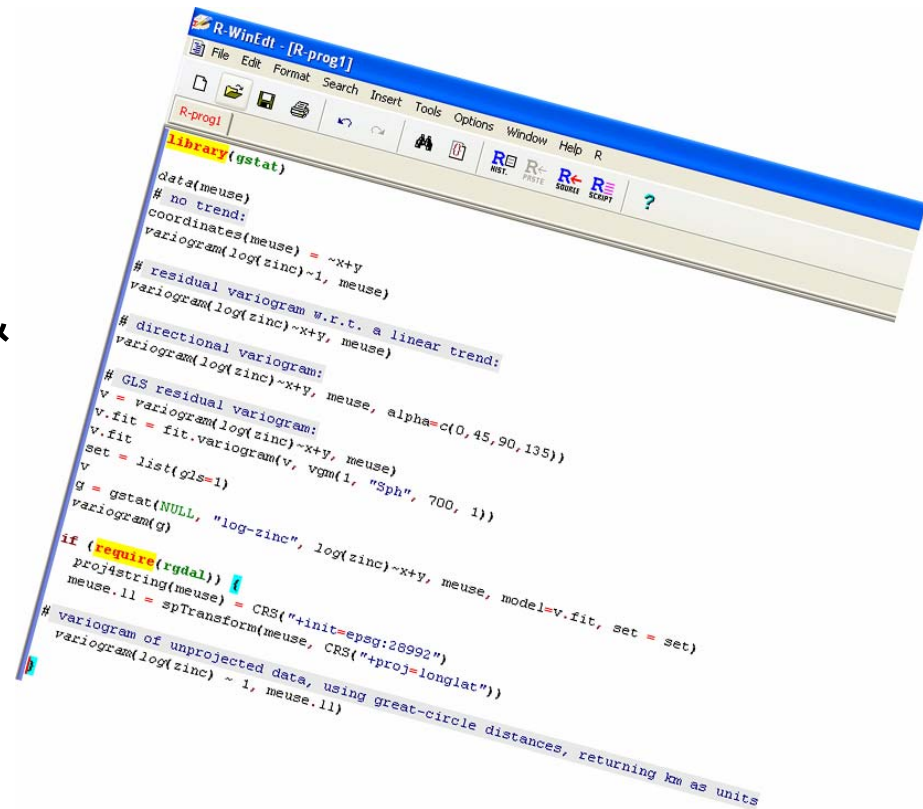
$$\hat{Y}_{UK}(s_0) = x_0' \hat{\beta}_{GLS} + \sigma' \Sigma^{-1} (Y - X \hat{\beta}_{GLS}) = \kappa' Y$$

- Universal Kriging Variance

$$\sigma_{UK}^2(\hat{Y}(s_0)) = C(0) - \sigma' \Sigma^{-1} \sigma + (x_0 - X' \Sigma^{-1} \sigma)' (X' \Sigma^{-1} X)^{-1} (x_0 - X' \Sigma^{-1} \sigma)$$

- same properties as SK predictor
- homogeneous BLUP

&



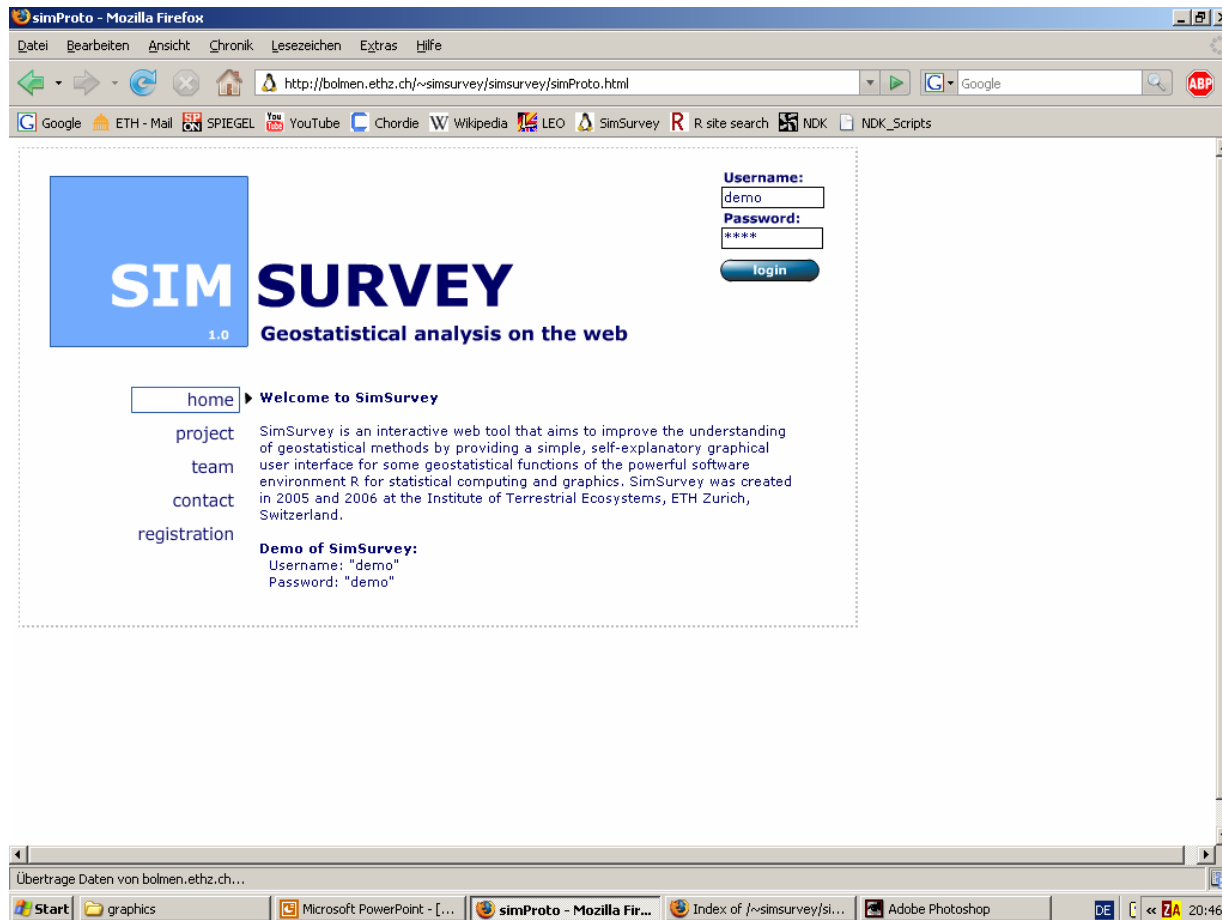
```
R-WinEdt - [R-prog1]
File Edit Format Search Insert Tools Options Window Help R
R-prog1
Library(gstat)
data(meuse)
# no trend:
coordinates(meuse) = ~x+y
variogram(log(zinc)~1, meuse)
# residual variogram w.r.t. a linear trend:
variogram(log(zinc)~x+y, meuse)
# directional variogram:
variogram(log(zinc)~x+y, meuse, alpha=c(0,45,90,135))
# GLS residual variogram:
v = variogram(log(zinc)~x+y, meuse)
v.fit = fit.variogram(v, vgm(1, "Sph", 700, 1))
set = list(gls=1)
g = gstat(NULL, "log-zinc", log(zinc)~x+y, meuse, model=v.fit, set = set)
variogram(g)
if (require(rgdal)) {
  proj4string(meuse) = CRS("+init=epsg:26992")
  meuse.ll = spTransform(meuse, CRS("+proj=longlat"))
# variogram of unprojected data, using great-circle distances, returning km as units
variogram(log(zinc) ~ 1, meuse.ll)
```

= often too difficult

2. What is *SimSurvey*?

2. What is *SimSurvey*?

SimSurvey is a graphical user interface with *R*. It can be used for (geo-) statistical analyses and teaching. The user interacts with *R*, running on a web server, using a menu in a browser window.



2. What is *SIMSURVEY*?

SimSurvey allows the user to test sampling strategies using a virtual sampling environment and simulated data.

The screenshot displays the SIMSURVEY 1.0 software interface. At the top, the status bar shows 'User: newuser', 'Project: project_001', 'Session: session_001', and a green 'dataset exists' indicator. The main menu includes 'Data', 'Graphics', 'Regression', 'Geostatistic', 'R-console', 'Windows', and 'Help'. The 'Data' menu is open, showing options: 'Import data', 'Take samples', 'View data', 'Edit data', and 'Export data'. The 'Take samples' option is highlighted. A 'session level' label is visible at the bottom left of the main interface.

The 'Take data samples' dialog box is open, featuring a tabbed interface with 'map1', 'map2', 'map3', and 'map4'. The 'map1' tab is active, showing an aerial orthophoto of a city area with numerous yellow sampling points. A red circle highlights a specific point with a 'Click me' label. On the left side of the dialog, there is a 'Select target variables' section with the following checked items: 'cd', 'cu', 'developed', 'forest', and 'geology'. Below this is a 'Sampling:' section with 'START' and 'STOP' buttons, and a 'Map licences' button at the bottom.

3. Used Software

3. Used Software

SimSurvey is installed on a Linux-Server and runs in a browser window.

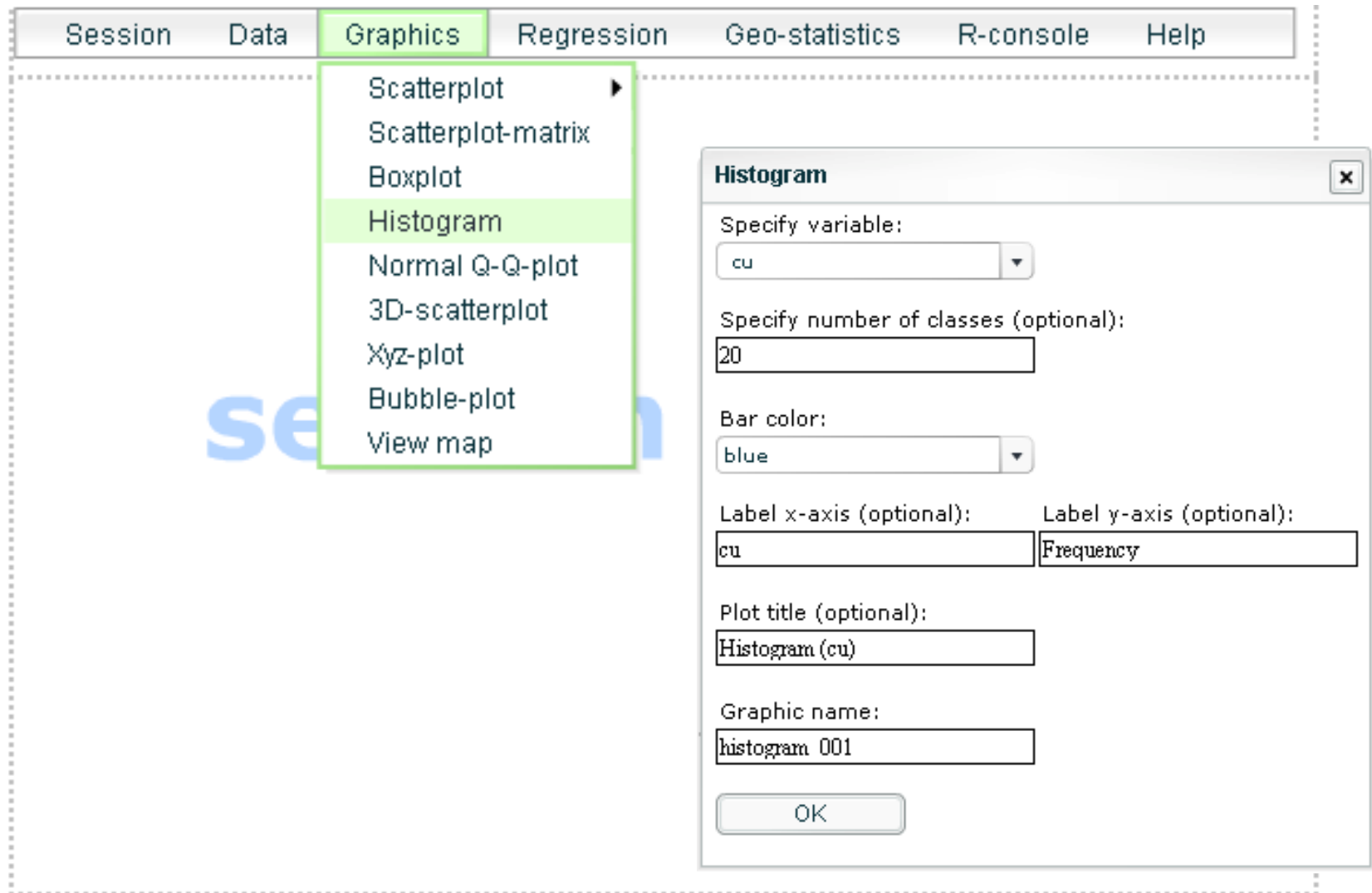
The following software is used to run SimSurvey:

- Linux (operating system)
- Apache (web-server)
- Macromedia Flash Player (visualisation)
- MySQL (database to block 'dangerous' commands in the R-console)
- PHP (interface between Flash, R and MySQL)
- R (statistics, text-output, graphics-output)

4. The GUI

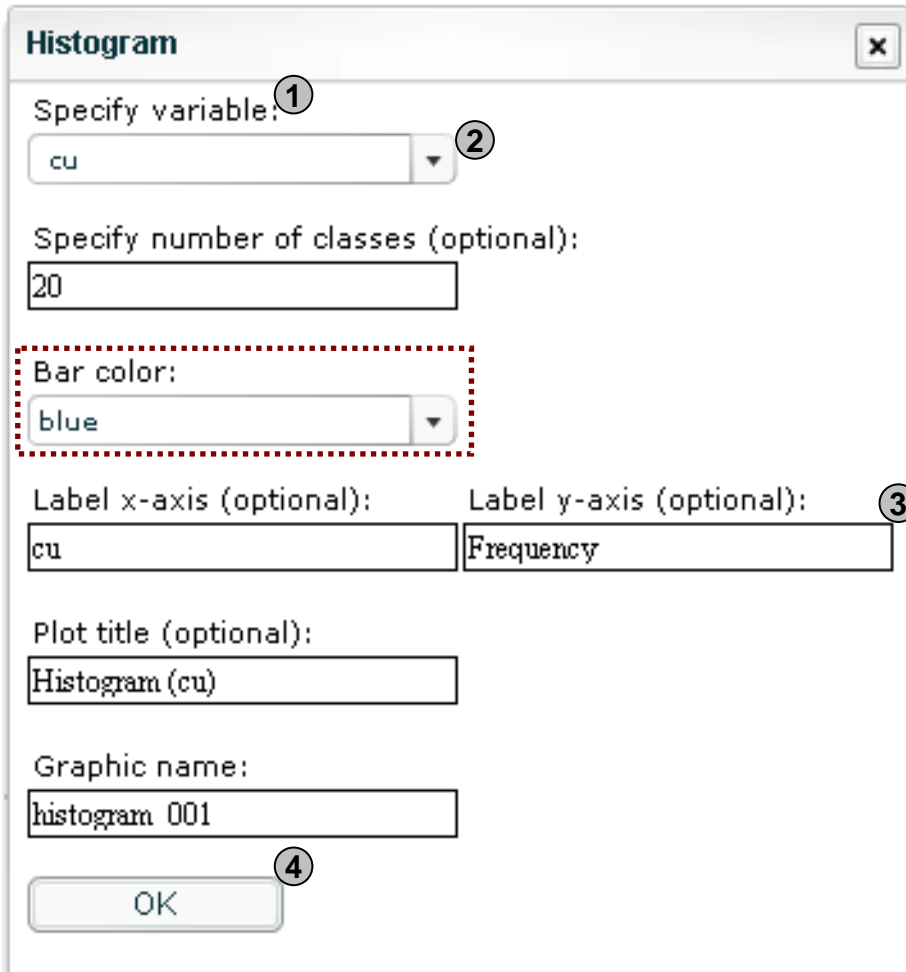
4. The GUI

At the heart of *SimSurvey*'s GUI is a statistics menu.



4. The GUI

The statistics menu and dialog boxes are created by Flash using an XML-file.



The image shows a 'Histogram' dialog box with the following fields and controls:

- Specify variable:** A label (1) followed by a dropdown list (2) containing 'cu'.
- Specify number of classes (optional):** A text box containing '20'.
- Bar color:** A dropdown list (3) containing 'blue', highlighted with a red dashed border.
- Label x-axis (optional):** A text box containing 'cu'.
- Label y-axis (optional):** A text box (3) containing 'Frequency'.
- Plot title (optional):** A text box containing 'Histogram (cu)'.
- Graphic name:** A text box containing 'histogram 001'.
- OK button:** A button (4) labeled 'OK'.

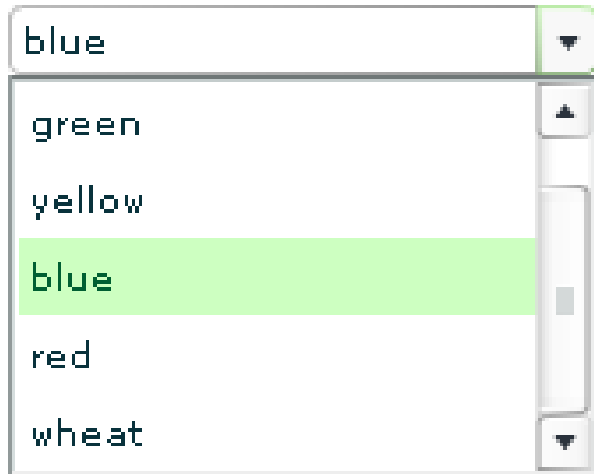
Dialog box items:

- ① label
- ② dropdown list
- ③ textbox
- ④ button

4. The GUI

In the XML-code of the dialog box, each item/variable has a specific value.

Bar color:



XML-code

```
<outputField column="left" text="Bar color:"
name="hisTxt003"/>
  <dropDown column="left" name="hisDdmCol">
    <dropDownItem text="gray" value="gray"/>
    <dropDownItem text="green" value="green"/>
    <dropDownItem text="yellow" value="yellow"/>
    <dropDownItem text="blue" value="blue"/>
    <dropDownItem text="red" value="red"/>
    <dropDownItem text="wheat" value="wheat"/>
  </dropDown>
```

Flash passes the (user-defined) values of the different variables to PHP.

4. The GUI

PHP takes the value of the Flash-variable and assigns it to a PHP-variable.

```
if ($_REQUEST["hisDdmV01"]) //graphics, histogram
{
$ausdruck=$_REQUEST["hisTinGrn"]; //title
if (!$ausdruck)
{
$ausdruck="tmp";
}
$variable1=$_REQUEST["hisDdmV01"]; //variable
$eigenschaften1=$_REQUEST["hisTinCla"]; //number of classes
$eigenschaften2=$_REQUEST["hisDdmCol"]; //barcolor
$eigenschaften3=$_REQUEST["histTinXax"]; //label x
$eigenschaften4=$_REQUEST["histTinYax"]; //label y
$eigenschaften5=$_REQUEST["hisTinTit"]; //plot titel
$arbeitsverzeichnis=$projektpfad;
$maske="$pfad/include/graphics/histogram.R";
if (!file_exists("$projektpfad/$ausdruck/"))
{
mkdir("$projektpfad/$ausdruck/", 0777);
chmod ("$projektpfad/$ausdruck/", 0777);
}
$grafikverzeichnis=$ausdruck;
}
```

PHP-code

\$eigenschaften2=\$_REQUEST["hisDdmCol"]; //barcolor ←

4. The GUI

The PHP-variable is part of an R-script containing 'dynamic' script parts.

```
setwd("**Arbeitsverzeichnis**/")
load(".RData")
library(grDevices)

user.workspace    <- "**Arbeitsverzeichnis**"

##### dynamic part 1 ##### start
graphic.name      <- "**Ausdruck**"
##### dynamic part 1 ##### end

path.to.new.workspace <- paste(c(user.workspace, "/", graphic.name), collapse="")
setwd(path.to.new.workspace)
load(paste(c(user.workspace, "/", ".RData"), collapse=""))

##### dynamic part 2 ##### start

sel.var           <- "**variable1**"
number.classes    <- "**eigenschaften1**"
bar.color         <- "**eigenschaften2**"
label.xaxis       <- "**eigenschaften3**"
label.yaxis       <- "**eigenschaften4**"
plot.title        <- "**eigenschaften5**"

##### dynamic part 2 ##### end

if(number.classes == "") {
hist(dat[[sel.var]], nclass=nrow(dat[[sel.var]])/10, col=bar.color,
     xlab=label.xaxis, ylab=label.yaxis, main= plot.title, cex.main=1)
}

if(number.classes != ""){
hist(dat[[sel.var]], nclass=number.classes, col=bar.color,
     xlab=label.xaxis, ylab=label.yaxis , main= plot.title, cex.main=1)
}

save(list = ls(all=TRUE), file = ".RData")
setwd(user.workspace)
quit(save = "yes",status=0,runLast = FALSE)
```

R-code

bar.color <- "eigenschaften2**"**

4. The GUI

PHP translates the 'dynamic' R-script into a 'standard' R-script and runs it.

```
setwd("/home/mariog/public_html/simsurvey/tmp/ruedi/dornach_project/dornach/")

load(".RData")
library(grDevices)
user.workspace      <- "/home/mariog/public_html/simsurvey/tmp/ruedi/dornach_project/dornach"

##### dynamic part 1 ##### start

graphic.name        <- "histogram_001"

##### dynamic part 1 ##### end

path.to.new.workspace <- paste(c(user.workspace, "/", graphic.name), collapse="")
setwd(path.to.new.workspace)
load(paste(c(user.workspace, "/", ".RData"), collapse=""))

##### dynamic part 2 ##### start

sel.var             <- "var4"
number.classes      <- 20
bar.color           <- "blue"
label.xaxis         <- "cu"
label.yaxis         <- "Frequency"
plot.title          <- "Histogram (cu)"

##### dynamic part 2 ##### end

if(number.classes == ""){
  hist(dat[[sel.var]], nclass=nrow(dat[[sel.var]])/10, col=bar.color, xlab=label.xaxis, ylab=label.yaxis, main= plot.title, cex.main=1)
}

if(number.classes != ""){
  hist(dat[[sel.var]], nclass=number.classes, col=bar.color, xlab=label.xaxis, ylab=label.yaxis , main= plot.title, cex.main=1)
}

save(list = ls(all=TRUE), file = ".RData")
setwd(user.workspace)
quit(save = "yes",status=0,runLast = FALSE)
```

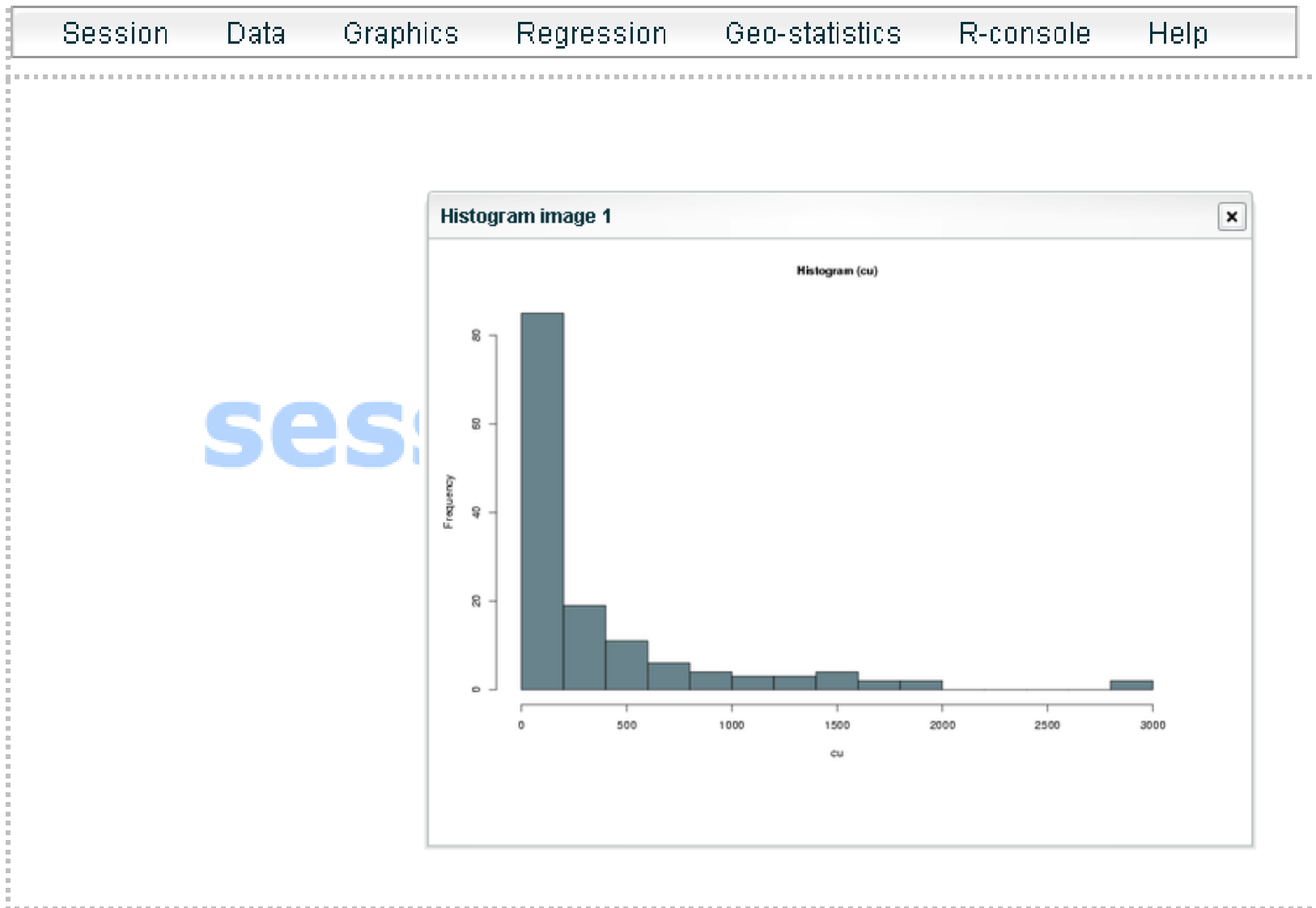
R-code

→ **bar.color <- "blue"**

→ **col=bar.color**

4. The GUI

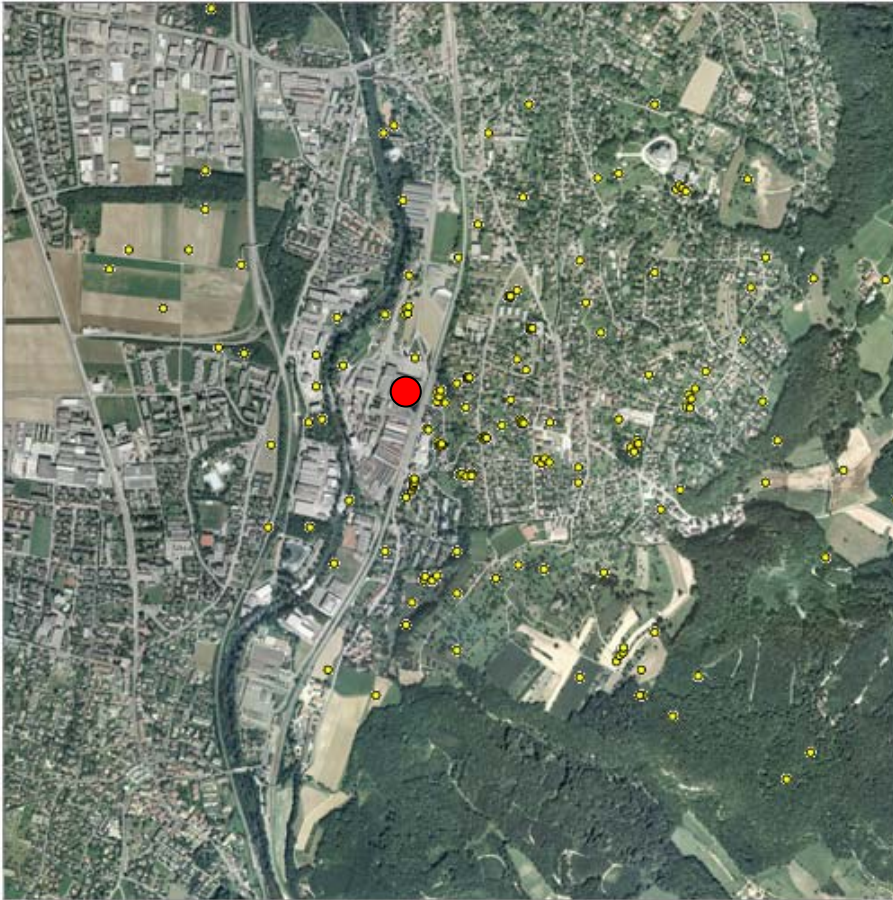
Flash is used to visualize the R-graphic in a browser window.



5. Example analysis

5. Example analysis

Geostatistical analysis using topsoil samples from an area surrounding a metal smelter in the municipality of Dornach (Switzerland).



Objective: delineate those areas, where the copper content exceeds guide, trigger, and cleanup values of the ordinance.

- location metal smelter
- topsoil samples (simulated data)

5. Example analysis: create your account & log in

The screenshot shows a Mozilla Firefox browser window with the following elements:

- Browser Title:** simProto - Mozilla Firefox
- Menu Bar:** Datei, Bearbeiten, Ansicht, Chronik, Lesezeichen, Extras, Hilfe
- Address Bar:** http://bolmen.ethz.ch/~simsurvey/simsurvey/simProto.html
- Toolbar:** ETHZ, ETH - LUE, Blackboard, LEO, Google, ETH - Journals, ArcGIS 9.2 Help, R site search, Telefonbuch Schwei

The main content area is enclosed in a dashed border and contains:

- Logo:** A blue square with the text "SIM SURVEY 1.0" in white.
- Text:** "Geostatistical analysis on the web"
- Navigation Menu:** home, project, team, contact, registration (highlighted with a blue box and a right-pointing arrow).
- Text:** "For free use of SimSurvey, create your own account. Choose username, password and press OK !"
- Form Fields:** Username: newuser, Password: *****
- Buttons:** login (blue button), OK (grey button)

5. Example analysis: create new project and session

The screenshot displays the SIM SURVEY 1.0 application interface. At the top left, the logo 'SIM SURVEY 1.0' is shown. The top right corner displays the user 'newuser', the project 'project_001', and the session field. A menu on the left side includes options like 'Project resources', 'Buy new resources', 'Start session', 'Delete session', 'Help', and 'Quit project'. The 'Start session' menu is open, showing 'Create new session' and 'Load existing session'. A dialog box titled 'Create new session' is open, prompting the user to specify a session name, with 'session_001' entered in the input field and an 'OK' button below.

SIM SURVEY
1.0

User: newuser Project: project_001 Session:

Project resources ▶
Buy new resources
Start session ▶
Delete session
Help
Quit project

Create new session
Load existing session

Create new session [X]

Please specify a session (i.e. session_1) and press OK

session_001

OK

project level

5. Example analysis: import soil sample data

The screenshot displays the SIM SURVEY 1.0 software interface. At the top, the status bar shows 'User: newuser', 'Project: project_001', and 'Session: session_001'. A red button labeled 'no dataset' is visible in the top right corner. The main menu bar includes 'Data', 'Graphics', 'Regression', 'Geostatistic', 'R-console', 'Windows', and 'Help'. The 'Data' menu is open, showing options: 'Import data', 'Take samples', 'View data', 'Edit data', and 'Export data'. A dialog box titled 'Import data from web-server' is open, containing the following fields and options:

- File location on web-server (URL):
- Specify separator:
- Include Header?:
- OK button

On the left side, a sidebar contains the following options: 'Project resources', 'Buy new resources', 'Delete session items', 'Save session as...', 'Help', and 'Quit session'. The text 'session level' is located at the bottom left of the sidebar area.

5. Example analysis: if necessary, manipulate data

The screenshot displays the SIM SURVEY software interface. At the top, the status bar shows 'User: newuser', 'Project: project_001', 'Session: session_001', and a green indicator 'dataset exists'. The main menu includes 'Data', 'Graphics', 'Regression', 'Geostatistic', 'R-console', 'Windows', and 'Help'. The 'Data' menu is open, showing options like 'Import data', 'Take samples', 'View data', 'Edit data', and 'Export data'. The 'Edit data' sub-menu is also open, listing actions such as 'Create or transform variable', 'Rename variable', 'Remove variable', 'Convert numeric to factor', 'Standardize variable', 'Remove NA's', and 'Create subset'. The 'Convert numeric to factor' option is further expanded to show 'Cut variable into classes' and 'Label classes'. On the left side, a sidebar contains project management options: 'Project resources', 'Buy new resources', 'Delete session items', 'Save session as...', 'Help', and 'Quit session'. At the bottom left, the text 'session level' is visible. Below the menu, a 'View data' window is open, displaying a table with 13 rows and 11 columns (A-K). The table contains numerical and categorical data for each row.

A	B	C	D	E	F	G	H	I	J	K
1	x	y	dist	azimuth	project	developed	forest	geology	cd	cu
2	613524.2	258894.5	761.58	66.8	"kl"	"d1964"	"n"	"other"	0.99	32.5
3	613274.2	258769.5	482.83	68.75	"kl"	"d1945"	"n"	"other"	2.16	169.2
4	614074.2	259144.5	1365.65	66.25	"kl"	"no"	"n"	"other"	1.4	42.7
5	613249.2	258119.5	637.38	138.18	"kl"	"no"	"n"	"other"	1.87	112.6
6	612874.2	258344.5	254.95	168.69	"kl"	"d1877"	"n"	"other"	2.24	1840.5
7	612549.2	258594.5	275	270	"kl"	"d1976"	"n"	"other"	1.25	46.5
8	612574.2	258714.5	277.31	295.64	"kl"	"d1970"	"n"	"other"	0.59	379.9
9	612574.2	258819.5	336.34	311.99	"kl"	"d1964"	"n"	"other"	0.93	286.8
10	612644.2	258944.5	393.57	332.78	"kl"	"d1970"	"n"	"other"	0.81	251.1
11	612424.2	258519.5	406.97	259.38	"kl"	"d1988"	"n"	"other"	0.78	97
12	612249.2	258844.5	627	293.5	"kl"	"no"	"n"	"other"	0.46	40
13	612324.2	259119.5	725	316.4	"kl"	"no"	"n"	"other"	0.54	108.8

5. Example analysis: graphically examine data

User: newuser Project: project_001 Session: session_001 dataset exists

SIM SURVEY 1.0

Data **Graphics** Regression Geostatistic R-console Windows Help

- Scatterplot
- Scatterplot-matrix
- Boxplot
- Histogram**
- Normal QQ-plot
- 3D-scatterplot
- XYZ-plot
- Bubble-plot
- View map
- Graphic settings

Project resources ▶
Buy new resources
Delete session items
Save session as...
Help
Quit session

session 1

Scatterplot-matrix image 1

Boxplot image 2

Bubble-plot image 3

undefined image 5

Frequency

copper

5. Example analysis: fit variogram model(s)

SIM SURVEY 1.0

User: newuser Project: project_001 Session: session_001 dataset exists

Data Graphics Regression **Geostatistic** R-console Windows Help

- Autocorrelation ▶ Sample variogram
- Kriging ▶ Variogram model(s)

Project resources ▶
Buy new resources
Delete session items
Save session as...
Help
Quit session

Variogram model(s)

Select sample variogram:
samp_variog

Covariance function(s):
 Exponential
 Spherical
 Gaussian
 Matern

Start values for nugget, range and partial sill:
Nugget: 0 Range: 150
Partial sill: 0.08 Kappa (for Matern model): 0.5

Anisotropy parameters (for directional variograms only):
Main axis direction: Anisotropy ratio:

Graphic title (optional):

Save variogram model as:
vgm_001

OK Help?

Server output message

```
-----  
Fitted parameters: vgm_001  
-----  
Exponential model  
nugget= 0.0236  
psill= 0.0692  
range= 132.477  
SSErr= 7.994059e-06  
-----  
Spherical model  
nugget= 0.0292  
psill= 0.0626
```

Variogram model(s) image 1

The plot displays semi-variance on the y-axis (ranging from 0.00 to 0.10) against distance in meters on the x-axis (ranging from 0 to 2000). Black dots represent the observed data points. Four colored lines represent different variogram models: Exponential (green), Spherical (blue), Gaussian (red), and Matern (orange). All models show a rapid increase in semi-variance that levels off as distance increases, with the Spherical model providing a very close fit to the data points.

5. Example analysis: predict values at unobserved sites

The screenshot displays the SIM SURVEY 1.0 software interface. At the top, the user is identified as 'newuser', the project as 'project_001', and the session as 'session_001'. A green status bar indicates 'dataset exists'. The main menu includes 'Data', 'Graphics', 'Regression', 'Geostatistic', 'R-console', 'Windows', and 'Help'. The 'Geostatistic' menu is open, showing 'Autocorrelation' and 'Kriging'. The 'Kriging' sub-menu is also open, listing 'Maps', 'Predictions', and 'Cross-validation'. On the left, a sidebar contains options: 'Project resources', 'Buy new resources', 'Delete session items', 'Save session as...', 'Help', and 'Quit session'. Below the sidebar, it says 'session level'. The 'Kriging predictions' dialog box is open, with the following fields and options:

- Specify x- and y- coordinates and variable:
 - x-coordinate: x
 - y-coordinate: y
 - Variable: log10cu
- Spatial trend (optional):
 $\log_{10}(\sqrt{(x-612800)^2 + (y-257600)^2})$
- Variogram model:
vgm(nugget=0, range=132, psill=0.069, model="Exp")
- Define grid of prediction points:
 - EITHER use default grid
 - OR enter coordinates:
 - x-coordinate (from, to, by):
 - y-coordinate (from, to, by):
 - OR read grid and further data from file (URL):
- Save kriging predictions as:
kriging_pred

Buttons for 'OK' and 'Help?' are at the bottom of the dialog. To the right, a 'Kriging predictions image 1' window displays the instruction: 'Go to Geostatik – Kriging – Maps, to plot Kriging–predictions!'.

5. Example analysis: visualize model predictions

SIM SURVEY 1.0

User: newuser Project: project_001 Session: session_001 dataset exists

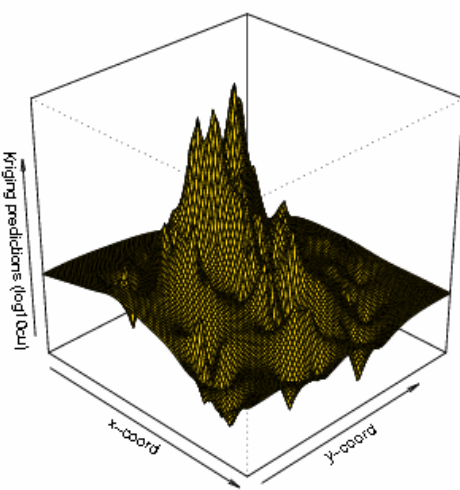
Data Graphics Regression **Geostatistic** R-console Windows Help

- Autocorrelation
- Kriging**
 - Maps
 - Levelplot
 - Contourplot
 - 3D-plot
 - Predictions
 - Cross-validation

Project resources
Buy new resources
Delete session items
Save session as...
Help
Quit session

Kriging 3D-plot image 3

Kriging predictions



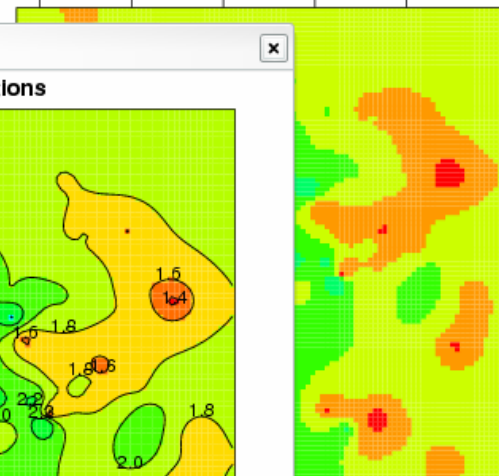
Kriging predictions (log scale)

x-coord

y-coord

Kriging levelplot image 1

Kriging predictions

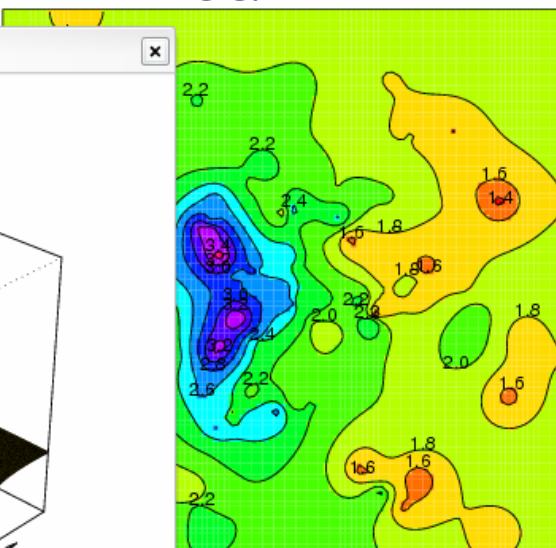


13500 614000

3.0

Kriging contourplot image 2

Kriging predictions



x-coord

2.0 2.5 3.0 3.5

Kriging 3D-plot

Select kriging predictions:
kriging_pred

Select color:
white

Label y-axis (optional):

Plot title (optional):

Vertic. angle (default 30):
30

Help?

Many thanks for your attention!

Visit SimSurvey on:

<http://bolmen.ethz.ch/~simsurvey/simsurvey/simProto.html>

Contact:

Andreas Papritz: andreas.papritz@env.ethz.ch

Mario Gellrich: mario.gellrich@env.ethz.ch