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

Mario Gellrich[1], Rudolf Gubler[2], Andreas Papritz[1], Andreas Schönborn[3], Rainer Schulin[1]

[1]Institute of Terrestrial Ecosystems, ETH Zürich (http://www.ites.ethz.ch)
[2]Terraplan Gubler, Lucern (gubler@terraplan.ch)
[3]Armadillo Media GmbH, Udligenswil (http://www.armadillo-media.ch)

ETH Eidgenössische Technische Hochschule Zürich Swiss Federal Institute of Technology Zurich

ITES Institute of Terrestrial Ecosystems

SIM SURVEY 1.0 Geostatistical analysis on the web

Outline

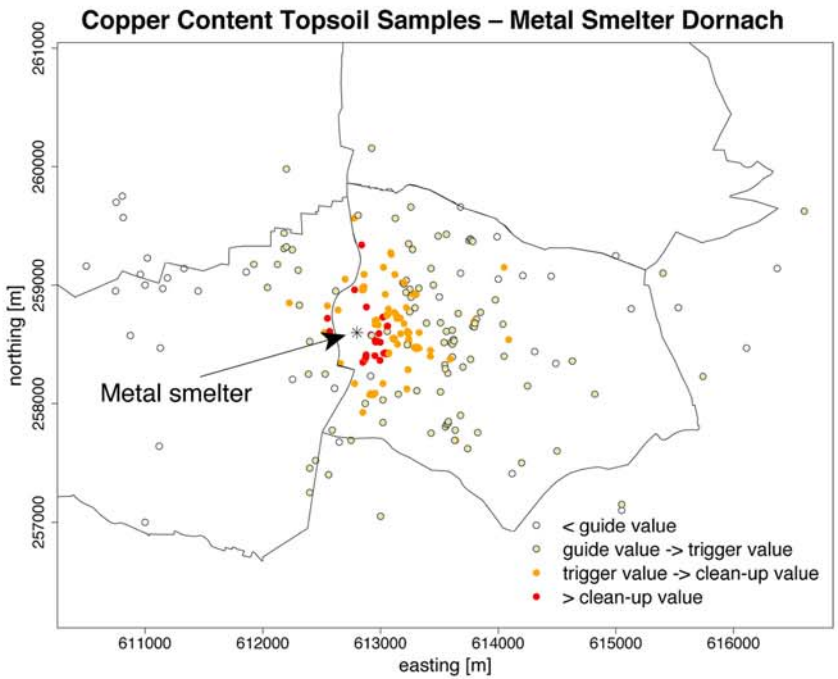
1. Background & Problem
2. What is *SimSurvey*?
3. Used Software
4. The GUI
5. Example analysis

1. Background & Problem

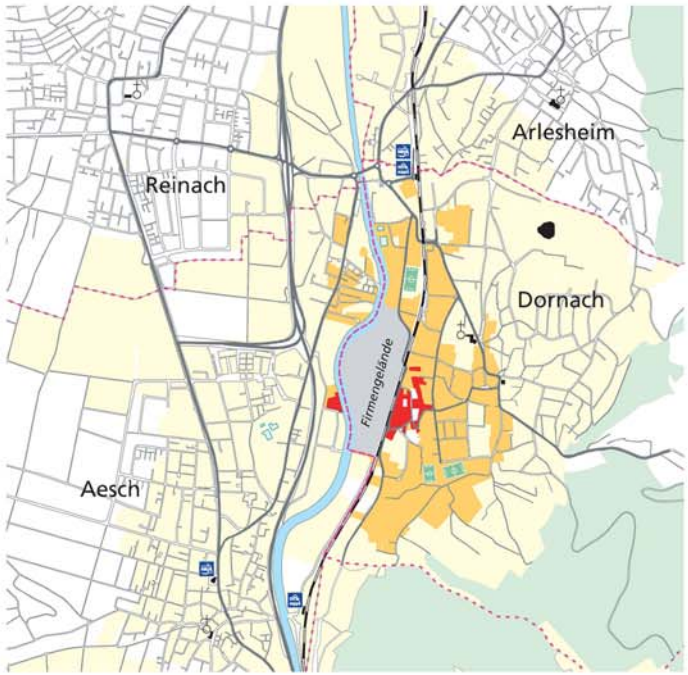
1. Background & Problem

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.

(1)



Geostatistical Analyses



(1) source: http://www.so.ch/fileadmin/internet/bjd/bumaa/pdf/boden/245_mb_02.pdf

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

$$\begin{aligned}\widehat{Y}_{\mathrm{UK}}(s_0) &= x_0'\widehat{\beta}_{\mathrm{GLS}} + \sigma'\Sigma^{-1}(Y - X\widehat{\beta}_{\mathrm{GLS}}) \\ &= \kappa' Y \ ,\end{aligned}$$

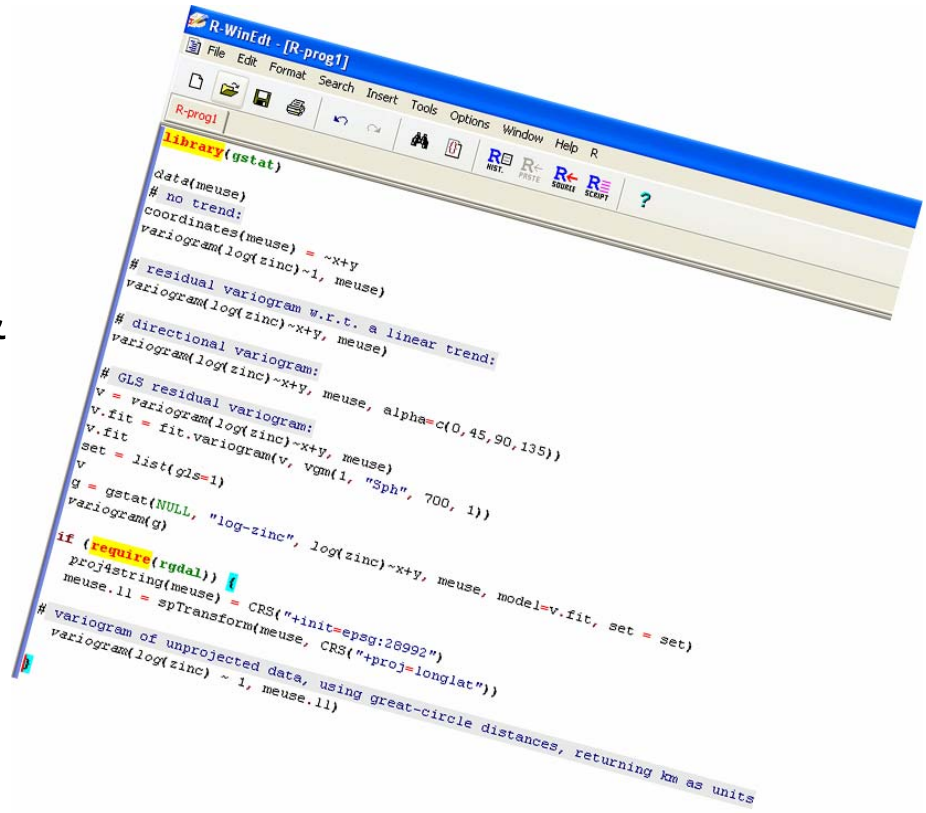
$$\kappa' = \{\sigma + X(X'\Sigma^{-1}X)^{-1}(x_0 - X'\Sigma^{-1}\sigma)\}'\Sigma^{-1}$$

- Universal Kriging Variance

$$\begin{aligned}\sigma^2_{\mathrm{UK}}(\widehat{Y}(s_0)) &= C(0) - \sigma'\Sigma^{-1}\sigma \\ &\quad + (x_0 - X'\Sigma^{-1}\sigma)'(X'\Sigma^{-1}X)^{-1}(x_0 - X'\Sigma^{-1}\sigma)\end{aligned}$$

- same properties as SK predictor
- homogeneous BLUP

&

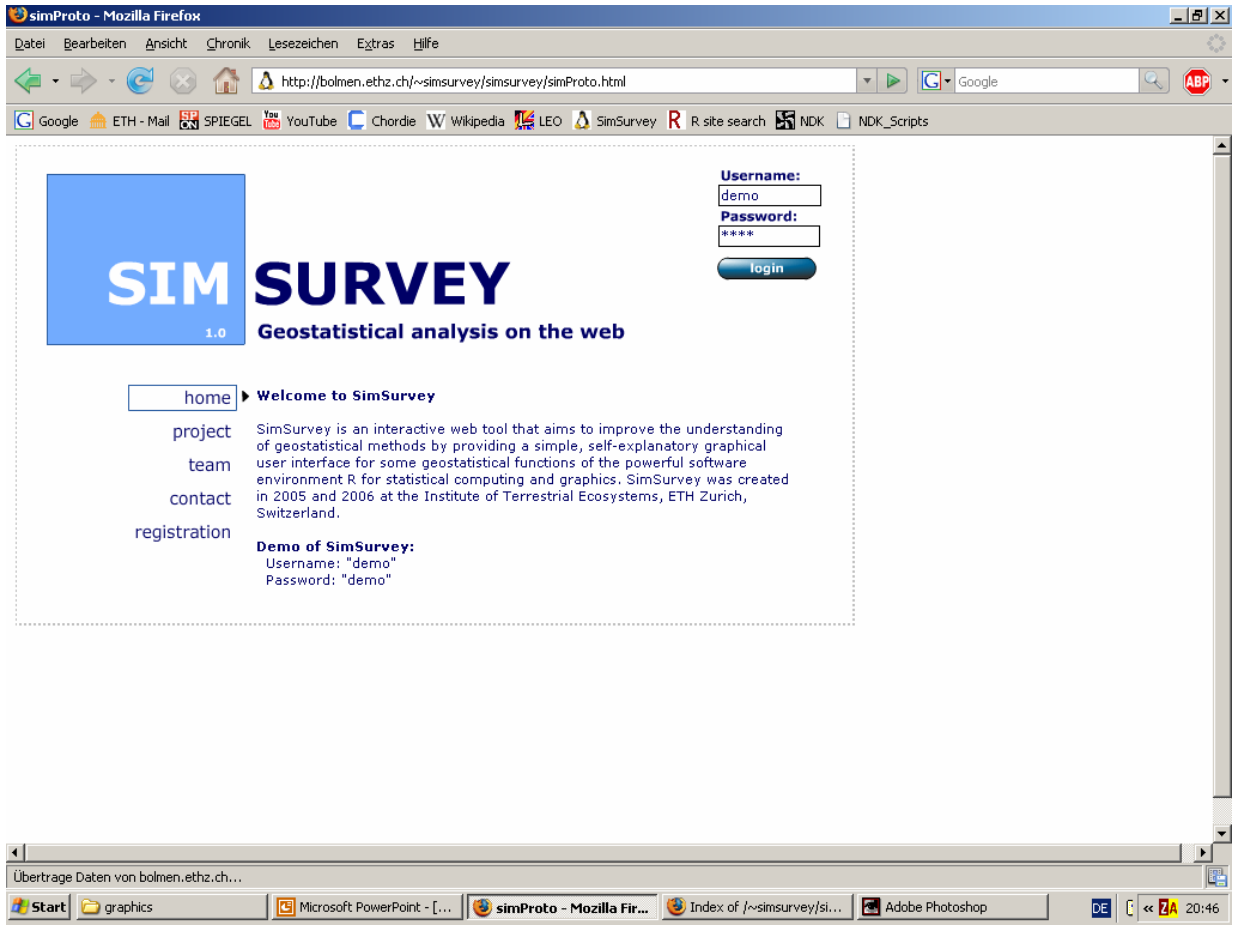


= often too difficult

2. What is *SimSurvey*?

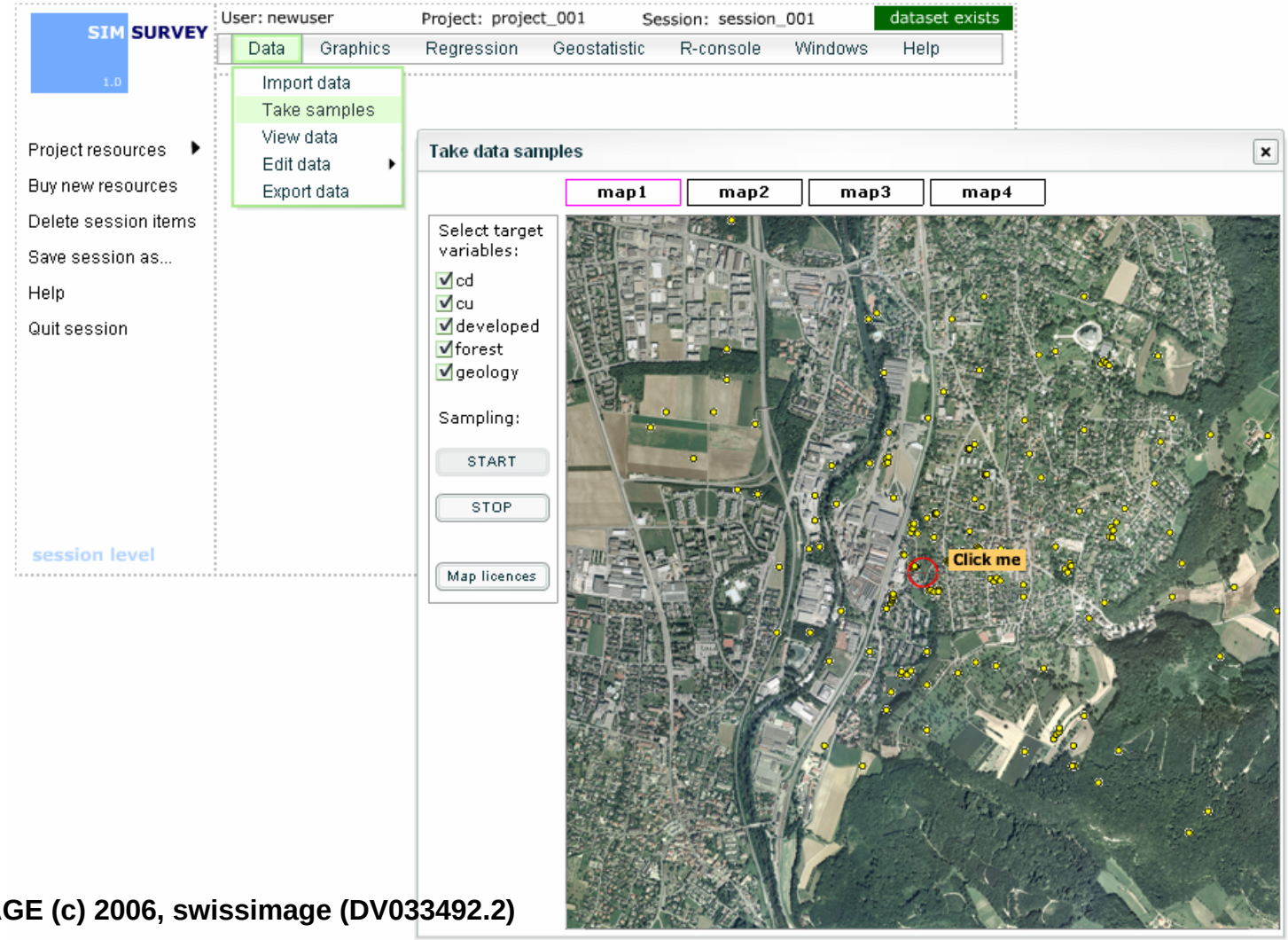
2. What is *SimSurvey*?

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.



Orthophoto : SWISSIMAGE (c) 2006, swissimage (DV033492.2)

3. Used Software

3. Used Software

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

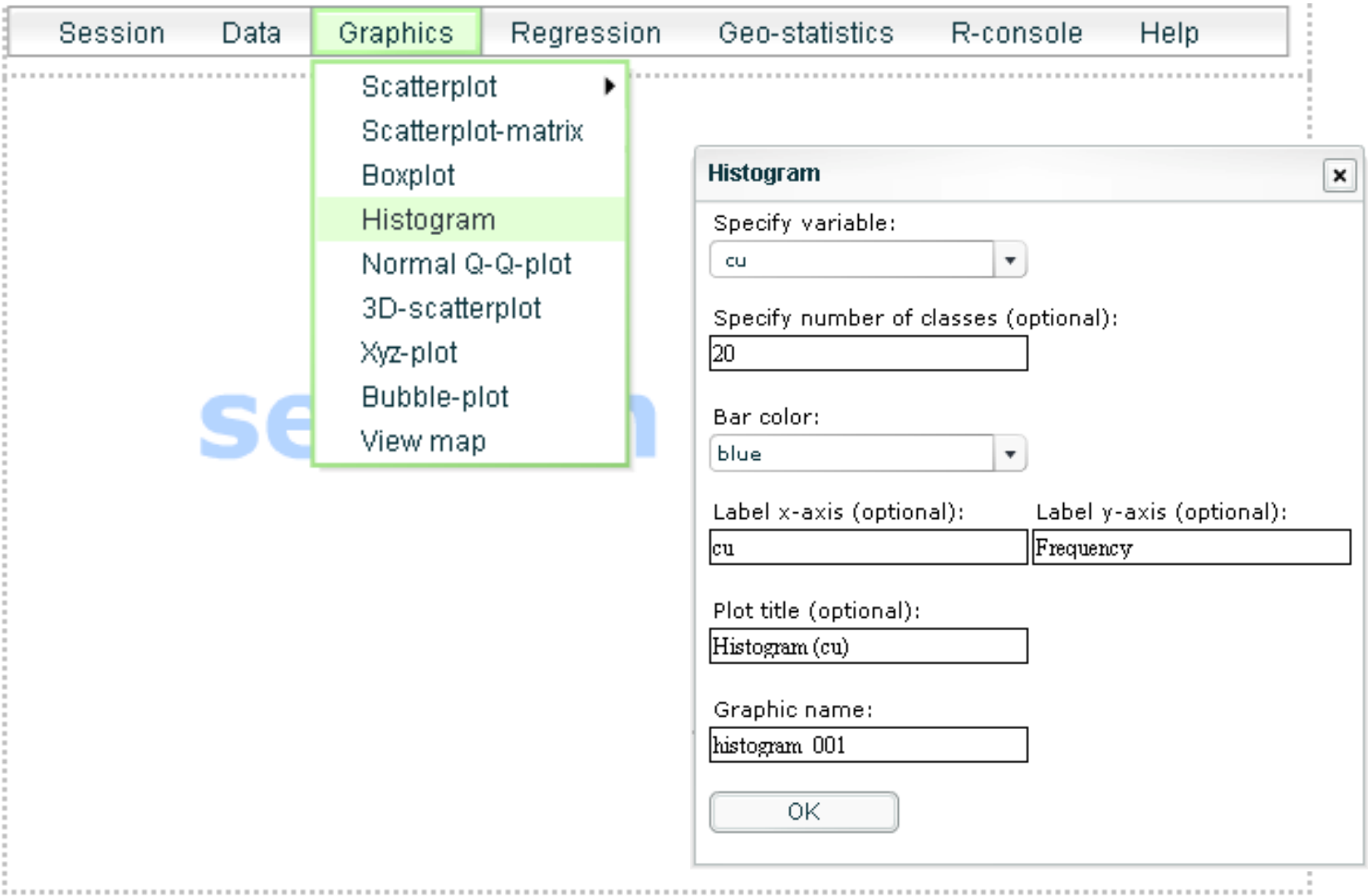
The following software is used to run SimSurvey:

- 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

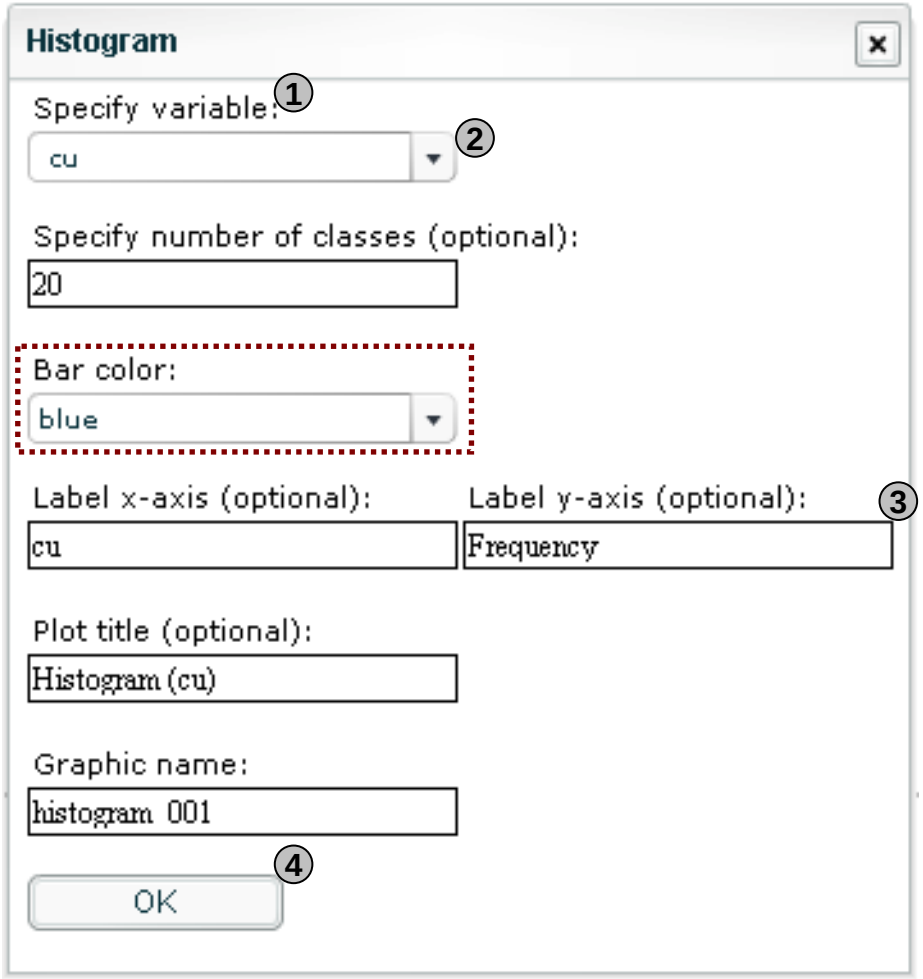
4. The GUI

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.

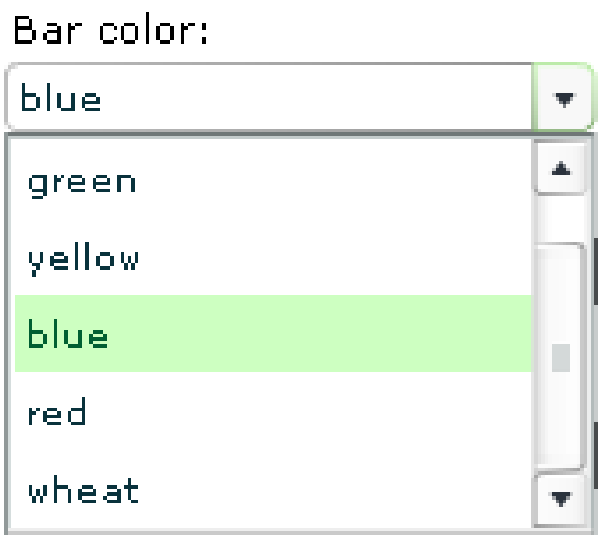


Dialog box items:

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.



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.

PHP-code

```
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;
}
```

$eigenschaften2=$_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.

R-code

```
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)
```

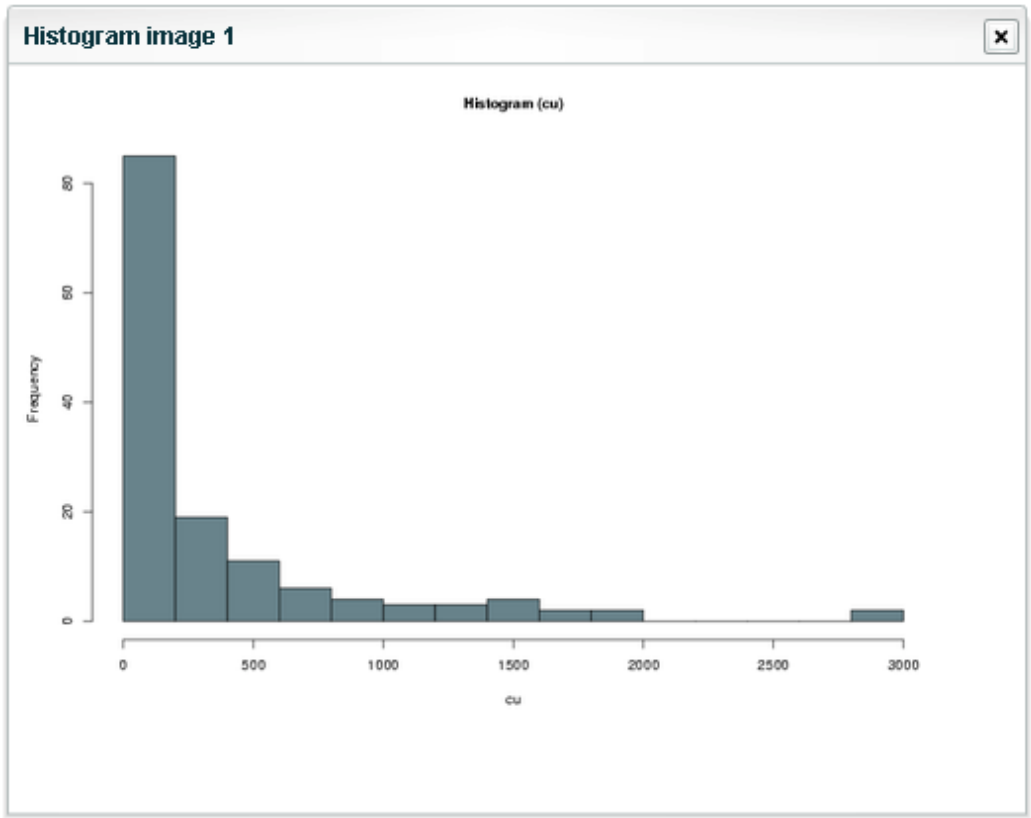
bar.color <- "blue"

col=bar.color

4. The GUI

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

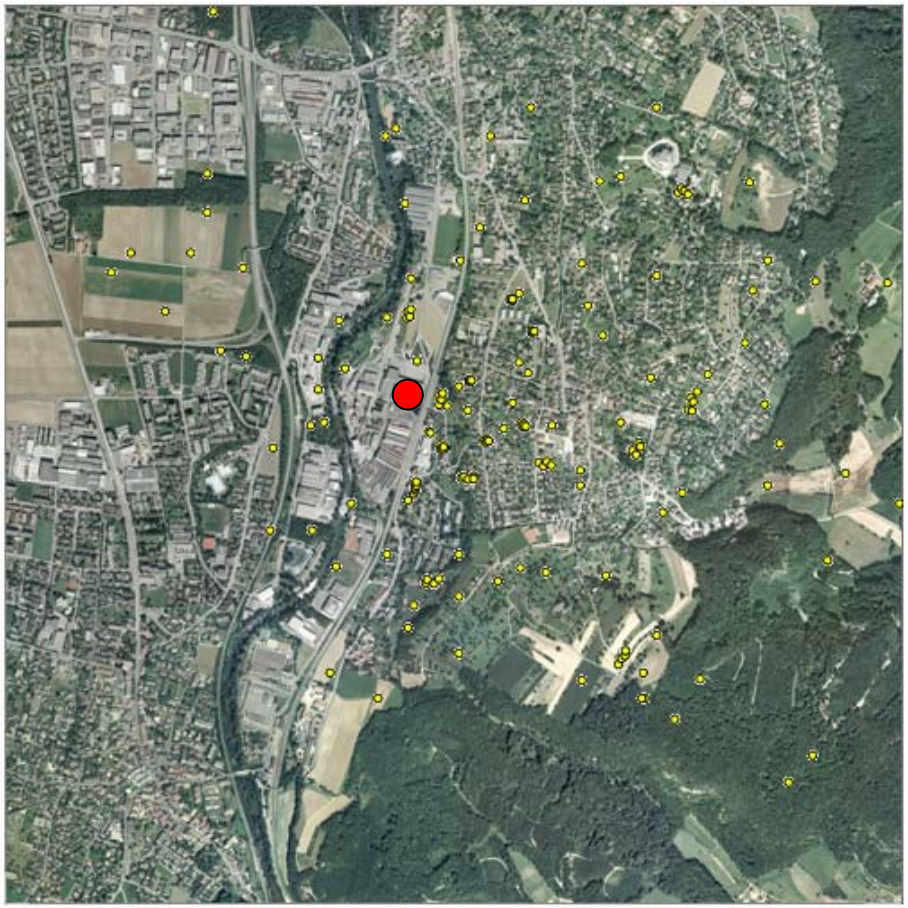
Session Data Graphics Regression Geo-statistics R-console Help



5. Example analysis

5. Example analysis

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

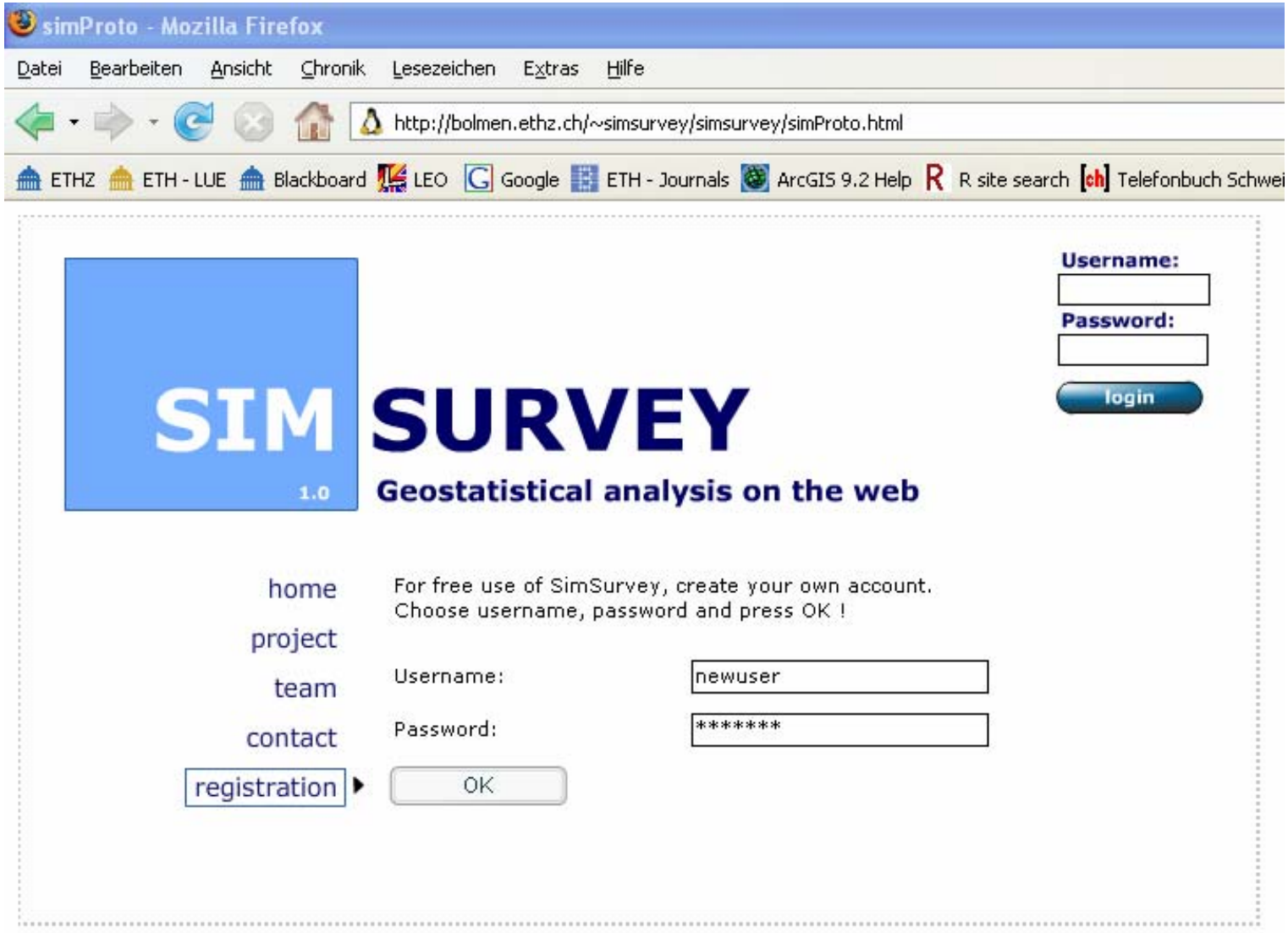


Objective: delineate those areas, where the copper content exceeds guide, trigger, and cleanup values of the ordinance.

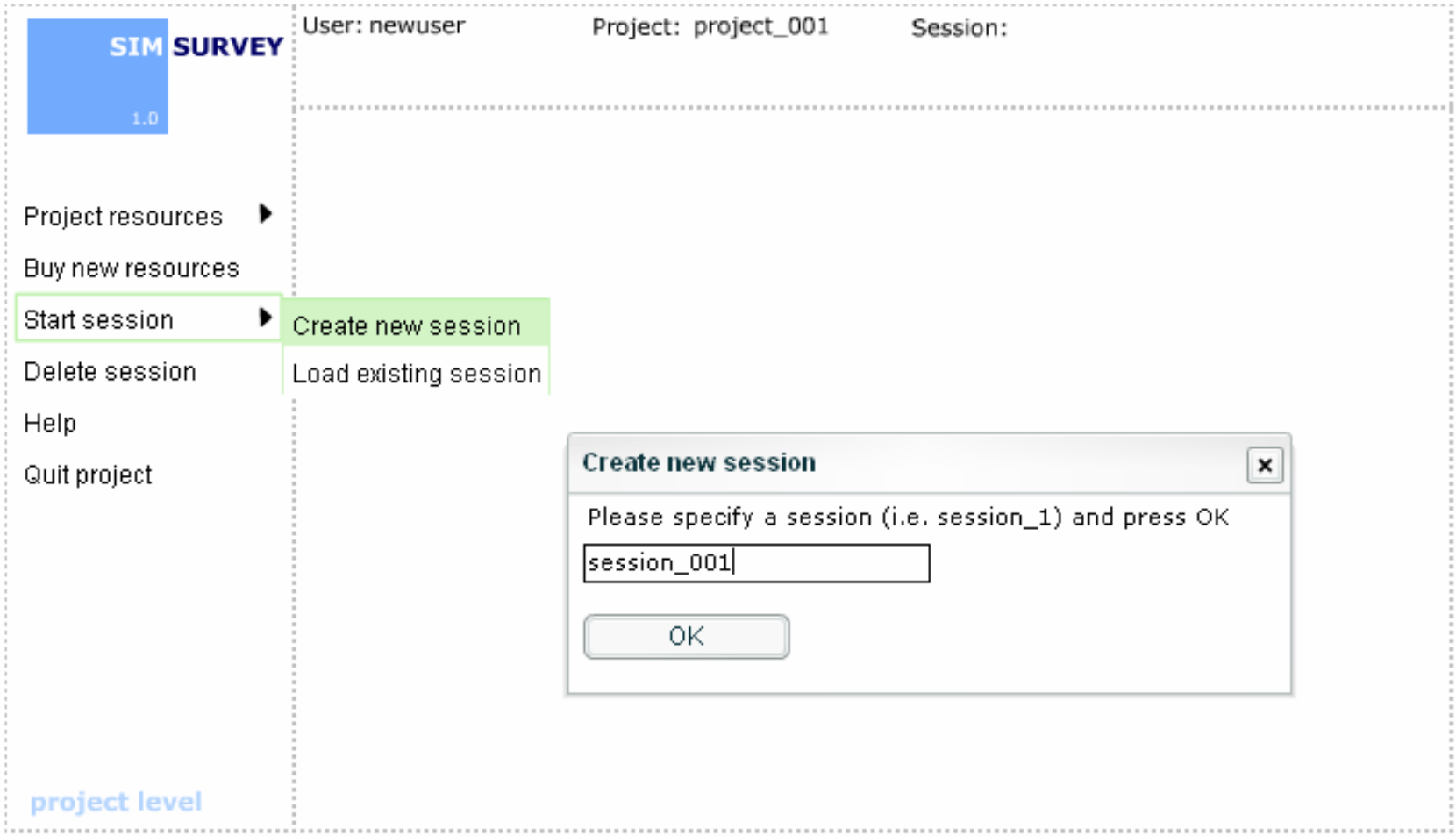
- location metal smelter
- topsoil samples (simulated data)

Orthophoto : SWISSIMAGE (c) 2006, swissimage (DV033492.2)

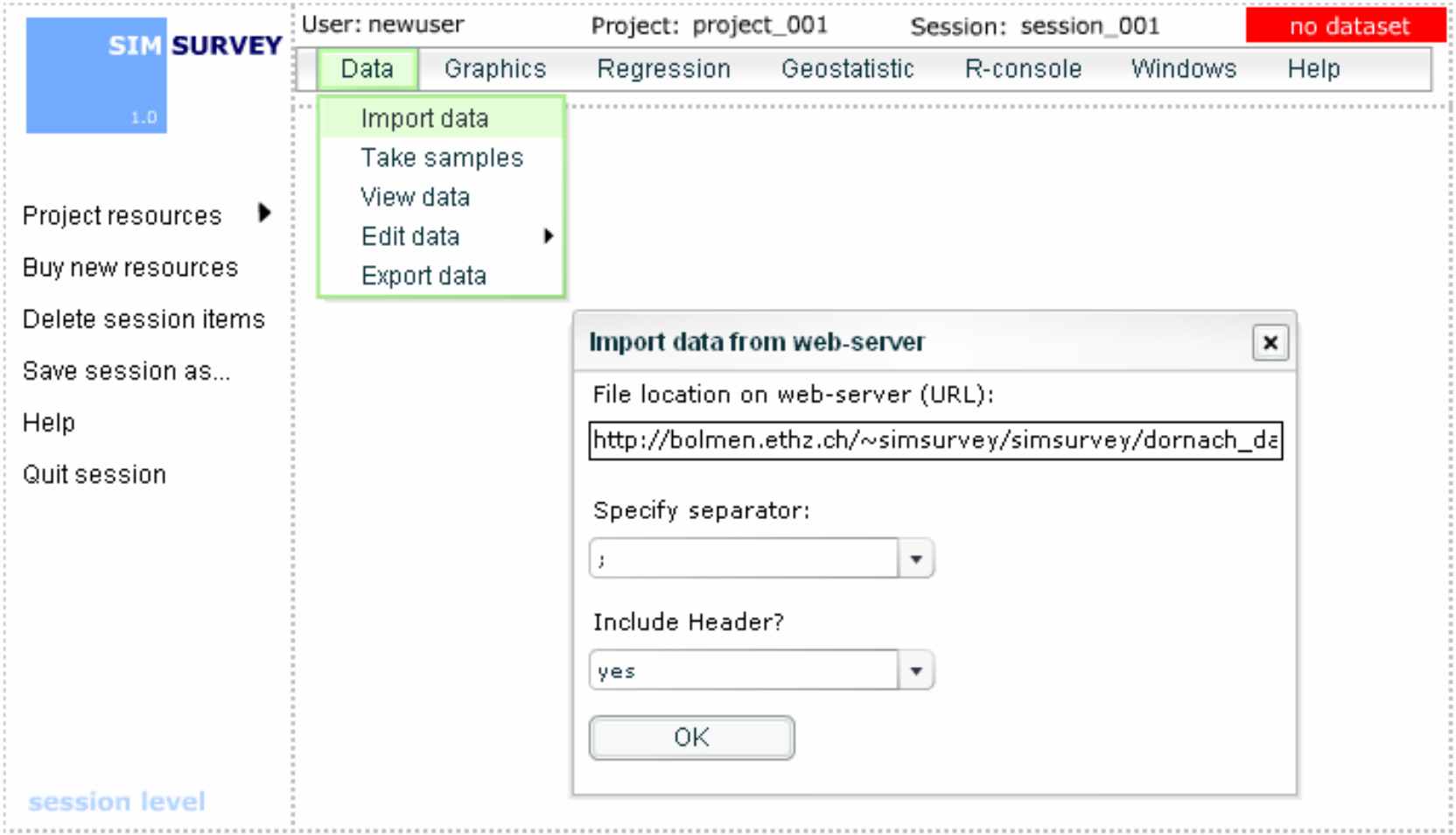
5. Example analysis: create your account & log in



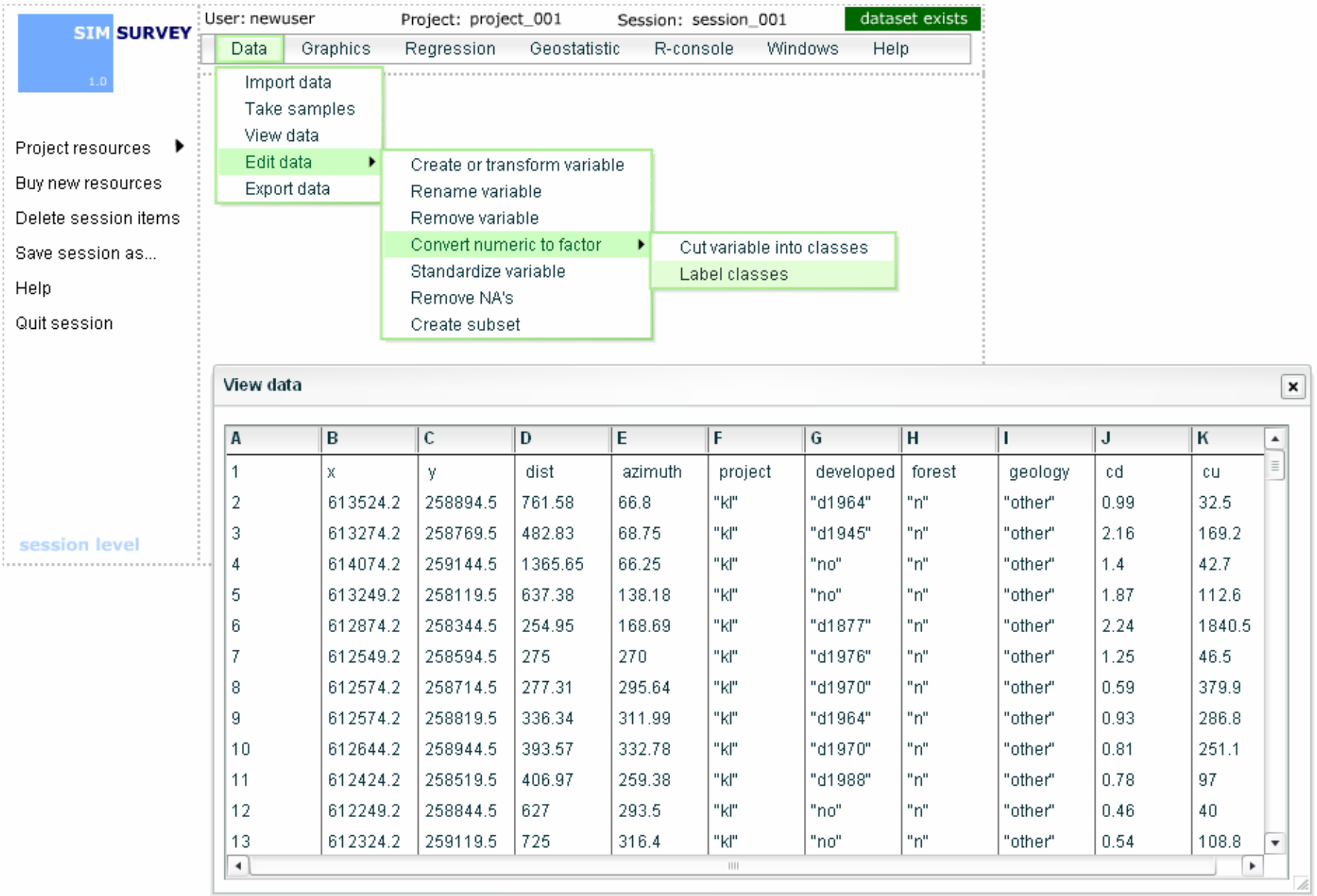
5. Example analysis: create new project and session



5. Example analysis: import soil sample data

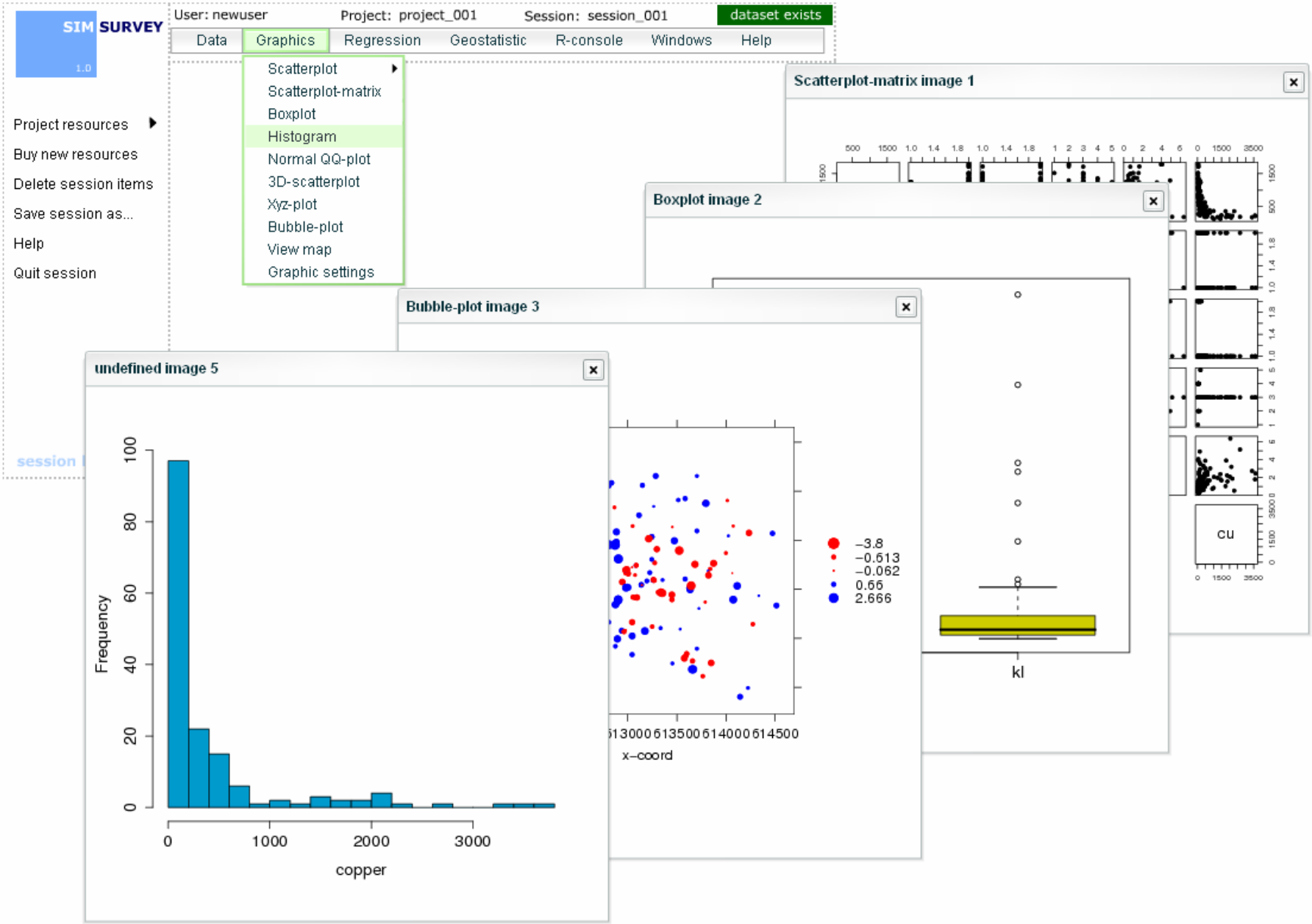


5. Example analysis: if necessary, manipulate data

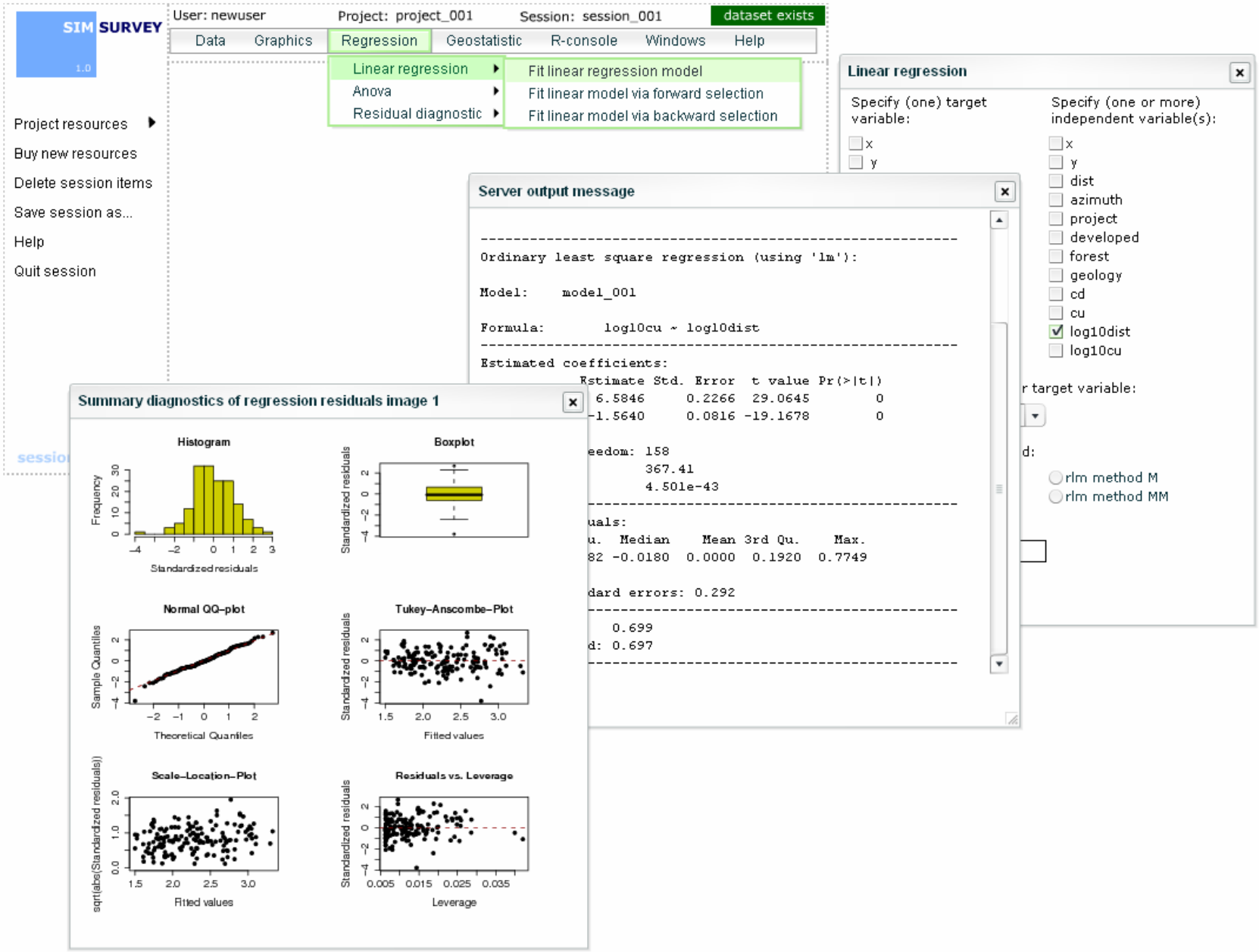


A	B	C	D	E	F	G	H	I	J	K
1	x	y	dist	azimuth	project	developed	forest	geology	cd	cu
2	613524.2	258894.5	761.58	66.8	"kl"	"d1964"	"n"	"other"	0.99	32.5
3	613274.2	258769.5	482.83	68.75	"kl"	"d1945"	"n"	"other"	2.16	169.2
4	614074.2	259144.5	1365.65	66.25	"kl"	"no"	"n"	"other"	1.4	42.7
5	613249.2	258119.5	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

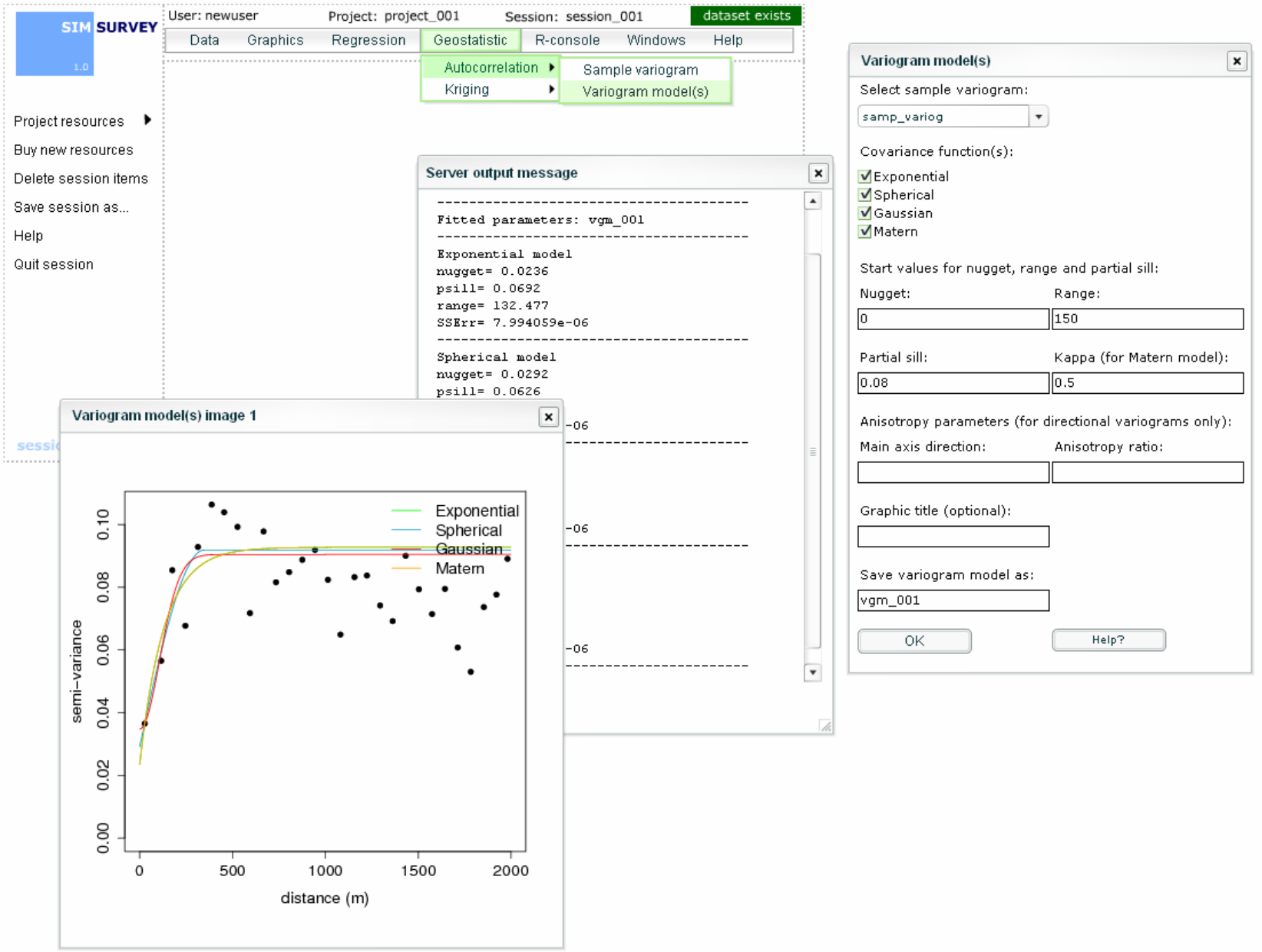
5. Example analysis: graphically examine data



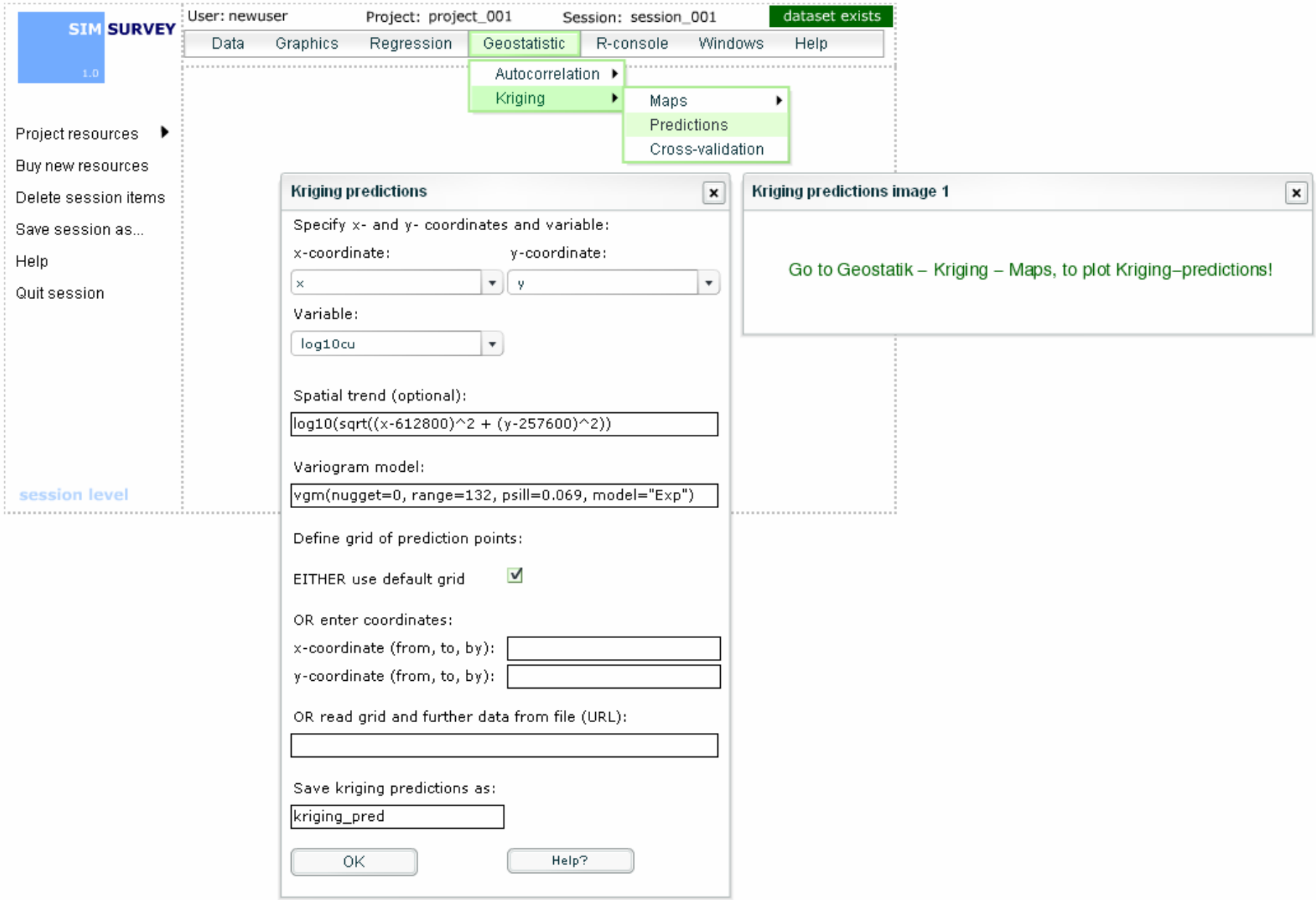
5. Example analysis : employ regression (lm, lqs, or rlm)



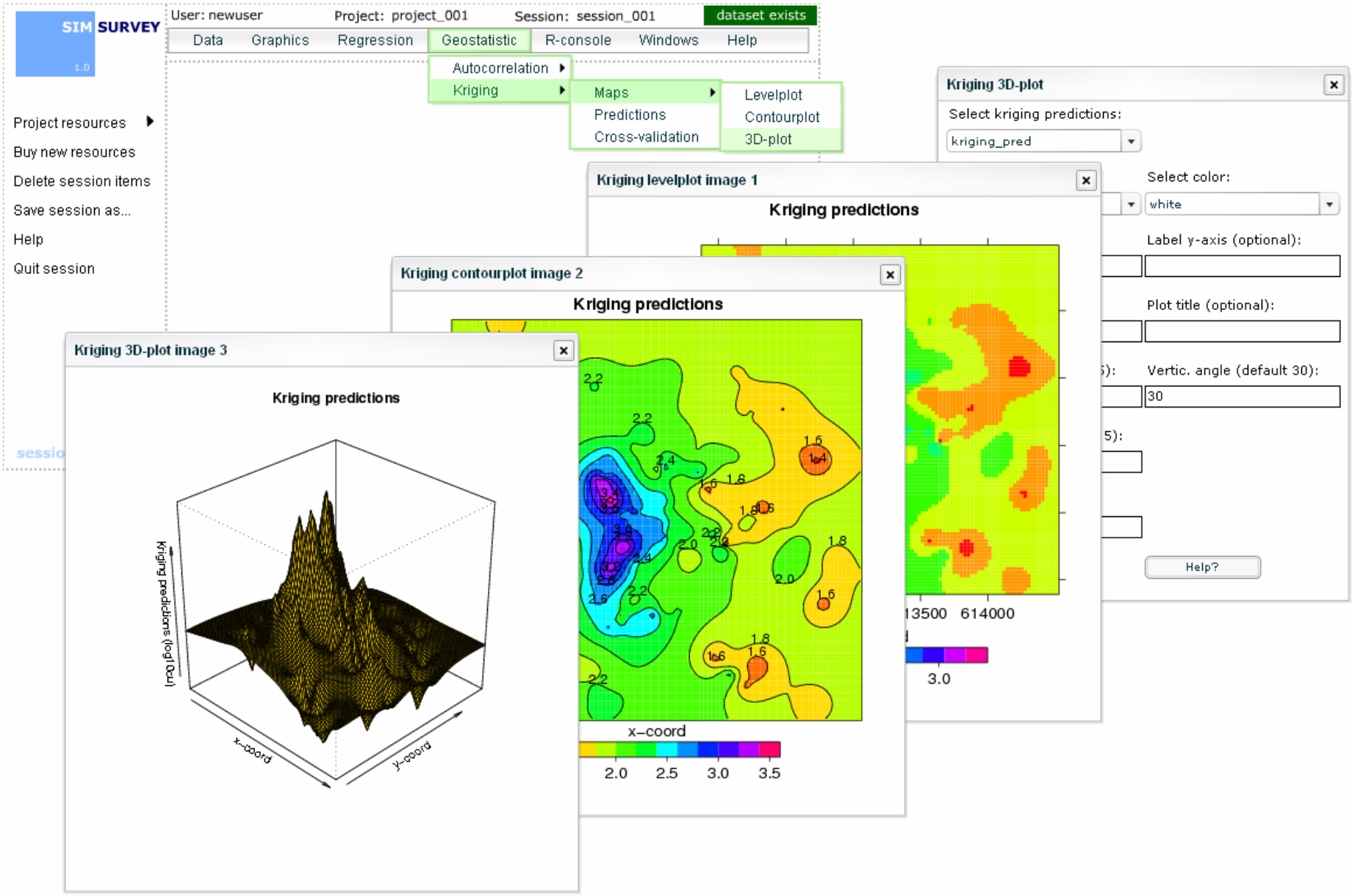
5. Example analysis: fit variogram model(s)



5. Example analysis: predict values at unobserved sites



5. Example analysis: visualize model predictions



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