SIMSURVEY— a tool for (geo-) statistical analyses with *R* on the web

Mario Gellrich *ETH Zurich, Institute of Terrestrial Ecosystems, gellrich@env.ethz.ch,*
Rudolf Gubler*,Terraplan Gubler, gubler@terraplan.ch,*
Andreas Schönborn*, armadillo media gmbH, schoenborn@armadillo-media.ch,*
Andreas Papritz *ETH Zurich, Institute of Terrestrial Ecosystems, papritz@env.ethz.ch*

Geostatistical methods are used in many branches of environmental research and applications for the statistical analyses of spatially referenced measurements and for the interpolation and mapping of data measured at a limited number of locations in a study domain. Courses in geostatistics are therefore part of the curriculum in environmental sciences and engineering at many universities. However, experience shows that geostatistics is a rather difficult subject to teach. Apart from the mostly limited prior knowledge in statistics, a lack of flexible, but at the same time easy-to-use software adds to the problems many students have with this topic. Commercially available statistics and GIS software either offers no or only limited geostatistical functionality, or it is expensive (and in addition often quite demanding to use). *R* includes several powerful packages for geostatistical analyses, but as a script-based programming language, *R* is difficult to use in introductory courses.

To mend this deficiency we developed *SIMSURVEY*, a graphical user interface (GUI) for geostatistical analyses with *R*. Unlike other *R* GUIs no software (apart from a browser) is required *as SIMSURVEY* runs on a web server (http://bolmen.ethz.ch/~simsurvey/simsurvey/simProto.html). Currently, *SIMSURVEY* offers the following functionality:

- Data transformation and management,
- exploratory analysis of spatial data,
- linear regression analysis of spatial data and analysis of variance,
- estimation and modelling of variograms, and
- universal kriging.

Various kinds of graphical tools are available for all these tasks. All these analyses can be run by using the GUI. For experienced *R* users, SIMSURVEY contains in addition a command window. For educational purposes, *SIMSURVEY* allows one to sample and to analyse simulated soil pollution data.

SIMSURVEY was implemented by an interplay of *Adobe Flash*, *PHP* and *R*. The GUI by which a user interacts with *R* is a *Flash* animation in a browser window. The dynamically changing structure of the GUI is largely controlled by *XML* code. The actions of a user are passed to *R* by *PHP*. Based on template *R* code *PHP* dynamically generates complete *R* scripts that are processed by *R* processes running permanently on the web server. To improve the performance *PHP* and *R* communicate with each other by a socket connection. The output that *R* generates (text and graphic files) are then routed back to the Flash animation by *PHP* and are then presented in the browser to the user.

Thanks to its modular architecture, *SIMSURVEY* can be easily modified and extended. To this aim the following steps are required:

- Define the new items of the GUI by adding to the *XML* code. To facilitate this task predefined elements for text input fields, radio buttons, check boxes etc. can be used.
- Write the template *R* code for the new tasks.
- Extend the *PHP-R* interface to pass the required information from the GUI to *R* (by dynamically generating *R* scripts) and route the *R* output back to the GUI.

This architecture provides a novel and flexible framework for general computations with *R* on a web server.

In our presentation we shall demonstrate the use of *SIMSURVEY,* and we shall show by an example how SIMSURVEY can be extended for new tasks.