Date:
Monday, June 28, 2010 - 09:15 to Friday, July 2, 2010 - 18:15

Body:

We arrived on Sunday to a small historic city in the hearth of the Iberian peninsula. It was hot night in Plasencia as usual, but it was an unusual sight for the restaurant staff at the terrace bar near the Campus of Plasencia who suddenly had to serve a large group of foreigners. ?Who are these people?? must have crossed the mind of the skilled waitress that had to serve us large quantities of ice cold beer and Mediterranean food (she managed to carry six glasses of beer at once ? just like at the Oktoberfest). The answer she could not know.

This group of foreigners were engineers, programmers, PhD students, researchers, scientists and university professors that came from all around the world to participate in an event called GEOSTAT ? a summer school for PhD students (and those who feel alike) that periodically travels around different universities of Europe. The 2010 edition of the GEOSTAT was held from 28.06. to 02.07. at the University of Extramadura in Plasencia. This event is organized by Tomislav Hengl in cooperation with Victor Olaya, the local host of the summer school. As the last four summer schools, GEOSTAT 2010 consisted of 5 days of lectures. The first day of course was common for all participants, while the 2nd, 3rd, and 4th days went to parallel sessions, and the 5th day was reserved for the workshop. My (unofficial) report describes in detail the three sessions that I followed, while the other three sessions I only describe from what I got from other colleagues.

For those of you who never heard of GEOSTAT, I should mention that GEOSTAT is a summer school that focuses on important aspects of statistical analysis of spatial and spatio-temporal data using open source spatial data analysis tools: primarily R and open source GIS (SAGA, GRASS). It is a knowledge-sharing event where the leading scientists and software developers gather to run theoretical and practical sessions. These sessions mainly focus on solving open problems in the field of analysis of spatial and spatio-temporal data, with a leitmotiv of ?spreading the word? about the recent developments in the open source community. If you are even more curious about how does GEOSTAT look in practice, you should continue reading this report to find out how I experienced the GEOSTAT 2010.
Spatio-temporal data analysis: from concepts to applications provided by Roger Bivand

I was first time introduced to the Bivand’s book (Applied Spatial Data Analysis with R [3]) by Tomislav Hengl, during a PhD course in Belgrade in December 2009. The ASDAR book was for me a new view on the (spatial data analysis) world. Before the ASDAR, I thought that standard GIS applications were only tools for spatial data manipulation. So I was sure that Roger’s session would be a great opportunity to expand the horizons. I wanted to catch every word in order to grasp a better understanding of how R and spatial packages in R work. At the end, Bivand’s session offered me much more that I was expecting. It was a perfect theoretical overview of GIS fundamentals, cartographer’s representation of phenomena, ontology and conceptualization. A perfect introduction for anybody that would like to rethink the ways to represent geographical features using mathematical/digital models.

The session continued with a detail explanation of spatial objects in R, coordinate reference systems, packages for spatial data manipulation, object exchange between packages etc. Bivand showed also many examples and R codes, and explained how to resolve same problem in different ways, and how to control GIS application like GRASS and SAGA from R.

The second part of session was about modeling spatio-temporal data, about spatial and temporal processes as separate features, and using models that consider spatial and temporal effects as combined. After a theoretical introduction, Bivand presented several examples of spatio-temporal data analysis with various objectives. Every example basically followed some research paper and data used in the research paper.

I was amazed by the devotion with which Bivand explains problems theoretically, and then suggests solutions e.g. through developing new spatio-temporal objects and functions in collaborative software. I was also impressed that he used examples from around 30 research papers, which I think is a great opportunity for young researchers because the right way to do good science is to read a lot of good articles.

At the end of the the first day we had enough time to chill out and refresh ourselves (again) with ice-cold beer and watch the World Cup Football games. Surprisingly, the spare time we spent together was comfortable as though we have been friends for long time (most of us at the summer school never met before).

Remote data access and processing. The SEXTANTE library provided by Victor Olaya

Olaya run a session on SEXTANTE [4], an independent Java library for advanced geographical analysis (it was largely based on SAGA GIS, but then with many new original contributions). SEXTANTE contains two main parts: a set of around 230 algorithms and a graphical component to run and use those algorithms in different software environments. He explained how to use SEXTANTE like extension in gvSIG GIS software in three different ways. The first as toolbox with graphical user interface similar like in standard GUI GIS application, the second through scripting BeanShell-based command-line interface, and the third was Batch processing interface. This is the first time I saw an open source GIS application that allows creation of complex models using drop and draw approach. In addition, the SEXTANTE
history session has many advantages in comparison with other GIS application that offer no scripting capabilities and history record. The SEXTANTE is available as an independent library which means that it can be attached as extension in many open source GIS applications written in Java. SEXTANTE is also implemented as extension of Web Processing Server 52N and is being implemented in the Geoserver.

After the SEXTANTE tutorial session, Olaya gave us training on how to create own geo-algorithms using components of SEXTANTE (toolbox, graphical modeler, etc.) in the Eclipse development environment. It was an extraordinary useful block for everyone who would like to develop own GIS application or extend an existing one.

Victor did not come through only as a good lecturer but also as a great host (organized a jamm session and many other events in Plasencia), and a fantastic musician (guitar, sax), which he demonstrated together with Edzer Pebesma (saxophone) on the day 3.

**Spatio-temporal modelling and prediction from sensor data provided by Edzer Pebesma**

Pebesma, the co-author of ASDAR book, and together with Bivand is one of the main maintainers and moderators of the the R-sig-GEO community, had a full day session on spatio-temporal statistics and geostatistics. This session was largely interactive discussion about methods for modeling of autocorrelation for spatial, temporal and spatio-temporal data, illustrated with numerous sensor-data exercises. The theoretical part focused on: detection and quantification of trends in data, spatial correlation, covariance, semivariance, temporal correlation only, spatio-temporal correlation and similar. Exercises were based on the two research papers for which R codes are available in the gstat package. It was some kind of re-engineering problem, we were investigating what was the meaning of the code, and how and why the authors made choice about various analysis steps.

The session finished with discussion about complexity of handling and modeling spatio-temporal data, memory requirements for solving space-time systems (problem of inversing space-time covariance matrix) and possible solutions. After the Pebesma’s session we had a gala dinner in hotel Parador, a gorgeous historic castle/monastery with unforgettable interior, delicious food and vine, served in Spanish style.

The rest of sessions I’m going to describe only superficially, because unfortunately I couldn’t attend. These report is based on the information I got from the participants that did attend these sessions and from the GEOSTAT 2010 web site materials.

**Geostatistical mapping using gstat, geoR and SAGA GIS by Tomislav Hengl**

This was a session by the person who basically initiated GEOSTAT (the first one he organized in 2004 in Zagreb, Croatia, then in Naples in 2007, Amsterdam in 2008, Split in 2009). Tom is the author of A Practical Guide to Geostatistical Mapping, which contains materials used over years for the GEOSTAT summer schools. This book is an open access publication, freely available in pdf format from his web page. The theoretical part of Tom’s session focused on basic concepts about feature/variables (environmental variables), digital formats of data, coordinate reference systems, spatial prediction methods in R and SAGA.
GIS. The second part, as usual in GEOSTAT sessions, was dedicated to running computer exercises in R, SAGA and Google Earth.

Olaf Conrad, the main creator and maintainer of SAGA [9] GIS, contributed demo sessions on geostatistics in SAGA GIS using the Meuse data (used in the gstat package): visualization of point and raster data, deterministic interpolators, variogram clouds and variogram surface options, regression analysis and KML export. This session was a continuation of the SAGA session from the first day of the summer school where he presented the SAGA software system architecture, modules and functionality. Olaf put some effort to actually prepare some real data for the Extremadura region and then demonstrated how various remote sensing and topographic layers can be analyzed and combined in SAGA GIS.

GRASS GIS tutorial provided by Markus Metz

Markus Metz gave a short GRASS GIS tutorial with exercises based on Spearfish and North Carolina sample datasets available from GRASS web site [10]. The tutorial contained: GRASS history and background; GRASS database concept (GISDBase, Location, Mapset); computational region in GRASS; raster format intro and principles of raster processing; vector format intro and principles of vector processing; data import/export; raster/vector reprojection. Exercises: data import/export, computational regions for raster processing, vector topological cleaning.

Spatial uncertainty propagation provided by Gerard Heuvelink

This session focused on theory and practice of spatial uncertainty propagation analysis. It was given by the one of the founders of this theory and probably one of the most influential GIS scientists of the era. The session started with a theoretical introduction to various uncertainty propagation techniques. As with all other blocks, there were also several exercises and case studies with available R codes. The impression of participants of Heuvelink’s session was that he was extraordinary good lecturer and moderator that can explain complex theory in very accessible and easy way.

Virgilio Gomez-Rubio, the 3rd coauthor of the ASDAR book, joined us the last day of the workshop with a lecture on parallel computing in R. Virgilio included code examples with universal block kriging and the detection of disease clusters. Finally, he described how a full parallel algorithm can be designed and implemented for Bayesian inference modeling using spatial data.

Workshop was the last part of the course where participants presented their work and asked specific questions. The questions were truly diverse. This session was also useful for people that had no question (such as myself) but got new ideas for research. Possible solution of problems were considered and discussed during the workshop day, and the day after in friendly atmosphere during the field trip to the Montafriage National park.

Having finished this summer school, I now fully understand the importance of being a part of the open source community. GEOSTAT and R-sig-geo are extremely useful communities if you come from countries that have limited resources and funds for young researchers. By
getting some initial guidance from such top scientists and developers I feel confident that I will also be able to contribute good publication and software in the coming years. I can only advise you to join the R-sig-geo mailing list to get a good idea about what do I mean by ?getting guidance from really smart and open minded people?. Which they kindly do without any formal commitment, as they did for the GEOSTAT school.

At the end, GEOSTAT 2010 was a great pleasure. I noticed that there are several people that have been coming back to GEOSTAT, so I have a good feeling that I might also become one.

In Belgrade, 15th July 2010  Milan Kilibarda

Preview: