SAGA vs GRASS: A Comparative Analysis of the Two Open Source Desktop GIS for the Automated Analysis of Elevation Data

T. Hengl[1], C. H. Grohmann[2], R. S. Bivand[3], O. Conrad[4], and A. Lobo[5]

[1]IBED University of Amsterdam, UvA B2.34, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands
Telephone: +31-(0)20-525-7379
Email: T.Hengl@uva.nl

[2]Institute of Geosciences, University of São Paulo, Rua do Lago, 562, Cidade Universitária 05508-080 São Paulo – SP, Brazil
Email: guano@usp.br

[3]Norwegian School of Economics and Business Administration, Helleveien 30, 5045 Bergen, Norway
Email: Roger.Bivand@nhh.no

[4]University of Hamburg, Institute for Geography, Bundesstraße 55, Raum 810, 20146 Hamburg, Germany
Email: conrad@geowiss.uni-hamburg.de

[5]Institut de Ciències de la Terra "Jaume Almera" (CSIC), Solé Sabarís s/n E–08028 Barcelona, Spain
Email: Agustin.Lobo@ija.csic.es

1. Introduction

Two of the most used open source desktop GIS software for the analysis of DEMs are SAGA[1] (System for Automated Geoscientific Analyses) GIS and GRASS[2] (Geographic Resources Analysis Support System) GIS (Wood, 2008; Steiniger and Bocher, 2009). SAGA has been under development since 2001 at the University of Göttingen (the SAGA development team, has since moved to University of Hamburg), Germany, with aim of simplifying the implementation of new algorithms for spatial data analysis. In 2004, most of SAGA's source code was published using an Open Source Software license. The functionality of SAGA is described in Böhner et al. (2002) and Böhner et al. (2008); the software design, methods, and usage are explained in detail in Conrad (2007).

GRASS GIS, now one of the eight initial Software Projects of the Open Source Geospatial Foundation (OSGeo), is probably the most known open source GIS software in the world. Its functionality and usage are described in detail in Neteler and Mitasova (2008). GRASS itself is a collection of modules (they vary from version to version). Although originally a Linux-based project, the most recent version of GRASS (6.3; development version) is now also available for MS Windows machines.

GRASS is a much larger project than SAGA considering the number of developers/institutions involved, although their functionality considering the DEM analysis is about similar. Both SAGA and GRASS are increasingly rich considering the functionality they offer: the latest version of SAGA (2.0.3.) contains 48 libraries with 300 modules; GRASS 6.3 contains over 350 routines. Both in fact provide more functionality for the analysis of DEMs than proprietary low-end products such as the basic installation of ArcGIS 9.2. By linking SAGA/GRASS with R environment for statistical computing, a powerful combination is created that allows fusion of GIS and statistical functionality in the same code (Grohmann, 2004; Brenning, 2008).

---

[1] http://www.saga-gis.org
[2] http://grass.osgeo.org

In this article we present the results of a comparative analysis of performances of the two GIS software for the analysis of elevation data. We focus on DEM generation, extraction of hydrological features (stream networks), and extraction of gridded DEM derivatives. We will base our comparison on objective and subjective criteria: measures of accuracy, processing speed, but then also on the user's satisfaction following questionnaires. Our intention is not really to name the winner, but to see what the basic differences are, and to suggest ways to combine the strengths of the two packages.

2. Methods and Materials

In order to make this software comparison objective, we will use reproducible methods and pre-defined technical criteria of principal interest to the users: accuracy, speed, ease of use. Guides and defined criteria to compare GIS software do not really exist. In order to be able to compare two packages, they need to at least overlap considering the functionality they offer. In fact, in order for a comparison to be fair, the software packages should implement the same mathematical models; if this is not the case, they should at least indicate the same type of service: e.g. DEM generation, extraction of hydrological network etc. In the case of SAGA/GRASS, a significant overlap in functionality exists. For each case study we run analysis using sampled elevation data, and then validate the outputs generated using SAGA and GRASS versus the ground truth data. The case study and the processing steps shown in this article are available from the http://geomorphometry.org website (R script).

2.1 Case Studies

We use three standard elevation datasets common for contemporary geomorphometry applications: point-sampled elevations (LiDAR), contours lines digitized from a topo map, and a raster of elevations sampled using a remote sensing system. All three datasets (lidar.shp, contours.shp and DEMSRTM1.shp) refer to the same geographical area — a 1×2 km case study *fishcamp* located in the eastern part of California.

The point-dataset (LiDAR ground reflections) consists of 273,028 densely sampled points. The original LiDAR dataset consist in fact of over 5 million of points, which were sub-sampled to speed up the processing. A very fine resolution (2.5 m) DEM derived from LiDAR measurements was used as a ground truth layer for validation purposes. For the raster-dataset we use the SRTM 1 arcsec (25 m) Shuttle Radar Topography Mission (SRTM) DEM. The complete dataset was obtained from the USGS National Map seamless server[3].

2.2 Comparative Criteria

We have decided to base our comparison on four criteria: (1) absolute accuracy of the DEMs (surface) generated using default settings; (2) spatial accuracy of hydrological features extracted using default settings; (3) processing speed; (4) extendibility i.e. ease of implementation of new algorithms (from mathematical model to a new routine).

The accuracy of generated elevations was assessed versus the most accurate DEM available for the study area of interest (LiDAR-based DEM) using standard accuracy measures (RMSE, MSE). To assess the spatial accuracy of the derived hydrological (stream) networks we used the mean distance from the point line sets that can be derived by overlaying the predicted stream network over the buffer map generated using

[3] http://seamless.usgs.gov

the actual stream network. Processing speed was measured using the "system.time" method in R.

3. Results

3.1 DEM Generation Accuracy

We have generated DEM surfaces from contours.shp using spline interpolation, which has been recommended by both SAGA and GRASS developers as the most suited DEM gridding technique for contour data (Conrad, 2007; Neteler and Mitasova, 2008). This looks for closest 10 points in a local search radius and fits the Thin Plate Spline over a 25 m grid. This initial DEM can be hydrologically adjusted using the deepen drainage route. The resulting DEM surface can be seen in Fig. 1.

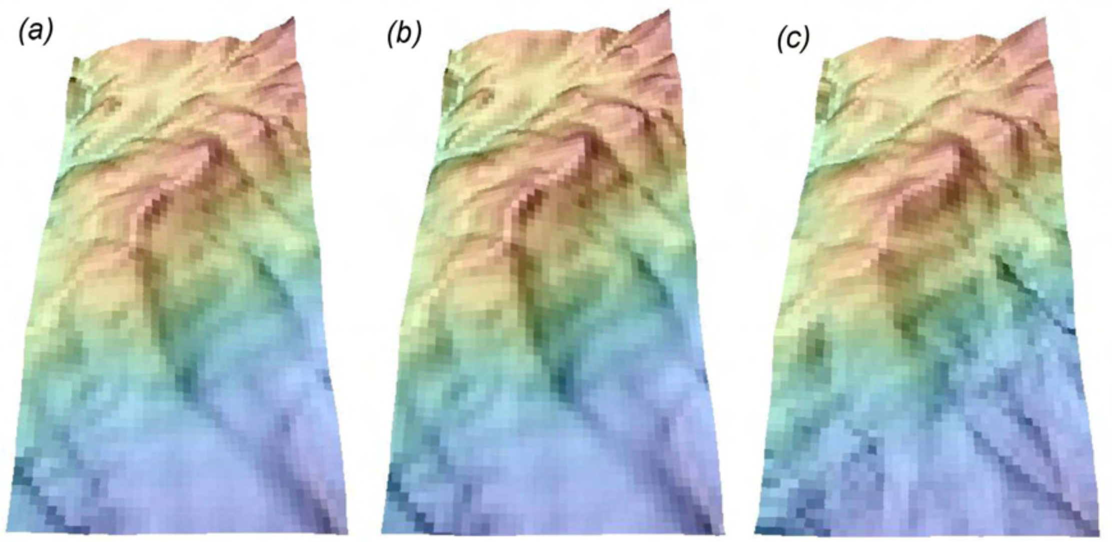

Figure 1. Comparison of DEMs derived in (a) SAGA and (b) GRASS using the 1:24k topo-map contour lines (contours.shp), as compared to the (c) DEM derived using all LiDAR points. Perspective view on the study area from the West–East direction.

The final comparison of the two DEMs shows that both software generate DEMs of approximately equal quality: the RMSE for the SAGA-derived DEM is 5.31 m, as compared to the 5.28 m for GRASS. Note also that, because the study area has a distinct topography, and because we use a smooth interpolator, both maps do not show artefacts (Fig. 1). Although the RMSE of both maps is satisfactory, it is obvious that many hydrological features in the area were missed. Neither SAGA nor GRASS are able to incorporate information on existing hydrological features (streams, and water bodies) into the generation of DEMs; compare with the ANUDEM procedure implemented in the TOPOGRID function of ArcInfo (Hutchinson, 1989).

3.2 Spatial Accuracy of Extracted Hydrological Networks

In the next exercise, we compare the drainage networks derived in SAGA and GRASS versus the stream network digitized from the topo-map. For this comparison, we use the 1 arcsec SRTM DEM dataset (DEMSRTM1.asc). In the case of SAGA, the stream network can be extracted in few steps: first we filter the spurious sinks, then extract channel network as shape file. In the case of GRASS, we first need to read the DEM into the GRASS format, and then extract the watershed using the "r.watershed" function. In the next step, we can "thin" the generated streams, so that we can also convert the stream map to vector lines.

Figure 2. Hydrological networks extracted using the 1 arcsec SRTM DEM (DEMSRTM1.asc) dataset in SAGA GIS and GRASS GIS, as compared to stream network digitized from the 1:24k topo-map (grey bold lines).

The final comparison is shown in Fig. 2. The statistical comparison of the difference between how closely the predicted streams match the streams at the topo-map shows that GRASS-derived streams are in average (median) 35.4 m from the actual stream network; for the SAGA-derived network we get 90.1 m. This difference is statistically significant (t-test statistics p-value=1.36e-06).

3.3 Processing Speed

To compare the processing speed, we tested several processing steps using the relatively large LiDAR dataset (lidar.shp). First, we use it to generate DEM using spline interpolation. This shows that there are quite some differences in the computing times: SAGA takes 279 seconds to generate a DEM (800×400 grid nodes) using Thin Plate Splines (closest 10 points) using 273,028 LiDAR points; GRASS takes about 2-3 times more. Even to derive DEMs from contour lines (2555 points) takes 81 seconds in GRASS, as compared to <1 seconds in SAGA. Second, we 480 test generation of DEM derivatives for the 800×400 grid DEM. SAGA takes 0.6 seconds to derive a slope map vs 0.4 seconds in GRASS.

There seems to be not much difference considering the image processing modules. Generation of the TWI using the multiple flow algorithm takes: 6 seconds with SAGA vs 31 seconds with GRASS. To derive the total incoming solar radiation for one year, with a hourly step of 2 hours, and a daily step of 5 days, takes 170 seconds in SAGA and 494 seconds in GRASS.

In summary, SAGA is in average about 2–3 times computationally more efficient for generation of DEMs using splines, and for hydrological and solar irradiation modeling. Functions in GRASS, on the other hand, in general provide more possibilities — they allow you to adjust the parameters and/or combine two operations within a single command line. The solar irradiation modelling in GRASS is much more sophisticated than in SAGA: it includes shadowing effects, reflected and diffuse radiation etc. GRASS also prints out the progress of processing in percentage.

3.4 Extendability

SAGA makes it easy to implement new algorithms and plug-in new libraries using Python; GRASS has an extensive and well-documented support for scripting and extension of routines. The "g.parser" module will parse specific variables in scripts (e.g. Shell, Python, Perl) and provide auto-generated graphical user interface, help page template and command line options checking, easily making simple scripts in full-featured GRASS modules. Some of the modules shipped with GRASS are in fact scripts, like "r.shaded.relief", which redirects the user options as parameters to the raster algebra module, "r.mapcalc". The GRASS-wiki AddOns lists over 100 users contributions.

In summary, GRASS supports scripting (original syntax), AddOns modules, and has a larger and more active community than SAGA. Nevertheless, modules in SAGA can also be easily extended or built from scratch but programming skills are required (C++, Python).

4. Discussion and Conclusions

The results of this set of comparisons shows that there are indeed some differences between the two software: SAGA seems to be more computationally efficient (2–3 times faster), GRASS generates more accurate streams networks, and in general offers more sophistication considering analysis of elevation data. On the other hand, much of the functionality (DEM generation, image processing, vector/raster conversion) in the two software is comparable.

The real differences exist between SAGA and GRASS in controlling the process from R and extending the functionality. SAGA seems to be slightly more user friendly considering the possibilities to manipulate maps, zoom in into the data and control the processing from external applications. GRASS, on the other hand is powerful as a tool for processing, but its interactive display characteristics are limited. SAGA's user community is smaller, less international than in the case of GRASS. SAGA is also missing (completely!) help documents that explain different functions, how to set-up different parameters etc. In summary, there are quite big "view of the world" differences between GRASS and SAGA.

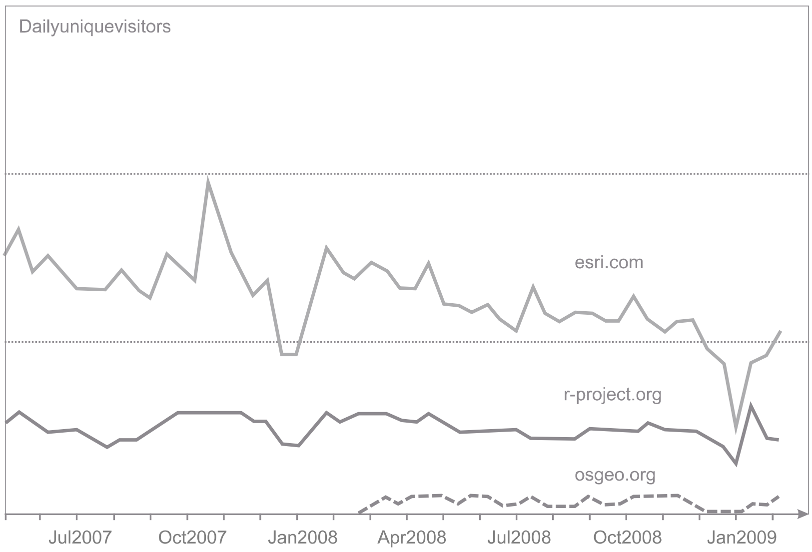

Figure 3. Trends in the web-traffic for http://esri.com, http://r-project.org and http://osgeo.org (following the http://trends.google.com statistics). Note that R users are typically more active also during the holiday periods.

Since the 1980's, the GIS research community has been primarily influenced by the (commercial) software licence practices that limited sharing of ideas and user-controlled development of new functionality (Steiniger and Bocher, 2009). With the initiation of the Open Source Geospatial Foundation (OSGeo), a new area started: the development and use of open source GIS has experienced a boost over the last few years; the enthusiasm to share code, experiences, and to collaborate on projects is growing. By comparing the web-traffic for commercial and open source GIS (Fig. 3), we can notice that there is indeed something going on — are we close to an inflection point when the open source GIS community will exceed e.g. the ESRI user's community? The open source GIS is certainly more powerful, more professional and more vital than five or six years ago. Further integration of packages such as SAGA and GRASS would increase this impression even more.

References

Bivand, R, 2005, Interfacing GRASS 6 and R. Status and development directions, *GRASS Newsletter*, 3, 11–16.

Böhner, J, McCloy, K, and Strobl, J, (eds) 2002, *SAGA — Analysis and Modelling Applications*, vol. 115 of Göttinger Geographische Abhandlungen, Verlag Erich Goltze GmbH, Göttingen.

Böhner, J, Blaschke, T, and Montanarella, L, (eds) 2008, *SAGA — Seconds Out*, vol. 19, Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie, Hamburg.

Brenning, A, 2008, Statistical Geocomputing combining R and SAGA: The Example of Landslide susceptibility Analysis with generalized additive Models, In: Böhner, J, Blaschke, T, and Montanarella, L., (eds) *SAGA — Seconds Out*, vol. 19, Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie, Hamburg, pp. 23–32.

Conrad, O, 2007, *SAGA — Entwurf, Funktionsumfang und Anwendung eines Systems für Automatisierte Geowissenschaftliche Analysen*. Ph.D. thesis, University of Göttingen, Göttingen.

Grohmann, C, 2004, Morphometric analysis in Geographic Information Systems: applications of free software GRASS and R, *Computers & Geosciences*, 30, 1055–1067.

Hutchinson, M, 1989, A new procedure for gridding elevation and stream line data with automatic removal of spurious pits, *Journal of Hydrology*, 106, 211–232.

Neteler, M and Mitasova, H, 2008, *Open Source GIS: A GRASS GIS Approach*, Springer, New York, 3rd edn.

Steiniger, S and Bocher, E, 2009, An Overview on Current Free and Open Source Desktop GIS Developments, *International Journal of Geographical Information Science*, in press, NA.

Wood, J, 2008, Overview of Software Packages Used in Geomorphometry, In: Hengl, T and Reuter, H, (eds), *Geomorphometry: Concepts, Software, Applications*, vol. 33 of Developments in Soil Science, Elsevier, pp. 257–267.