Environmental Data Management, Analysis and Modeling in GRASS6

Helena Mitasova
Department of Marine, Earth and Atmospheric Sciences
North Carolina State University

Markus Neteler
Center for Scientific and Technological Research

Lubos Mitas
Department of Physics, North Carolina State University
Russell S. Harmon, US Army Research Office

Abstract. Geographic Resources Analysis Support System (GRASS, http://grass.itc.it) has evolved into one of the most comprehensive, general purpose open source geoinformation systems since its original design as a land management software tool for military installations. Support for environmental applications has been an integral part of its 20+ years of development. Recent GRASS6 releases reflect new efforts of the international development team to bring GRASS closer to a modern geospatial software system and make it one of the core components of the open source geospatial software stack. Overview of the GRASS6 capabilities relevant for environmental applications are provided, including the updated and new tools for 2D and 3D raster data, the vector modules based on the redesigned topological 2D/3D vector engine, and SQL-based attribute management. Approaches to linking with other open source geospatial tools and environmental models are discussed. Case studies from North Carolina and Panama are used to illustrate the GRASS6 capabilities for environmental applications. The examples include watershed analysis, modeling impacts of terrain modifications on water flow and assessment of effectiveness of various approaches to runoff and sediment control. Tools for processing and analysis of multi-temporal lidar-based elevation models and their environmental applications are also presented. To facilitate better understanding of impacts of terrain change and improve face-to-face collaboration the possibilities to use Tangible GIS that integrates digital landscape representation with a physical tangible model are explored.

1. Introduction

Geographic Resources Analysis Support System (GRASS, http://grass.itc.it) is the oldest and one of the largest geospatial software systems. It was originally designed for environmental applications at military installations with special focus on land use management. Recent GRASS6 releases reflect the efforts of the international development team to modernize GRASS and make it one of the core components of the open source geospatial software stack.

GRASS development is coordinated at ITC-irst, Italy and distributed through the server at http://grass.itc.it and at more than 20 international mirror sites. Worldwide downloads are estimated at more than 25,000 per month. Quality management is enforced by transparent development methods such as a centralized source code repository (in a CVS server), immediate peer-review of source code changes by email broadcast, and an automated software quality assurance system (Bouktif et al. 2006). The GRASS software itself is organized in modules that can be easily combined to create powerful applications depending on the user's needs. The code and data format is portable across all major operating systems including GNU/Linux (GRASS is now official part of several Linux distributions such as Mandrake or Fedora), Mac OSX, MS-Windows (native and with Cygwin), and across hardware platforms including 64bit.

Environmental applications often require integration of georeferenced data from multiple sources in different formats and coordinate systems, extensive data exchange capabilities are therefore essential. GRASS uses the GDAL/OGR library (http://www.gdal.org) that supports import, export and manipulations of a wide range of raster and vector formats and the PROJ library to support coordinate system transformations. GRASS offers several ways for user interaction, from command line and internal graphical user interface to external tools such as QGIS (http://www.qgis.org). Every GRASS module now generates its graphical user interface (GUI) on the fly. This keeps descriptions always up to date and permits for internationalization of the system. The parameters/flags explanation section of the manual pages is also rendered from the standard help texts with additional descriptions including examples and algorithm references.

Although GRASS has evolved into a general purpose GIS, its major areas of application remain various aspects of environmental modeling and analysis. In the following sections we provide a brief overview of GRASS functionality, then we introduce the first OSGeo Educational data set (Open Source Geospatial Foundation, http://www.osgeo.org) followed by several examples of GRASS applications related to environment.

2. Methods and tools for environmental analysis and modeling in GRASS6

2.1 Overview of raster and vector based analysis and modeling

Raster-based spatial analysis and modeling provides the most mature, core GRASS functionality. Besides the standard raster data management and manipulation tools – such as spatial queries, profiles, basic statistics, reclassification and recoding, buffers, and neighborhood operators, it includes a powerful map algebra module *r.mapcalc* for customized spatial analysis and modeling

(Neteler and Mitasova, 2007). Cost surfaces, representing cumulative costs of moving between different locations support various optimization tasks (see section 2.3, Figure 1) including a module specially designed for walking person applications (such as locating lost persons, Ciolli et al. 2006).

Comprehensive set of tools for terrain modeling and watershed analysis allows users to compute terrain parameters and features that are essential for environmental modeling and management. GRASS includes modules for computing elevation surface gradients and curvatures, partitioning landscapes into watersheds and extracting stream networks (see section 3.2, Figure 3.), computing flowpath length and upslope contributing areas using different algorithms, filling depressions and computing surface water extent (lake) for a given elevation. Terrain parameters can be extracted at different levels of detail (Mitasova et al., 2005b, Wood 1996) and combined to create maps of land forms. Line of sight module can be used to provide input for location of communication towers, wind turbines, emergency management or assessment of impact of high buildings. Shading, sun illumination and computation of solar energy maps are covered by a related article in this volume (see Hofierka et al. 2007).

GRASS also includes a limited number of modules for modeling dynamic processes such as surface and subsurface flow (see section 3.3, Figure 4) and simulation of the spread of wildfires. Series of raster maps either produced by models or based on measured data can be analyzed using a module that computes each cell value in a new map as a statistical function of the values of the cells in the time series. More sophisticated support for management and analysis of multitemporal data is planned for the future GRASS releases. Important component of geospatial modeling is uncertainty analysis, several modules have been recently added to support this task, especially to study impact of random errors on surface analysis (Ehlschlaeger 1996). GRASS provides a basic support for 3D raster data (voxel volumes) including a 3D map algebra and 3D volumes interpolation (IDW and spline algorithms) and 3D flow (partial differential equation library with OpenMP parallelization support).

The set of modules for quantitative analysis of landscape structure and patches in general has been recently upgraded to a client-server, multiprocess implementation. It provides important tools in studies of impacts of clear cutting, urbanization, fires etc. on ecosystem structure. Two main groups of indicators are already implemented: patch and diversity indices. Map input to calculate landscape metrics include landuse/landcover maps, satellite or aerial imagery data. Imagery is in GRASS represented as raster data although several modules are specifically designed for image processing, such as geocoding, image rectification, classification, and fusion, calculation of vegetation indices, and visualization of color composites. Common multiband satellite data formats are also supported.

Transformation between the vector and raster data models is performed based on the type of the feature/phenomenon. Categorical data representing geometric features (points, lines and areas) are transformed using line and area vectorization and rasterization, with resampling performed using nearest neighbor method. For continuous fields, several methods for spatial interpolation are available, including bilinear, bicubic, bspline, IDW, and regularized spline with tension

(RST). RST is unique in that it optionally computes topographic parameters or partial derivatives simultaneously with the interpolation and it supports spatially variable smoothing. Raster data can also be spatially aggregated with the new cell values based on univariate statistics computed over all of the input cells whose centers lie within the output cell.

The GRASS vector engine has been completely redesigned and vector analysis and modeling capabilities have been substantially enhanced. Geometry and attribute storage are clearly separated, giving advantage to more flexibility in the data organization. The vector geometry was extended to manage 2D and 3D topological vector data. The new internal vector data format is portable across 32bit and 64bit platforms. A new spatial indexing system as well as category indexing accelerates data access. The attribute management includes full and flexible integration of database management systems (DBMS; currently supported are PostgreSQL, MySQL, DBF, SQLite and ODBC). SQL statements are used to manage and query attributes. Support for vector map overlays, intersections and extraction of features is implemented. Conversion between combined coordinate lists with attributes and maps is available. The integrated Directed Graph Library provides support for vector network analysis. Available algorithms are shortest path, traveling salesman (round trip), allocation of sources (subnetworks), minimum Steiner trees (star-like connections), and iso-distances (from centers). Traveling costs may be assigned both to nodes and arcs. Both directions of a vector line can be used, which permits to define a forward and a backward direction and to store their attributes into the related attribute table. Furthermore, integrated linear reference system (LRS) tools use linear features and distance measured along those features to locate objects or events, tasks which are often needed for management of infrastructure such as roads and utilities.

GRASS includes a relatively sophisticated 3D visualization module *nviz* that allows users to interactively visualize multiple surfaces along with color maps, and vector data, explore the 3D spatial relationships using moving cutting planes, adjust the light to study the structure of the modeled surface and create dynamic surfaces using animations. Volume visualization supports isosurfaces, profiles, and volume renderings.

2.2 GRASS as component of OSGeo software stack: R, WebGIS

The complete GRASS source code is available on the GRASS Web site. The code base is written in ANSI C programming language. To make the development of GIS tools more efficient, GRASS provides a large GIS library with documented application programming interface (C and C++ API). Two levels of programming are supported. Average users will use 'script programming' to simplify repeating processes, while advanced users can extend existing code or even develop new modules. The modular concept of GRASS provides huge potential for development. Most modules are also usable from command line which allows the user for integration into UNIX shell, PERL, PHP or Python scripts. The C API is exposed to other programming languages through a SWIG interface (currently PERL and Python).

Since GRASS does not provide high level geostatistical functionality, a couple of interfaces is available. These interfaces link to external software packages such as R statistical language, MATLAB, Octave, and gstat. Within GRASS data can be pre-processed, then taken over to the external package and results stored back. We used this framework to integrate a machine learning software suite within GRASS for predictive habitat suitability modelling of trees (Benito Garzón et al., 2006).

On top of this, Web based spatial services are developed. In particular, opportunities for integrating Spatial Data Infrastructure (SDI) into Internet based services have grown. Interoperability and standardization is granted by following standards of the Open Geospatial Consortium (OGC) for geodata and related information technologies. While internally data storage and processing may differ from proposed OGC standards, all recent FOSS for Geomatics systems provide data exchange interfaces to industrial standards. The open source UMN/Mapserver can be easily integrated with GRASS as it reads GRASS raster and vector data through the GDAL/OGR library directly from a GRASS database. Alternatively, GRASS vector data can be stored into a PostGIS spatial database which is then linked to UMN/Mapserver. A related project is PyWPS (Python Web Processing Service, http://pywps.wald.intevation.org/) which implements the Web Processing Service standard from Open Geospatial Consortium. PyWPS offers an environment for programming own process (geofunctions or models) for WebGIS applicaions. It is written with native support for GRASS to make GRASS modules directly accessible via Web. An application is the support of wireless hand-held devices (field mapping etc.) which request data subsets or give a push to complex and data intense algorithms on the server side and then just receive the results.

2.3 OSGeo Educational data set

GRASS development team has provided a sample GRASS project data set (location) for users and developers to get started with GRASS, learn and teach, develop tutorials and test new modules. This data set, that covered an area near Spearfish, SD (USA) is now over 10 years old and is being replaced by more extensive data that reflect the latest advances in mapping technologies and data distribution. The new data set will be used to develop comprehensive educational materials for GRASS and geospatial modeling in general, based on modern data including lidar, high resolution imagery, real-time environmental measurements and many others. It will also offer more possibilities for software testing, debugging, and new module development. The data set includes basic physical and socioeconomic data for he central North Carolina (NC) obtained from the public data repositories provided by the state and local governments. It is highly extendable including real time data such as USGS river gauges, weather and soil moisture, or multitemporal coastal topography. As an example of a spatial analysis course material (Neteler and Mitasova 2007) we present a task that involves response to a potential environmental disaster – burning truck with toxic chemicals. First we use the module *r.cost* and *r.drain* applied to the area (with speeds ranging from 5km/hr in off-streets areas to 120km/hr on highways) to identify fire stations that can provide rapid response and delineate the shortest path to the accident location (Figure 1). If only the street and road network is used, the same task can be solved using the vector networking tools. Then we run flowrouting tool *r.flow* on a DEM to evaluate potential spill impact that shows that the spill will be accumulated in a nearby wetland where it can be contained (Figure 1, insert). Finally, we can use the module *r.spread* along with the wind direction information to do a simplified estimate of the spread of the smoke, overlay it with residential areas and use this information to issue warnings or evacuation orders.

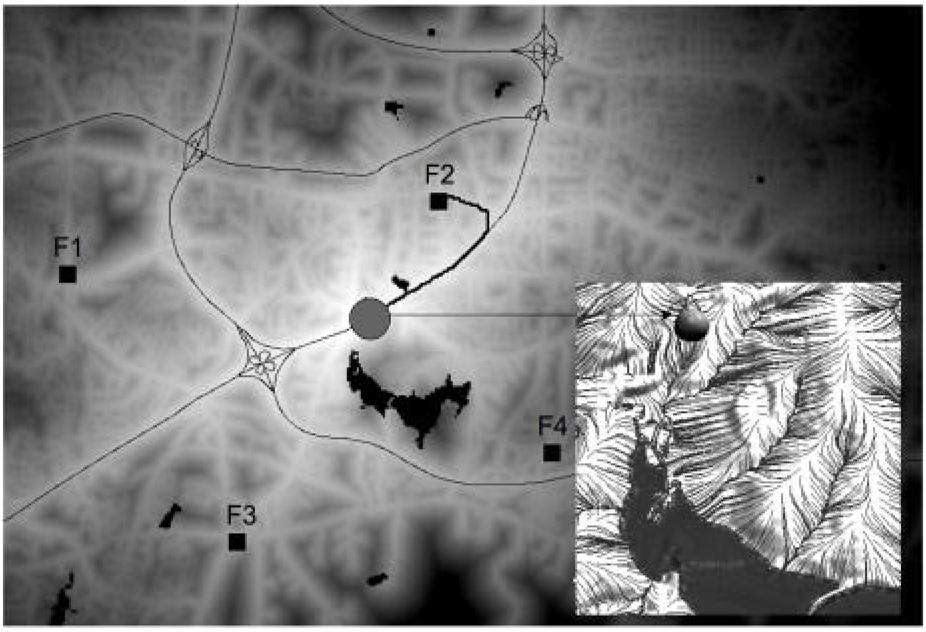

Figure 1. Example of a cost surface, shortest path and flow distribution analysis: circle represents a location of accident with potential toxic spill, shaded map represents the cost surface to get to the accident (the darker the color the higher the cost or time), wider black line shows the shortest path from the closest fire station F2 and insert shows 3D terrain near the accident and distribution of flow.

3. Applications

3.1 Coastal

NC has extensive lidar data sets that cover 10 years of annual surveys (with the exception of 2002 and 2006). This data set offers a new and unique view of barrier islands dynamics, impact of rising sea levels, and increased storm intensity on coastal erosion. The data set poses numerous challenges for processing as the surveys were performed during the period of rapid development of lidar technology. In addition to massive volume of data, the point density and coverage varies from year to year and integration of data at common resolution and spatial coverage has been essential for successful spatial-temporal analysis. We have used the module *r.in.xyz* that reads the ascii x,y,z points and computes raster maps representing number of points, mean, max, range and other values for each cell (the increase in point cloud densities was from 1pt/3m to 1pt/0.3m). The result of this analysis is used to select suitable resolution and interpolation method. Computed DEMs are then analyzed by *r.series*, in our case the results revealed that the longshore dune ridges were the most variable feature in terms of vertical change, temporal gradient parameter allows us to identify locations and time periods of most dramatic changes and study their link to storms or engineering projects.

Figure 2. High resolution (0.3m) lidar-based DEM interpolated and smoothed by the spline module *v.surf.rst* and visualized in 3D by the *nviz* module.

The resolution of some of the latest data is sufficient to reconstruct building geometry (Figure 2) and rapidly and efficiently map the damage to building structures after large storms and hurricanes. New set of tools allows users to extract buildings from multiple return data (Brovelli et al. 2004). Examples of additional coastal application include analysis of evolution of a threatened, large coastal dune located in a state park (Mitasova et al. 2005b), or impact of tsunami (Federici and Cosso 2006).

3.2 Watershed analysis

Urban development in Panama is producing severe land-use pressures on its watersheds. Deforestation rates associated with agriculture have accelerated, resulting in soil erosion and river sedimentation that have the potential to compromise the quality of its water resources. Thus, an understanding of the hydrochemical characteristics of the watersheds is required as a foundation for sound water resources management. To support an ongoing study that compares the surface water chemistry of the largely pristine basins with the watersheds affected by different land-use practices, SRTM elevation data (90m resolution) were used to derive detailed, Panama-wide stream network that provided more detailed representation of streams than the existing maps (Figure 3., only the major rivers are shown). Watershed analysis required processing of DEM with over 25 million grid cells, therefore a new module *r.terraflow*, specially designed for massive terrains, was used (Arge et al. 2003). Several watershed characteristics such as area, mean slope, river length were computed using GRASS terrain analysis and raster data management modules as inputs for study of relationship between of river water chemistry and physical properties of the watersheds.

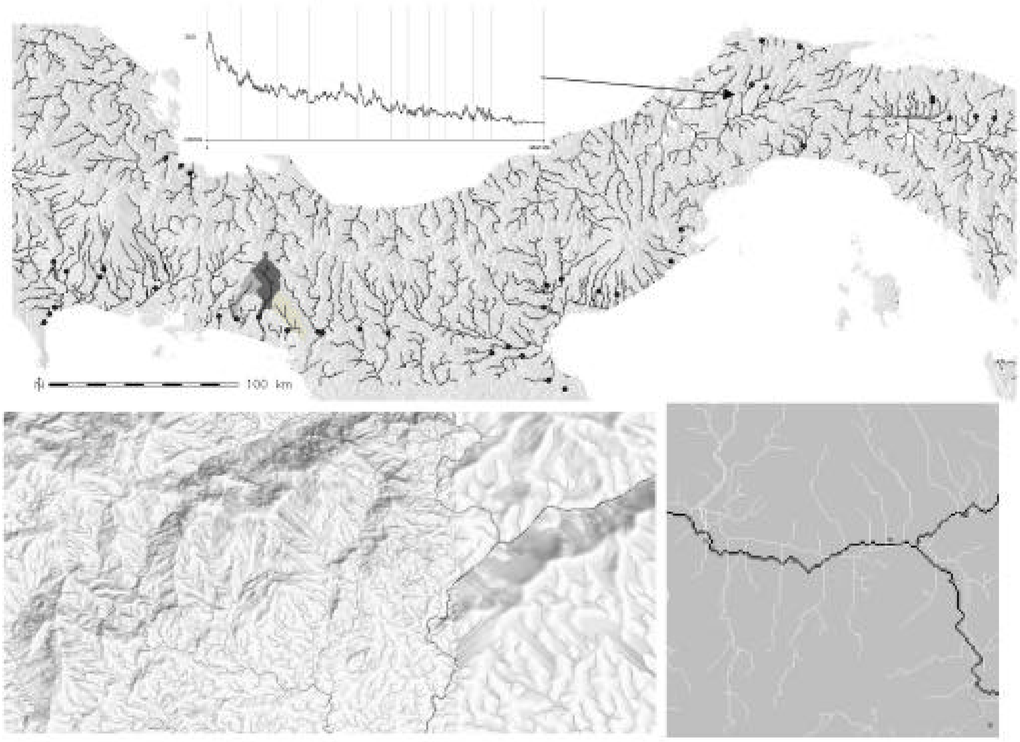

Figure 3. River network analysis for Panama using SRTM data (upper image), the profile illustrates the impact of vegetation cover on the terrain surface. IFSARE data provide higher level of terrain and stream network detail as shown on the image of patched IFSARE and DEM (lower left). Lowest cost algorithm performs better than the sink filling methods when appliesd to SRTM.

More detailed analysis was performed for central Panama by combining 10m resolution IFSARE with 90m SRTM data reinterpolated to 30m resolution by the RST method (Figure 3. illustrates the difference in the level of detail between the SRTM and IFSARE data in this area). The georeferenced sampling sites on rivers were used to evaluate the accuracy of derived stream network (Kinner et al. 2005). The accuracy was considered surprisingly good given that the elevation data included tropical forest cover (see the river profile Figure 3). Comparison with the standard depression-filling algorithms has shown that the lowest cost algorithm used in *r.watershed* produced more accurate stream network representation with this type of data where the depressions are mostly gaps in vegetation rather than artifacts of interpolation.

3.3 TanGIS

The concept of a Tangible GIS (Mitasova et al., 2006) is being explored as an intuitive interface for geospatial analysis and design based on interaction with a digital landscape through a physical model. The physical model is continuously scanned while the terrain parameters are recomputed and projected over the model surface. The system permits visualization of the impact of terrain modifications while they are performed so the user can continue to make adjustments until the desired outcome is achieved. The unique capability to easily add models of structures and buildings to the terrain model is important for exploring design alternatives for preparing plans for storm water and sediment control. To support real-world applications the system is being coupled with GRASS GIS and its simulation modules (Figure 4).

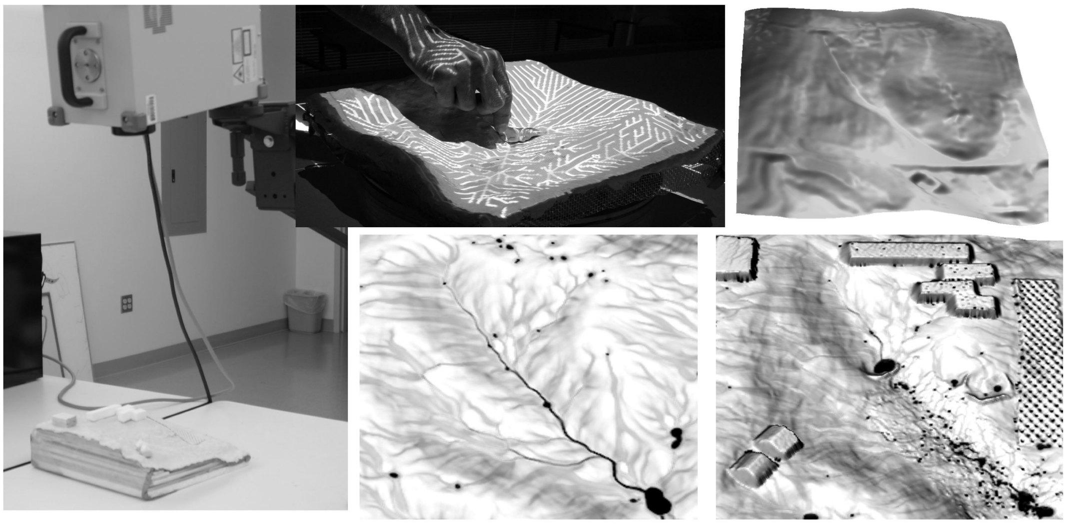

Figure 4. The physical model is scanned (left) and the result of analysis is projected over the surface (upper center). Simulated flow for the initial terrain model (lower center) that is based on the lidar-based DEM (upper left) and flow after modifications such as added buildings, rip-rap, a checkdam (lower right).

4. Conclusion

Over the past few years GRASS evolved from a specialist only to user friendly GIS with major advantages in interoperability and integration with related software. Peer-reviewed code leads to high quality algorithms with scientific input from research community. With the new vector and database support GRASS capabilities needed for environmental modeling have been substantially enhanced although several challenges remain, such as the scalability of the vector modules and more sophisticated support for multitemporal data.

Acknowledgements

ITC-irst is kindly acknowledged for the continous support of the GRASS development. Markus

Neteler was partially supported by the IAS Innovation Actions for Society project. We are also grateful for the continuing support by the U.S. Army Research Office and North Carolina Water Resources Research Institute.

References

L. Arge, J. Chase, P. Halpin, L. Toma, D. Urban, J. Vitter, and R. Wickresmesinghe. 2003. Flow Computation on Massive Grids. In *GeoInformatica*, 7(4) 283-313.

Benito Garzón M., Blazek R., Neteler M., Sanchez de Dios R., Sainz Ollero H., and Furlanello C, 2006. Predicting habitat suitability with Machine Learning models: the potential area of *Pinus sylvestris L.* in the Iberian Peninsula. Ecological Modelling, 197(3-4), pp. 383-393.

Bouktif S., Antoniol G., Neteler M. and Merlo E., 2006. A Novel approach to Optimize Clone Refactoring Activity. Proc. GECCO 2006, SBSE track, Seattle, WA, USA, July 8-12 2006, pp. 1885–1892.

Brovelli M.A., Cannata M., and Longoni U.M., 2004. LIDAR data filtering and DTM interpolation within GRASS, Transactions in GIS, vol. 8(2), pp. 155-174.

Ciolli M., Mengon L., Vitti A., Zatelli P., Zottele F., 2006. A GIS-based FOSS decision support system for the management of SAR operations in mountain areas. Proc. FOSS4G2006 - Free And Open Source Software for Geoinformatics, Lausanne, Switzerland, Sept. 2006, http://www.foss4g2006.org

Ehlschlaeger, C.R., and Shortridge, A.M., 1996, Modeling Uncertainty in Elevation Data for Geographical Analysis, Proceedings of the 7th International Symposium on Spatial Data Handling, Delft, Netherlands, August 1996.

Federici, B., Cosso, T., 2006. Tsunami inundation maps and damage sceneries through the GIS GRASS. Proc. FOSS4G2006 - Free And Open Source Software for Geoinformatics, Lausanne, Switzerland, Sept. 2006, http://www.foss4g2006.org

Hofierka J., Huld, T., Cebecauer, T., and Suri, M., 2007. Open source solar radiation tools for environmental and renewable energy applications, this volume.

Hofierka J., Parajka J., Mitasova H., Mitas L., 2002. Modeling impact of terrain on precipitation using 3-D spatial interpolation. Transactions in GIS, 135-150.

Kinner D., Mitasova, H., Harmon, R., Toma, L., Stallard R., 2005. GIS-based Stream Network Analysis for The Chagres River Basin. In: The Rio Chagres: A Multidisciplinary Profile of a Tropical Watershed, R. Harmon (Ed.), Springer/Kluwer, p. 89-95

Mitasova, H., Mitas, L., Ratti, C., Ishii, H., Alonso J., and Harmon, R.S., 2006. Real-time Human Interaction With Landscape Models Using a Tangible Geospatial Modeling Environment, IEEE Computer Graphics & Applications, 26(4), pp 55-63.

Mitasova, H., Mitas, L. and Harmon, R.S. 2005a. Simultaneous spline interpolation and topographic analysis for lidar elevation data: methods for Open source GIS, IEEE GRSL, 2(4), pp. 375- 379.

Mitasova, H., Overton, M., Harmon, R.S. 2005b. Natural Processes and Human Impacts: The Dynamic Evolution of a Coastal Sand Dune Complex at Jockey's Ridge, North Carolina, Geomorphology, 72, pp 204-221.

Neteler, M. and Mitasova, H. to appear in 2007. Open source GIS: A GRASS GIS Approach, Third Edition. Springer, New York.

Mitasova, H., Thaxton, C., Hofierka, J., McLaughlin, R., Moore, A., Mitas L., 2004. Path sampling method for modeling overland water flow, sediment transport and short term terrain evolution in Open Source GIS. In: C.T. Miller, M.W. Farthing, V.G. Gray, G.F. Pinder eds., Computational Methods in Water Resources, Elsevier.

Wood, J. 1996, The Geomorphological characterisation of Digital Elevation Models. Diss., Dept. of Geography, Univ. of Leicester, U.K. online at: http://www.soi.city.ac.uk/~jwo/phd/