Chapter Six

Evaluation Study

for some GIS Software packages
Chapter Six

Evaluation Study for some GIS Software packages

6.1 GRASS Release 5

6.1.1 What is GRASS?

GRASS (Geographic Resources Analysis Support System) is an open source, Free Software Geographical information system (GIS). As other commercial GIS software, GRASS 5 can handle image processing either in raster format or topological vector format as well as performing spatial modeling, and visualization. Although it has been developed on UNIX platforms in the beginning, GRASS 5 can operates on various platforms through a graphical user interface and shell in the X Window System. It is released under GNU General Public License (GPL). “Originally developed by the U.S. Army Construction Engineering Research Laboratories (USA-CERL, 1982-1995), a branch of the US Army Corp of Engineers, as a tool for land administration and environmental planning by the military, GRASS has evolved into a powerful utility with a wide range of applications in many different areas of scientific research. GRASS is currently used in academic and commercial settings around the world, as well as many governmental agencies including NASA, NOAA, USDA, DLR, CSIRO, the National Park Service, the U.S. Census Bureau, USGS,
and many environmental consulting companies.\footnote{http://en.wikipedia.org/wiki/GIS}

\section*{6.1.2 Spatial data analysis in GRASS 5}

GRASS 5 has the ability to perform all spatial data analysis that can be applied to geographical data such as storing, manipulating and visualizing of geographical data, and handling tasks such as positional information which describe where the points are located (place, context and space), given usually as coordinates in two dimensions either relative to local position or absolute position. Using Local position (coordinate) system is sufficient in some cases as analysis of medical imagery, and absolute or global coordinate system (degrees easting and northing, longitude and latitude), for example UTC, is required in case of spatial data analysis. GRASS 5 has the capability to perform either local or global coordinate tasks so that it can perform different spatial analysis such as calculating distances between observations, and/or for describing the network topology of their relative positions.

\section*{6.1.3 GIS data models}

In the early days most GIS data have been constructed from paper map analogies, because GIS systems were mainly designed for surveyors and others maintaining and renewing existing paper archives of maps, no major break with tradition occurred. With the increase needs for more analyzing ability, some drawbacks as realized such as incapability of...
handling three dimensional data sufficiently, another is the incapability to reflect time factor in the analysis process. More details may be found in (61) and (75).

Now days, two major families of GIS data models for managing spatial data are used, and both, GRASS 5 can deal with. One family is what so called the vector GIS family that meets the surveying standard requirements better, by focusing on the exact position of points, and objects derived from them. Vector formats usually represented as points, lines and polygons (see Figure 5a, pp 29). The other family is raster format, which usually represented as units, typically squares or rectangles, at a selected resolution (see Figure 5b, pp 29). Vector format is best to handle Property boundaries or utilities inventory data, while many environmental analyses, especially those concerning remote sensing, raster format suits them better, as long as resolution plays important rule. But if we considered working in a very high resolution, then these tow family can be considered to be equivalent. For example, the limits between the families are removed by the extensive use of scanned images, which are store in an image format such as TIFF.

Valuable sources of data and information are OpenGIS, which is an industry standardization group, where a number of the projects coordinated through the www.remotesensing.org website. In particular GDAL\(^2\) gives access to multiple raster formats, permitting conversion without loss of

\(^2\) http://www.gdal.org/index.html
georeferencing. The same site hosts the maintained version of PROJ.4\textsuperscript{3}, a library and helper programs for projection.

\textbf{6.1.4 The R-GRASS interface in GRASS 5}

As we have seen earlier that GRASS was originally developed within the US Army for the purpose of monitoring the environmental effects of military exercises. It was written in C as a set of programs run from the Unix shell prompt, where environment variables are defined on startup to indicate the current “LOCATION”, the study area, and the user’s “MAPSET”. To avoid users overwriting files of each other, GRASS 5 creates a user hierarchy under each location. Each such mapset will have either default window or user-defined window, the user also specify area and resolution to be used.

Comparing GRASS 5 with its earlier versions one can notice a lot of improvements. GRASS 5 becomes full floating point support, NULL-value support, and compilation under Cygwin becomes feasible.

\texttt{R} (statistical data analysis language) can interact with GRASS 5 well when running from inside the GRASS shell, and issues commands to GRASS. The current \texttt{R} version is 1.2.0, an older version for \texttt{R} versions before 1.2.0 is available on CRAN\textsuperscript{4}. The first version of the interface as described in (21) used ASCII temporary files to move data between \texttt{R} and GRASS for sites using functions \texttt{read.table()} and

\textsuperscript{3} http://proj.maptools.org/
\textsuperscript{4} http://cran.r-project.org/
write.table() on the R side, and scan() function to import raster layers as vectors transferred through ASCII. These functions could in turn be run from within R through system() command, allowing the interface to be encapsulated in a library of R functions. Figure 26 shows the nesting of GRASS 5 and R.

Figure 26: Nesting of shell levels in the R/GRASS interface
6.2 Spacestat-Arcview and S+Grassland

6.2.1 Introduction

The integration of spatial statistical methods and GIS has been an active topic of research in both the academic and commercial GIS communities in recent years (12). Different strategies have been proposed to establish a link between the two. These strategies can be conceptualized as ways to combine the "traditional" spatial analysis functionality of the GIS (e.g., spatial queries, buffering, overlay) in a "GIS Module" with spatial statistical and data analysis methods in a "Spatial Data Analysis Module" (6). One approach towards implementing the combination of these two modules into an overall framework for spatial analysis consists of incorporating elements of one into the other (a so-called encompassing approach), such as the addition of GIS capabilities to a statistical software package or the extension of a GIS with statistical functionality. The latter is typically not included in the standard software release but is made possible by taking advantage of macro or script languages supported in GIS software. Examples of this are the use of the AML (Arc Macro Language) for Arc/Info and the Avenue script language for ArcView to extend the functionality of the GIS with EDA (Electronic Design Automation) tools (17) or descriptive spatial autocorrelation statistics (33) (14) (28). The advantage of such an approach is that the added functions
are fully integrated into the familiar GIS data model and user interface. However, the user bears the full burden of both identifying the appropriate methods and developing effective scripts. Moreover, the scripting environments are somewhat limited in terms of the size of data sets that can be handled and often seriously deficient in terms of speed.

A different strategy consists of developing an efficient linkage between existing commercially available GIS and statistical software packages. A number of taxonomies have been suggested to implement such a linkage, such as loose coupling vs. seamless coupling, a unidirectional link vs. a bidirectional link and a static link vs. a dynamic link (for reviews, see, e.g., (2) and (43)). There are now several examples of such approaches, developed both in academic environments (e.g., (39) and (95)) as well as in the commercial sector such as the S+GISLink (71) and the S+ArcView Link (13).

In this section, we contrast two different approaches to linking a GIS and a spatial statistical module. One is based on a loose coupling strategy by means of an efficient interchange of input and output files between the SpaceStat spatial data analysis software (7)(8) and the ArcView GIS (36). The other is designed as a seamless integration (close coupling) between the S-PLUS statistical computing environment (73) and the Grassland GIS (65). For each of these approaches we next briefly describe the overall architecture, linkage mechanism and operational implementation. We finish with a comparison of the relative merits of these approaches and some thoughts on future developments.
6.2.2 The *SpaceStat-ArcView* Link

The link between *SpaceStat* and *ArcView* is an extension and implementation of the prototype suggested in Anselin *et al.* (1993)(6), based on loose coupling between *ArcView* as the visualization engine and *SpaceStat* as the spatial data analysis engine. Its main objective is to provide an efficient way to display the results of spatial statistical analyses by means of the GIS and to obtain locational information for use in the statistical analysis from the GIS.

*SpaceStat* (2)(7) is a software package for the analysis of spatial data developed in the *GAUSS* programming environment (9) for DOS operating systems. *SpaceStat* includes a broad range of test statistics for both global as well as local spatial autocorrelation, and econometric estimation methods and specification tests for regression models that incorporate spatial dependence (spatial autoregressive models). In addition, *SpaceStat* has extensive capabilities to construct, manipulate and analyze spatial weight matrices, an important tool in the analysis of spatial autocorrelation.

*ArcView* is currently one of the most popular desktop GIS software environments (36), primarily geared to the manipulation of vector data. *ArcView* can easily be customized by means of scripts written in the object oriented Avenue script language (supported by *ArcView* 2.1 and higher) and has recently been extended with optional modules for the analysis of raster data (the Spatial Analyst module) and network data (the Network Analyst module).
6.2.2.1 Architecture

The main features of the SpaceStat-ArcView Link are:

- A concentration on Exploratory Spatial Data Analysis (ESDA) of net data;
- Sharing the work between the ArcView and SpaceStat; and
- To be working in different platforms.

It is obvious that establishing direct interaction between ArcView and SpaceStat is not visible, simply because SpaceStat uses the DOS version of GAUSS while ArcView runs under Windows, so there is no way to call internal SpaceStat functions from the ArcView environment.

To overcome this limitation, an indirect communication was established between the two software through auxiliary files that have standard file names and data formats. So that spatial data is moved from ArcView to SpaceStat for analysis, and location-specific results are passed back from SpaceStat to ArcView for visualization.

6.2.2.2 Linkage Mechanism

This linkage mechanism was established by customizing the ArcView user interface and adding functions to perform the following tasks:

- Driving the output of ArcView to a file that have formats compatible with SpaceStat.
- Mapping of SpaceStat output to View window of ArcView.
In other words, the linkage mechanism established by creating an intermediate files with formats designed to be compatible with both SpaceStat and ArcView.

6.2.2.3 Operational Implementation of the SpaceStat-ArcView Link

Because the link between SpaceStat and ArcView was implemented primarily in the ArcView environment through customized menus, two additional menus and a few extra buttons and tools have been added to the standard View window of ArcView: a Data menu and an Explore menu.

Data menu, which have six functions in three groups:

➢ Functions to add the X-Y coordinates of polygons and constructing an indicator variable for selected locations.

➢ Functions to read ArcView Shape file to construct of spatial boundary files.

➢ Functions to transfer data between ArcView and SpaceStat.

Explore menu, which have eight functions in three categories:

➢ Functions to visualize the spatial distribution of the data. Such as Histogram, Box Plot and Box Map.

➢ Functions to visualize spatial autocorrelation in attribute variables. Such as the Spatial Lag Bar Chart and Spatial Lag Pie Chart.

➢ Functions to visualize local spatial association. Such as Moran Scatterplot and Map.
For more technical details please refer to (3),(5) and (4).

It is worth noticing that those functions are implemented as external DLL functions instead of using Avenue scripts, and the result is more speed and more flexibility.

6.2.3 The S-PLUS and Grassland Link

AT&T Bell Laboratories has developed a statistical software package based on the S language and known with the name S-PLUS(73). This software contains more than 2,000 functions and most of them can be is easily extended either by add-on modules, or by user-written functions. One of the good add-on module is S+SpatialStats (72), which adds functions for analysis of spatial data, geostatistical modeling, descriptive spatial statistics, and spatial regression analysis.

In the other hand, Sun Microsystems has developed a commercial version of GRASS and named it Grassland, which run under windows (65). The Graphical User Interface (GUI) for Grassland is developed using Tool Command Language /Tool Kit (Tcl/Tk) scripting language (83). Each Grassland communications component has an equivalent Tool Command Language extension loaded dynamically at boot time. And can manipulate raster and vector data to perform tasks such as overlay, buffer zone analysis, and terrain analysis.

6.2.3.1 Architecture

The main features of the S+Grassland Link are:
Almost full compatibility of $S$-$PLUS$ functions with Grassland.

The link can perform analysis on multiple data layers.

The link can perform statistical analyses either to raster data or vector data.

The $S$+$Grassland$ Link interface is based on a combination of the Tcl/Tk, OGDI (Open Geospatial Datastore Interface) (66) library and $S$+API (Application Programming Interface) techniques. The linkage between Grassland and $S$-$PLUS$ is a close coupling with a bidirectional conversation. Data are exported from Grassland to $S$-$PLUS$ and $S$-$PLUS$ functions are accessed directly from the Grassland environment.

6.2.3.2 Linkage Mechanism

The $S$+$Grassland$ Link is established by several functions in DLLs form. For example, exporting data and passing commands from Grassland to $S$-$PLUS$ was performed by functions in DLL form, and another DLL take care of creating an $S$-$PLUS$ data driver to access $S$-$PLUS$ objects from within Grassland. The GRASS data can be retrieved by $S$+$Grassland$ and transformed into the $S$-$PLUS$ object structure using $S$+API. The $S$-$PLUS$ data driver is a collection of Tcl/Tk script programs and C routines.

The link between Grassland and $S$-$PLUS$ is direct and dynamic, so that URL (Uniform Resource Locator) connection can be established between Grassland and the $S$-$PLUS$ workspace from the Librarian window. Once connection is established, $S$-$PLUS$ objects and results from statistical analyses can be accessed directly from Grassland.

122
6.2.3.3 *Operational Implementation of the S+Grassland Link*

Geographical data can be exported from *Grassland* to *S-PLUS* either from the Mapviewer window or the Librarian. Where raster data are exported to *S-PLUS* as a matrix object. And area and Point data are exported to *S-PLUS* as a data frame that contains X and Y coordinates along with associated attribute data. Text data, in turn, are exported to *S-PLUS* as a data frame that contains X and Y coordinates along with the attribute text data.

On the other hand, *S-PLUS* objects can be accessed directly from *Grassland* by establishing URL connection from the Librarian window. The “Open Connection” dialog in the “File” menu establishes URL connection to the workspace and lists the *S-PLUS* geospatial objects (Raster, Point, Text, Pmap) in the workspace.

6.2.4 *Main Differences and Similarities between the SpaceStat-ArcView and S+Grassland Link*

Major Differences and similarities between the SpaceStat-ArcView and S+Grassland Link can be listed as follow:

➢ They both provide functions to construct the spatial weight matrices contiguity.

➢ While *S+Grassland* link is direct and almost fully compatible, *SpaceStat-ArcView* link is indirect and much less compatibility and efficiency.
6.2.5 Strategies for Future Development

Although the linkages between GIS and statistical software packages have witnessed increasingly improvements, they considered to be fairly rudimentary. And more effective and compatible linkage is expected, such linkages should be simple and allow bidirectional data exchange, and more importantly, to set up a real-time functional integration between the two software packages.

Bao, S. and Anselin (12) suggested that “Instead of linking a comprehensive statistical (or spatial statistical) module with the GIS as a single piece of software, it may perhaps be more effective to implement selected methods in small self-contained software applets that can be invoked from within the GIS. Clearly, the reverse strategy is promising as well, to implement small applets incorporating GIS functionality. An API such as S+API may provide the first step towards such a decentralized approach that would allow the individual user to customize the spatial data analysis toolbox for each application”.
Overall Conclusions and Recommendations

From this study, we come out with several conclusions:

➢ Geographical Information system techniques found and proved to be very essential and very efficient when we concern to find answers for critical questions at the national level as long as the global level. Such questions could be: How to discover, assesses and use, at the national level, the natural and human resources optimally? How to discover, assesses and use the global resources optimally? How to monitor the earth globe so that we, as human being, get better knowledge, and accordingly, better management for our limited resources? How to use this knowledge to avoid certain disasters that might threaten the human being existence? It is obvious that the answers for such questions play the main role for any country to achieve development and wealth for citizen of that country.

➢ GIS techniques also proved to be helpful tool in the military field, just recall that GRASS was born on the Army Construction Engineering Research Laboratories (USA-CERL, 1982-1995), and all of us saw how GIS techniques (i.e. remote sensing, GPS, and missiles directions etc) are used effectively in the third Gulf War so that Saddam’s regime collapsed in just weeks.

➢ GIS techniques also proved to be helpful tool for medium and big
companies, specially the companies that its activities spread over wide geographical area such as telephone companies and water companies. Here GIS techniques can be used to make the appropriate decisions in the planning stage as well as when the company start its activity.

- Neural Networks also considered to be useful, as a classification technique in GIS, hence we consider NN as a very interesting part of Artificial Intelligence subject which we expect to achieve great improvements in the near future.

For all of these, and more, we conclude with the following recommendations:

- We recommend the governments in India and the Republic of Yemen, and all under-development countries, to pay more attention and high priority for GIS techniques so that achieving development planning becomes visible.

- We also recommend the private sectors in both countries and in all under-development countries to invest more in GIS techniques to gain its potentials.

- We recommend all educational institutions to be more concern about teaching Artificial Inelegant subject in general, and Neural Networks in particular. Simply because it is expected to be the future.