From SpaceStat to CyberGIS: Twenty Years of Spatial Data Analysis Software

Luc Anselin

2011

Working Paper Number 06

From SpaceStat to CyberGIS: Twenty Years of Spatial Data Analysis Software

Luc Anselin

GeoDa Center for Geospatial Analysis and Computation

School of Geographical Sciences and Urban Planning

Arizona State University

Tempe, AZ 85287

luc.anselin@asu.edu

Abstract

This essay assesses the evolution of the way in which spatial data analytical methods have been incorporated into software tools over the past two decades. It is part retrospective and prospective, going beyond a historical review to outline some ideas about important factors that drove the software development, such as methodological advances, the open source movement and the advent of the internet and cyberinfrastructure. The review highlights activities carried out by the author and his collaborators and uses SpaceStat, GeoDa, PySAL and recent spatial analytical web services developed at the ASU GeoDa Center as illustrative examples. It outlines a vision for a spatial econometrics workbench as an example of the incorporation of spatial analytical functionality in a cyberGIS.

1. Introduction

The software environment that supports spatial analysis in general and spatial data analysis in particular has changed dramatically since the initial ventures in this field appeared in the late 1980s and early 1990s. Whereas Haining (1989) once lamented the lack of specialized software as a major impediment to the adoption and application of the proper spatial statistical methods in empirical practice, this is no longer a valid concern. In part in response to calls by many scholars (e.g., Openshaw 1990, Anselin and Getis 1992, Goodchild et al 1992), a great variety of software tools are now available to carry out a wide array of spatial data analytical techniques on diverse platforms, developed both by the commercial and the non-commercial sector (Goodchild 2010). As I argued in Anselin (2010), spatial econometrics (and spatial analysis) has moved from the margins to the mainstream of the methodological portfolio in the social sciences. I see the development of easily accessible and user-friendly software as a major contributing factor in accomplishing this evolution.

In this essay, I offer an assessment of this software evolution, both retrospective and prospective. I review the important changes that occurred over the past two decades in terms of the type of functionality embodied in the software, the architectures on which it was implemented, and its connection with GIS. I link these aspects to broader trends in software development in general, such as the open source movement and the advent of cyberinfrastructure. The review is focused on the statistical analysis of so-called lattice or regional data (Cressie 1993) and mostly excludes software for the analysis of point patterns or geostatistical models. As I argued in Anselin (2005), the distinctive data models required for the analysis of lattice data pose a challenge to standard statistical

software, which is much less the case for the treatment of point coordinates.

The viewpoint presented is highly personal, in that I will draw extensively on my own experience in this regard that now spans more than 25 years. While I focus on the broader trends, I will illustrate my points with software tools whose design and development I was intimately involved with, including SpaceStat, GeoDa, PySAL and the current generation of spatial analytical web services being created at the GeoDa Center for Geospatial Analysis and Computation. In the process, I will draw extensively on previous reviews and assessments I offered in Anselin and Hudak (1992), Anselin and Bao (1997), Anselin (2000, 2005, 2010), Anselin et al. (2004, 2006), and Rey and Anselin (2006).

In the remaining sections, I start by providing a brief historical background for the evolution of spatial data analysis software up to the turn of the century, with a particular focus on its position relative to mainstream GIS. I next discuss three important aspects that I see as key influencing factors that stimulated spatial analytical software development in the 21st century. First, I consider the role of methodological innovations in both exploratory and confirmatory analysis. I focus on the effect of developments in the fields of geovisualization and exploratory spatial data analysis (ESDA) on the one hand and spatial econometrics on the other hand in driving the functionality of the software. Second, I review the role of the open source software movement in stimulating new development and broadening the community of developers and adopters. I then move to the role of the internet, in the form of web-based spatial analysis, spatial analytical web services, and the advent of a scientific cyberinfrastructure for geospatial analysis. I close with some concluding comments and speculation about future directions.

2. Background

In this section, I offer a historical perspective on the origins of the development of software tools for spatial data analysis primarily up to the beginning of the 21st century. I start with a mention of the initial discussions in the late 1980s and early 1990s dealing with the respective roles of spatial analysis and GIS. I next briefly review a number of pioneering software development efforts. This review is not intended to be comprehensive, but illustrative of a number of approaches taken to implement spatial methods in software tools during the 1990s. I end the section with a short discussion of the role of the SpaceStat package in this regard.

1.1. GIS and Spatial Analysis

The impetus for the development of specialized software that implements spatial analytical methods came from the sense among academics in the late 1980s that the emerging GIS technology was lacking in terms of analytical capability and thereby presented a missed opportunity for advancing science (e.g., Goodchild 1987). In part as a reaction to this, the U.S. National Science Foundation established the National Center for Geographic Information and Analysis (Abler 1987) and similar research centers followed in other parts of the world.

Initial discussion very much saw the GIS and the statistical functionality as separate but related entities and focused on the relationship between the two (e.g., Openshaw 1990, Ding and Fotheringham 1992). Analysis was seen as one of many functions of a GIS and the debate pertained to how much of this functionality should be part of the GIS itself as opposed to being a separate piece of software. For example, in Goodchild et al (1992) the distinction between close coupling and loose coupling was

emphasized. In Anselin and Getis (1992) three different ways of connecting the analytical functionality to the GIS database were outlined: full integration, which is referred to as "embedding," a "modular" approach with efficient links between the parts ("close coupling"), and a complete separation between the two with simple import and export functions ("loose coupling"). It should be noted that to date, this distinction has become largely moot (see also Goodchild 2010) and all three forms are present in modern products. In addition, a fourth type has emerged, in which a subset of the standard GIS functions (such as input, storage and output) is subsumed in a spatial statistical package that operates independent from a GIS. In part, this can be attributed to the explosion of activity in the open source GIS movement, which freed developers from the near-monopoly of commercial GIS software vendors that existed in the early 1990s. I return to this point in Section 3.

At the time, the dominant commercial GIS software essentially lacked any statistical capabilities. Partial exceptions with specialized functionality were niche packages such as SPANS, IDRISI and TRANSCAD, but other than data selection (queries) and manipulation (buffering, overlay), analysis in the statistical sense was absent in ESRI's market leader Arc/Info. Most initial efforts therefore focused on extending the GIS functionality of Arc/Info by providing a link to existing commercial statistical software, exemplified by the so-called "archeologists's workbench" (Farley et al. 1990). The key aspect of these products was that the (commercial) GIS played a central role as the provider of data storage and retrieval and as the main means of visualizing the results.

During the 1990s, a great many software tools were developed, mostly in the non-

commercial academic world. These took the form of freestanding packages with limited functionality, extensions of the GIS with statistical functions written in macro and script languages, and more efficient linkages between the relevant pieces of software (for a more extensive review, see Anselin 2000, 2005). By the early 21st century, the new technology of object orientation and the use of software components greatly facilitated the array of combinations that could be developed (e.g., Symanzik et al. 2000, Zhang and Griffith 2000, Ungerer and Goodchild 2002). In addition, a vibrant open source community had emerged (Bivand and Gebhardt 2000). Finally, with the inclusion of a spatial statistics toolbox in ESRI's ArcGIS version 9 in 2004, spatial statistics became prevalent in the core functionality of the leading commercial GIS software as well. To date, the number of software solutions available is virtually unlimited and the lack of specialized software can no longer be considered to form an impediment to the adoption of spatial analytical methods (Anselin 2010).

1.2 Pioneers

By the early 1990s, a number of software packages started to appear that offered spatial analytical functionality in one form or another. By the end of the decade, a tremendous amount of progress had been made (Anselin 2000), so that I choose to refer to these efforts as "pioneers." Tools developed after 2000 will be included in the review of three driving factors in Sections 2-4.

Arguably one of the first examples to provide "spatial" exploration was the application of the principles of dynamic linking and brushing (Becker et al 1987, Stuetzle 1987, Cleveland and McGill 1988, Monmonier 1989) in the Spider and Regard software developed by Haslett, Unwin and associates (Haslett et al 1990, 1991, Unwin 1994).

Implemented on an Apple Macintosh platform, these packages took advantage of the efficient graphics bitmap to obtain real-time brushing of linked maps and statistical graphs. However, the focus was primarily on data exploration and "statistical" functions were limited (an exception is exploratory variography in Bradley and Haslett 1992). Whereas Haslett, Unwin and associates approached the software development primarily from the perspective of a statistician, a similar effort originating in cartography led to implementations of dynamic maps in the cdv package of Dykes (1997, 1998) and interactive maps for visual data exploration in Andrienko and Andrienko (1999).

The first package to implement a suite of spatial statistics and spatial regression in a free-standing environment was SpaceStat, released in 1991 by NCGIA (Anselin 1992).[1] For a long time, this was the only software that included a full range of local and global spatial statistics, regression diagnostics and both maximum likelihood and method of moments estimation of spatial regression models. It was followed by S+SpatialStats (Mathsoft 1996), an add-on to the commercial S-Plus statistical programming environment that included point pattern analysis, geostatistics (variogram analysis, kriging) and limited spatial regression. Point pattern and variogram analysis had been available for some time as add-on functions to the S-Plus software, e.g., in the so-called MASS library of Venables and Ripley (1994). Info-Map, a software companion to the Bailey-Gatrell text (Bailey and Gatrell 1995) similarly included a limited set of statistical functions. In contrast to the other packages, Info-Map included its own visualization. The latter was absent from SpaceStat, whereas S+SpatialStats relied on a link with ESRI's ArcView or the GRASS GIS for mapping of results (Bao and Martin 1997, Bao et al.

[1] The original SpaceStat has no relationship to the current SpaceStat 3.0 distributed by Biomedware (http://www.biomedware.com).

2000). With the exception of S-Plus, which also ran in the unix operating system, these applications were implemented on the Microsoft Windows platform.

Several other efforts relied on ESRI's ArcView or Arc/Info software to manage the data handling and visualization, whereas the statistical functions were included in a closely coupled module. In a Windows operating system, in addition to the link between ArcView and S+SpatialStats, there was also a dedicated "extension" for ArcView that linked with SpaceStat as the computing engine, primarily for the visualization of local clusters and spatial outliers (Anselin 2000). In addition, developed primarily for unix workstations, there were a number of efforts to link XGobi and Arc/Info or ArcView for exploratory spatial data analysis and variography (Symanzik et al. 1994a, b). Finally, the SAGE package included spatial regression and regionalization linked to Arc/Info (Haining et al. 1996, Wise et al. 2001). In contrast to the previously mentioned MS Windows packages, the fact that SAGE and XGobi linkages were primarily implemented on unix workstations limited their adoption in empirical practice.

Two additional efforts are worth mentioning that took a different approach. In contrast to the previous packages, which were self-contained and closed, LeSage and Bivand spearheaded the development of open toolboxes to implement a range of spatial statistical and spatial regression functions. LeSage's spatial econometrics toolbox was implemented using the commercial MatLab software (LeSage 1999), whereas the work of Bivand and collaborators used the open source R environment for statistical programming (for an review of early efforts, see Bivand and Gebhardt 2000). Both of these development efforts continue unabatedly to date (e.g., Bivand et al. 2008, LeSage and Pace 2009). Others were more short-lived, such as the implementation of spatial

statistics in X-LispStat (Brundson 1998) which never saw widespread adoption.

By the end of the decade, a wide range of software options were available, although SpaceStat arguably remained the market leader. Before turning to the developments post 2000, I briefly describe the motivation for and features of this package.

1.3. SpaceStat

The origins of SpaceStat are the routines written to provide the empirical illustrations in my Spatial Econometrics text (Anselin 1988). While a number of individuals had written code to implement maximum likelihood estimation of spatial regression models, this code was not generally available at the time. Also, it was near impossible to implement these methods using standard commercial statistical or econometric software. Typically, these packages lacked functionality to deal with spatial weights and the generic nonlinear optimization routines included were not optimized for maximizing the particular likelihood functions needed for spatial regression. In addition, the available scripting languages or macro programming environments were not optimized and lacked features and performance. Also, none of the regression diagnostics for spatial correlation could be carried out using standard software (see Anselin and Hudak 1992, for some examples of scripting approaches).

The original set of Gauss routines for the estimation of spatial regression models (Anselin 1989) were distributed in limited circles and had gained a degree of adoption that warranted turning them into a more organized and self-contained package. When NCGIA was established, software dissemination was folded into its activities and the first official release of SpaceStat occurred in late 1991.

SpaceStat was completely written in GAUSS, a matrix language, and had a very

rudimentary DOS-based user interface. While the GAUSS language was very fast, it initially was constrained to manipulating matrices that took less than 64K of memory (in double precision), which limited analysis to N < 80 (this constraint was removed in later versions). A major update of SpaceStat was released in 1995, which included the use of sparse format weights and the newly developed local indicators for spatial association (Anselin 1995a, b).

The core of SpaceStat included spatial autocorrelation statistics (join counts, Moran's I, Geary's c, QAP), diagnostics for spatial autocorrelation in regression models (Lagrange Multiplier tests) and estimation of spatial regression models (maximum likelihood, method of moments, spatial regimes). The emphasis of the software was on computation and it lacked any means to visualize the results. This was accomplished in the SpaceStat Extension for ArcView, a link between the two packages that worked from within ArcView. It was implemented using a combination of ESRI's Avenue scripting language and C code. The main purpose of the extension was to visualize the results of the LISA statistics on a map. The extension also included functionality to create a spatial weights file from an ArcView Shape file format (Anselin and Smirnov 1997, Anselin and Bao 1997). The limitations inherent in the linking strategy motivated the decision to develop a self-contained package that included both the visualization and the spatial statistics and broke the dependence on a GIS. This eventually led to GeoDa (Anselin et al. 2006).

2. Methods: Exploration and Confirmation

By the beginning of the 21st century, there were sufficient software packages in circulation that the Center for Spatially Integrated Social Science (Goodchild et al 2000)

instituted a software clearinghouse as part of its dissemination efforts. In this section, I review the main developments during the decade since the beginning of the century in light of the notion offered in Anselin and Getis (1992) that the analytical capability associated with a GIS separates into exploratory and confirmatory analysis. In many ways, this methodological division has also been evident in the more recent development of software tools (past 2000). On the one hand, new insights from exploratory data analysis, geovisualization and visual analytics have led to a category of visualization tools, primarily driven by efforts in cartography and computer science. On the other hand, new methods for the estimation of spatial regression models and their computational demands have yielded a new generation of spatial statistical and econometric software. I close the section with a discussion of GeoDa and OpenGeoDa in which an attempt is made to combine both exploratory and confirmatory viewpoints.

2.1. ESDA, Geovisualization and Visual Analytics

In Anselin (1999, p. 258), I describe exploratory spatial data analysis as "a collection of techniques to describe and visualize spatial distributions, identify atypical locations or spatial outliers, discover patterns of spatial association, clusters or hot spots, and suggest spatial regimes or other forms of spatial heterogeneity." With roots in statistics and exploratory data analysis (EDA, Tukey 1977), ESDA can be conceived of as a superset of EDA techniques with a special focus on location, distance and spatial interaction. An essential characteristic of software for ESDA is therefore the possibility to interact with the data through graphical interfaces. Any graphical representation or summary of the data, such as a chart, a table, a graph or a map is considered a "view" of the data (Buja et al. 1996). The analyst interacts with the data by selecting observations, linking different

views, zooming in or focusing on subsets, changing selected observations, rotating views, etc.

A second disciplinary perspective is offered by cartography, where the traditional role of a map as a way to present research outcomes has similarly been extended to encompass the concept of geovisualization, which can been considered to be a special form of data exploration through the use of the map. For example, DiBiase (1990) outlines the geovisualization research process as a move from visual thinking (private realm) to visual communication (public realm), proceeding through the stages of exploration, confirmation, synthesis and presentation (see also MacEachren 1995). In the first decade of the 21st century, this led to considerable research on the design of new graphical tools to facilitate visualization and on ways to go beyond the static map and provide support for geographical data exploration (e.g., Kraak and MacEachren 2005, Rhyne et al. 2006).

A third perspective is grounded in computer science, where the need to handle massive data sets, particularly driven by security and surveillance concerns after the establishment of the Department of Homeland Security in the U.S. stimulated research on pattern recognition, data mining and their incorporation in visual tools. The integration of these new software tools with the inherent analytical capabilities of humans led to the coining of a new interdisciplinary field of "visual analytics" (Thomas and Cook 2005). As in ESDA and geovisualization, the emphasis is on the human-computer interaction and on designing effective methods and tools to facilitate analytical reasoning in order to "detect the expected and discover the unexpected™" (Kielman et al. 2009, p. 245). When the interest focuses on the unique characteristics of space and spatial data, the terms

geovisual analytics or geospatial visual analytics have been introduced. This has led to a vibrant research agenda dealing with space-time dynamics, movement and spatial decision support (e.g., Andrienko et al. 2007, 2010a,b, 2011).

The research efforts associated with ESDA, geovisualization and visual analytics have yielded a plethora of software tools, some still under active development, and several of which were also deployed over the internet (see Section 4). Almost all of these tools originated in the non-commercial sector and many were open source (see also Section 3). Besides GeoDa (see Section 2.3), an illustrative example of the evolution of exploratory software tools during this period are the various toolkits developed as part of the GeoVISTA project (Gahegan et al. 2002, 2008). The initial software consisted of a series of linked maps and graphs, such as a parallel coordinate plot. The particular architecture of GeoVISTA Studio was highly modular, in the form of software components that could be readily mixed and matched for customized applications, later also deployed on the web. GeoVISTA Studio has seen considerable application in the fields of public health and epidemiology. More recent descendants of this effort consist of toolkits for the exploration of space-time and multivariate association, such as the Visual Inquiry Toolkit (Chen et al. 2008) and the Geoviz toolkit (Hardisty and Robinson 2011).

Similar efforts, but constructed in an existing software environments are the ESDA tools GeoXP (Laurent et al 2009) and Arc_Mat (Liu and LeSage 2010). These packages are similar in design to GeoDa, in that they combine both exploratory techniques as well as some spatial statistics and spatial regression. GeoXP consists of linked windows implemented using the open source R platform (an earlier version was developed using MatLab). It is distinguished by the incorporation of some specialized

graphs, such as a Lorenz curve and by a link to the spatial statistical functionality of the R spdep package (see Section 3). The Arc_Mat toolbox is very similar in design and spirit, but is built using the MatLab software, including the spatial econometric tools from the LeSage-Pace library (e.g., LeSage and Pace 2009).

The tools described so far primarily deal with cross-sectional data. Examples of a special focus on space-time dynamics are STARS (Rey and Janikas 2006), written in the open source Python scripting language, and LISTA-Viz (Hardisty and Kippel 2010), part of the GeoVISTA family of tools. This continues to be an active area of research and development.

2.2 Spatial Statistics and Spatial Econometrics

Spatial statistics and spatial econometrics saw tremendous methodological advances during the first decade of the 21st century (for a recent review, see Anselin 2010). Three developments in particular stand out. In spatial statistics, the Bayesian approach became by far the dominant paradigm, in which forms of spatial dependence and spatial heterogeneity are embedded in hierarchical model specifications (e.g., Banerjee et al. 2004, Cressie and Wikle 2011). In spatial econometrics, there is increased attention to the computational and numerical requirements necessary to implement maximum likelihood estimation for spatial regression models with a large number of observations (e.g., Smirnov and Anselin 2001, 2009, Pace and LeSage 2004, 2009, Smirnov 2005). These new methods allow much larger data sets to be analyzed than was possible with the classic eigenvalue-based method (Ord, 1975), e.g., as implemented in SpaceStat. In addition, alternatives to the maximum likelihood approach were developed, based on the general method of moments (GMM) and semi-parametric techniques (e.g., Kelejian and

Prucha 2007, 2010, Lee 2007). The new methods quickly found their way into specialized spatial analysis software.

The Bayesian methods rely on the ability to specify the hierarchical model structures in a flexible manner as well as the efficient implementation of Markov Chain Monte Carlo (MCMC) simulations. Up until the BUGS project (Bayesian inference Using Gibbs Sampling) was launched, there was no comprehensive software package to implement this. Especially since its implementation for the Windows operating system in the form of the freestanding WinBugs package and its spatial component GeoBugs (later integrated into the main WinBugs), the software has become the main tool to implement Bayesian spatial statistical analysis (Lunn et al. 2009). In addition to the Windows version, there is an active OpenBugs open source development and many routines have been written to interface WinBugs (or Bugs) with other software, such as R. Also, Bayesian spatial hierarchical modeling is under active development in the R spBayes package.

On the spatial econometric side, the MatLab-based spatial econometrics toolbox developed by LeSage, Pace and co-workers (http://spatialeconometrics.com) contains Bayesian routines for the estimation of most spatial regression specifications, including those for limited dependent variables (probit and tobit). In addition to the Bayesian techniques, this toolbox also contains a number of the large data set methods for maximum likelihood estimation of spatial regression models (based on the sparse matrix implementation in MatLab). Sparse matrix techniques from maximum likelihood estimation are also implemented in the routines contained in Bivand's spdep package (Bivand 2006, see Section 3) and in GeoDa/OpenGeoDa (see Section 2.3).

Up until recently, software to carry out the new GMM estimation methods was not readily available. A partial implementation had been added to spdep by 2006, and routines were used for individual researchers' analyses, but a broad adoption of these techniques was largely stymied due to the lack of software. Recent additions have remedied this situation. Programmed for the open source R platform, a new package sphet (Piras 2010) contains all the latest GMM estimation methods as well as the spatial HAC. The GMM methods were also recently implemented in the commercial Stata software (Drukker et al 2011), which represents the first time spatial econometric methods are adopted by a mainstream econometric software package. In addition, PySAL release 1.3 of the open source Python library for spatial data analysis (see Section 3.3) contains the full range of all GMM methods in the updated spreg module.

In addition to software that followed the new methodological and theoretical developments, a number of niche packages gained wide acceptance in the empirical practice of specific subfields, such as public health, epidemiology and criminology. CrimeStat (Levine 2006, 2010) deals with the analysis of crime incident locations and contains spatial distribution statistics, rate smoothing methods and techniques to estimate crime travel demand and journey to crime. SaTScan (Kulldorff 2011) has become the de facto standard for cluster detection in epidemiology and public health. It implements Kulldorff"s scan statistic and its extensions (e.g., Takahashi et al. 2008) for a range of statistical models. Finally, the GWR software distributed by the Irish National Centre for Geocomputation implements the geographically weighted regression method of Fotheringham and co-workers for several regression specifications (Fotheringham et al. 2002). It has an open source counterpart in the form of the gwr package in the R software

environment. In addition, GWR for ordinary least squares regression has been included in the spatial statistical toolbox of ESRI's ArcGIS starting with version 9.3 (2009).

2.3 GeoDa and OpenGeoDa

GeoDa was conceived as an alternative to existing software toolboxes to provide an easy and intuitive path from geo-visualization, through exploratory spatial data analysis and the study of spatial correlation as descriptive statistics to the specification and estimation of spatial regression models. It was primarily aimed at a social science audience interested in "spatial questions" and at users who were not familiar with or did not have access to a GIS. The origins of the idea go back to the SpaceStat integration with Arcview (Anselin and Smirnov 1997), but the limitations due to the lack flexibility from the linkage with ArcView dictated a different approach. GeoDa was built using ESRI's MapObjects for all mapping and query functions and written in C++ for the Microsoft Windows operating system. The package therefore did not require a GIS, but used industry standard ESRI shape files as the main data format.

At the time of its initial launch in 2002, GeoDa was unique in implementing full dynamic linking and brushing between all maps and charts in the user interface. It also included an eclectic combination of methods that were not available in other packages at the time, such as Lagrange Multiplier tests for spatial autocorrelation in regression models, maps for clusters and spatial outliers derived from the Local Moran statistic, efficient treatment of spatial weights and efficient maximum likelihood estimation of large spatial regression models (see Anselin et al. 2006, for details). The package was free and quickly gained broad acceptance worldwide. To date (fall 2011), there have been more than 63,000 uniquely identified downloads. In addition, GeoDa has been employed

in over 500 published empirical articles and working papers.

It quickly became clear that the reliance on the legacy ESRI MapObjects code and some other closed source libraries could become a constraint. The ESRI libraries only worked on Windows XP and never migrated to the more recent Windows platforms. Overall, the architecture and design of the program was tied to old technology and prevented the adoption of the software in other operating systems.

In 2005, a decision was made by the developers to port the code to a cross-platform infrastructure and to eventually open source it. The new project was referred to as OpenGeoDa. With a few exceptions (most notably the addition of the Gi and Gi* statistics), the core functionality of OpenGeoDa is the same as that of Legacy GeoDa. However, its design "under the hood" is totally different. OpenGeoDa is modular and extensible and refactored to the latest software standards. It is completely written in C++ and uses standard open source libraries such as the C++ Standard Template Library (STL), wxWidgets for the cross-platform GUI library (http://www.wxwidgets.org), and the Boost C++ libraries for computational geometry and some mathematical operations (http://www.boost.org). The program runs on all operating systems supported by wxWidgets. Since early 2009, frequently updated binaries of a beta version have been released that run on all Windows platforms, Mac OS X and Linux. OpenGeoDa Version 1.0 was officially released in October 2011 under the open source GPL 3.0 license.

It is hoped that the modular structure of the open source OpenGeoDa will attract developers from the larger community interested in adding specialized functionality. In addition to extending functionality, the core development team is also focusing on making the software suitable for use in High Performance Computing (HPC)

environments, such as compute clusters, the cloud and the grid (see Section 4).

3. The Open Source Movement

A major factor in facilitating the transition from methodological innovation to software code has been the rapid growth of the open source community (Rey 2009). Open source has effectively removed the quasi monopoly on GIS functionality held by a handful of commercial vendors. To date, it has yielded a large array of free software tools that can serve as building blocks for more advanced analysis. In this section, I briefly review some salient aspects of the open source GIS movement and open source software for spatial data analysis. I close with a discussion of the PySAL library for spatial analysis, currently under development under the auspices of the GeoDa Center.

3.1. Open Source GIS

Many of the analytical tools developed pre-2000 had to rely on a commercial GIS for data storage, mapping and visualization. Paralleling the evolution in the larger software development community (e.g., Raymond 1999), the open source movement gained increased acceptance among GIS developers as well. To date, hundreds of active projects are in the process of being carried out (for recent reviews, see Hall and Leahy 2008, Steinger and Bocher 2009). As a result, several open source desktop GIS environments currently exist as viable alternatives to the commercial options, and they are increasingly adopted for teaching and research. One of the oldest such platforms is the GRASS GIS, which has its origins as closed source software, but converted to an open source license in the late 1990s (Neteler and Mitasova 2008). Other examples include Quantum GIS, SAGA, gvSIG and uDig. An extensive listing can be found on the web site of the Open Source Geospatial Foundation (http://www.osgeo.org) and at http://opensourcegis.org.

An exhaustive review of open source GIS is beyond the current scope, but three important characteristics can be highlighted that greatly facilitated the development of open source spatial analytical software. First is the existence of extensive open source libraries for spatial data handling that serve as foundational building blocks for other software. Examples include the geometry engine encompassed in the JTS Topology Suite (formerly known as the Java Topology Suite) and its C++ implementation in the GEOS library, which implement many standard (non-statistical) spatial analytical operations (such as intersection and overlay), the GDAL/OGR Geospatial Data Abstraction Library, a cross platform C++ translator library that covers most existing vector and raster geospatial data formats (http://www.gdal.org), and the GeoTools java GIS toolkit (Turton 2008). These libraries encompass the core functionality to deal with the computational geometry aspects behind a GIS. They also greatly facilitate interoperability by removing barriers to access different data formats. In addition, the libraries implement industry standards for geographic content and functionality promoted by the Open Geospatial Consortium (http://www.opengeospatial.org), thereby greatly facilitating the adoption and enforcement of these standards.

A second important aspect has been the spatial indexing of open source industry standard relational databases, such as PostGIS for the PostgreSQL object-relational database. This allows spatial analytical software to take advantage of the built-in query and computational functionality in the relational database rather than having to replicate such functions (usually in a less efficient manner). It also greatly facilitated the growth of web mapping and internet GIS, a large and growing share of which is build on open source software.

A final aspect of open source GIS that facilitates the extension of the software with spatial analytical capability is the implementation of an open architecture and well documented application programming interfaces (API). This allows for extension of the functionality and the addition of specialized operations by means of plug-ins or close coupling with other software (e.g., between Quantum GIS and GRASS). Again, this avoids the need to "reinvent the wheel" and allows analysts to leverage the GIS functionality and focus on the implementation of specialized methodological advances.

3.2 Open Source Spatial Data Analysis

By and large, the open source GIS initiatives remain limited in terms of spatial data analytical functionality (Rey 2009). For example, in the review of GIS functionality in ten leading open source desktop GIS in Steinger and Bocher (2009), spatial statistical analysis is absent. The development of such functionality has been primarily driven by the statistics community through the widespread adoption of the R programming environment for statistics and graphics (Ihaka and Gentleman 1996). While open source econometric software is also gaining increased acceptance (Yalta and Yalta 2010), to date it is still totally without any spatial functionality.

As illustrated in Bivand et al. (2008), there now exists a wide array of specialized "packages" implemented in the R language to deal with various aspects of spatial statistics. The most familiar among these is Bivand's spdep package, which was the first to implement spatial autocorrelation statistics and spatial regression methods for lattice data. As pointed out earlier, several packages existed for the analysis of point patterns and to carry out geostatistical analysis in the S-Plus language (e.g., Venables and Ripley 1994) and these were quickly ported to the R environment. In addition to spatial

autocorrelation analysis in spdep, other packages deal with cluster detection and an array of specialized methods, such as geographically weighted regression (gwr), Bayesian spatial hierarchical modeling (spBayes) and advanced spatial econometric methods (sphet), discussed above.

In addition to spatial statistical functionality, R packages have been developed to deal with fundamental geospatial data structures, such as the r-spatial foundation classes that build upon the GDAL/OGR libraries. Together with the graphics capabilities in R, this allows the development of graphical and interactive interfaces to the statistical functionality, as illustrated by the GeoXP package mentioned in Section 2. Also, there are several links between R and open source GIS, such as GRASS and SAGA. The development of open source spatial statistical functions in R continues unabated.

3.3. PySAL

PySAL is a project to develop an open source cross platform modular library of spatial analytical functions written in the Python scripting language (http://pysal.org). As outlined in Rey and Anselin (2006), it grew out of the software development activities at the Spatial Analysis Laboratory of the University of Illinois (Anselin) and the STARS developer group at San Diego State University (Rey). The effort is now based in the GeoDa Center at Arizona State University. The official release of PySAL 1.0 occurred on August 1, 2010. The program follows a six month release cycle, with PySAL 1.3 slated to be available on January 31, 2012.

PySAL is highly modular and implemented as pure Python, avoiding any dependencies on other languages or non-Python libraries. It leverages the wealth of existing code available for the Python language, such as the work by the GIS and Python

lab (http://gispython.org) and the scientific Python community (http://www.scipy.org). The latter in particular, with origins in astrophysics and large-scale climate modeling, has yielded efficient code for linear algebra and numerical computing, taking advantage of sparse matrix data structures to allow the manipulation of very large data sets. This avoids some of the constraints on data set size and data structures inherent in the R environment.

PySAL is conceived as a library and is therefore aimed at programmers or end users who are comfortable with a command line environment (similar to R). It is completely open source under the new BSD license, with all code posted on google code (http:// code.google.com/p/pysal). There is no graphical user interface, nor an environment that groups functionality together. The code is fully object oriented and highly modular, organized around functional requirements. Besides a core (including, among others, the abstraction of file input/output) and a computational geometry module, it currently includes specialized functionality for exploratory spatial data analysis (spatial autocorrelation statistics and rate smoothing), spatial dynamics (space-time analysis), regionalization algorithms, spatial inequality measures, spatial weights manipulation and spatial regression. As mentioned before, Version 1.3 includes the full range of advanced spatial econometric estimators based on the general method of moments.

The PySAL library is viewed as the foundational framework to deliver spatial analytical functionality in many different forms. For example, the spatial econometric routines are also encompassed behind a graphical user interface as the GeoDaSpace package. Similarly, some of the spatial weights functionality forms the basis for the delivery of spatial analytical web services and web applications. PySAL functionality can

readily be added as a plug-in to existing GIS software such as the ArcGIS spatial analytical toolbox, GRASS or Quantum GIS. Specific functionality can be wrapped in a user interface targeted to specialized applications, such as crime cluster tracking, and delivered on a range of output devices, including iPads, iPhones and similar smart phones and tablets.

4. Cyberinfrastructure

As Goodchild (2010) points out, the explosive expansion of the internet was a major factor unanticipated in the discussions of spatial analysis and GIS in the early 1990s. With hindsight it is clear that the advent of the internet significantly affected the evolution of GIS and spatial analytical applications. In this Section, I start by briefly reviewing some recent developments in web GIS as well as the migration of spatial analytical capability to the web in the form of spatial analytical web services. I next discuss the notion of a cyberinfrastructure in general and cyberGIS in particular. I close with an outline of a vision for a spatial econometrics workbench as part of a cyberinfrastructure for spatial data analysis.

4.1. Web GIS

The initial efforts to extend GIS with spatial analytical capabilities were focused on desktop GIS systems. However, by the mid 1990s, the internet had become increasingly more pervasive and browsers were sufficiently powerful that a number of early attempts at delivering maps over the web had been developed (Plewe 1997, Dragicevic 2004). This very rapidly evolved into a new way of delivering geographic information, especially after it was embraced by commercial vendors. Products such as ESRI's ArcIMS and its open source counterparts (e.g., MapServer) became increasingly adopted

and allowed for the integration of maps (and a limited number of map operations) with other web content. Given the scope of this essay, it is impossible to provide a comprehensive review of the evolution of web GIS. However, it is useful to highlight a number of important characteristics that provided the groundwork and context for the notion of cyberinfrastructure, to which I turn in the next section.

Web mapping and web GIS fundamentally changed the way geographic information could be accessed. After largely being constrained by proprietary formats used in commercial desktop GIS, geographic information became distributed and widely accessible to anyone with a browser (Peng and Tsou 2003). In several countries, extensive efforts at establishing a spatial data infrastructure led to sophisticated data portals that provide digital information in georeferenced form (Tait 2005, Goodchild et al. 2007). The distributed and unrestricted nature of geospatial data also enabled the advent of participatory GIS, allowing ready access to information as well as ways for the public to intervene in spatial decision processes (e.g., Caldeweyher et al. 2006). Google maps (and later Bing maps) further revolutionized the way information could be accessed, mapped and annotated. In the current era of the Web 2.0, interaction with information has become the norm and many elementary GIS operations have become available to the general public, leading to the notion of volunteered geographic information (Goodchild 2007, Batty et al. 2010). Web atlases have proliferated (e.g., MacEachren et al. 2008) and in a small number of cases have begun to include limited spatial analytical functionality (e.g., Anselin et al. 2004, Tiwari and Rushton 2010).

Two other important aspects of internet GIS are the move towards a service oriented architecture (SOA) in software delivery and the creation and wide adoption of

open standards for the associated web services. The growing attention towards services imply that the geospatial software no longer needs to be delivered on a desktop, but can be running on powerful servers accessed over the internet. This access is either by the user, through a web browser (as a web application) or directly at the software level (machine to machine interaction). In order to make such a framework operational, a rich set of standards has evolved through the auspices of the open geospatial consortium (OGC), providing a common structure for data provision (e.g., web feature standard or WFS) and geoprocessing (e.g., web processing standard or WPS) to enable distributed internet geographic information systems and web services (e.g., Peng and Zhang 2004, Michaelis and Ames 2009, Yue et al. 2010, Li et al. 2011a, Dragicevic et al. 2011). In addition, considerable research attention has been devoted to semantic annotation of geospatial data and operations, resulting in the so-called geospatial semantic web (e.g., Scharl and Tochterman 2007, Wiegand et al. 2010). While much of the attention in the web GIS literature so far has focused on data delivery and basic geoprocessing functions, the extension of these concepts to spatial analytical operations is beginning to gain traction, and a number of prototype implementations are starting to appear (e.g., Li et al. 2011b).

4.2 Cyberinfrastructure and CyberGIS

In the so-called "Atkins Report" (NSF 2003), a much-cited blue-ribbon report to the National Science Foundation, the need to establish a cyberinfrastructure was invoked to support the nation's scientific advances required to serve the emerging knowledge economy. This argument was situated in a broader context in which the two traditional approaches to scientific research, i.e., theoretical and experimental/observational are

extended with a third branch consisting of computation (modeling and simulation). The cyberinfrastructure, referred to as e-science in the U.K., integrates "enabling hardware, algorithms, software, communications, institutions and personnel" (NSF 2003, p. 5). In addition to the enabling technology that provides for high performance computing and access to distributed data and sensor information, other important aspects of the cyberinfrastructure are visualization and data analysis and collaborative networks (see also, Goodchild 2010). In other words, cyberinfrastructure is an integrated system designed to support the solution of complex scientific problems in a collaborative fashion.

While often associated with the physical sciences, e.g., with applications to astrophysics, earthquake and global climate change modeling, the importance of cyberinfrastructure to support efforts in the social sciences and humanities should not be understated (e.g., ACLS 2006). Similarly, given the importance of location (in both space and time) in so many scientific domains, it is not surprising that cyberinfrastructure has been referred to as a potential driving force to support advances in the geospatial sciences as well. Alternatively termed spatial cyberinfrastructure (Wright and Wang 2011), geospatial cyberinfrastructure (Yang et al. 2010), or cyberGIS (Wang 2010), the "infrastructure" consists of a combination of distributed spatial data (e.g., data repositories, spatial data infrastructures as well as evolving sensor networks) and distributed geoprocessing (e.g., data manipulation, geovisualization, pattern detection, process modeling) as well as the software systems (spatial middleware, or "glue") needed to allow seamless integration between these resources. A growing number of applications take advantage of and contribute to an emerging geospatial cyberinfrastructure (for a

recent review, see Yang et al. 2010), although they tend to emphasize the development of technical solutions to carry out distributed geoprocessing (Yang et al. 2008, Yang and Raskin 2009), or to take advantage of the high performance computation capacity provided by grid networks of supercomputers (Wang and Armstrong 2009, Wang and Liu 2009, Wang et al. 2009) or cloud computing infrastructure (Yang et al. 2011).

In terms of spatial data analytical applications, the integration with a geospatial cyberinfrastructure is still in its infancy. Some early applications include the estimation of hierarchical Bayesian space-time models (Yan et al. 2007), the computation of a local spatial autocorrelation coefficient (Wang et al. 2007) and spatial interpolation (Wang 2010) implemented using the GISolve middleware to access the U.S. TeraGrid. A comparable illustration using the U.K. computational grid resources is the implementation of geographically weighted regression in Harris et al. (2010). Overall, these applications tend to be prototypes, focused on illustrating the use of the high performance computational resources. They do not yet constitute a seamless infrastructure to support advanced spatial data analysis.

4.3. Towards a Spatial Econometrics Workbench

In this final section devoted to cyberinfrastructure, I outline some ongoing efforts at ASU's GeoDa Center to develop a "spatial econometrics workbench," as a platform to support spatial analytical cyberinfrastructure. The workbench is built on the functionality contained in the open source PySAL library (Section 3.3), but it is delivered in the form of spatial analytical web services and web applications through the addition of specialized middleware.

The goal behind the spatial econometrics workbench is to design a platform that

allows users to seamlessly link distributed data sources to carry out a wide range of spatial data analytical functions through a web interface, without having to run any specialized software on their desktop. In addition to the familiar spatial autocorrelation and spatial regression analyses, the workbench also includes functionality to carry out a range of simulation experiments using synthetic data that reflect different types of spatial structure (spatial dependence and spatial heterogeneity). All the computing is carried out on a server, either in the form of a single server (with limited scalability), a compute cluster, or using the full high performance computing resources of the TeraGrid or cloud environments. In addition to a traditional user-interface as a web application, the web services can also be accessed directly by code operating on other computers.

Moving functionality from a library to a web service is not simply a matter of adding the appropriate middleware to implement the communication over the internet between the various pieces of code. Careful choices need to be made to define meaningful "atomic" functionality in such a way that more complex analyses can be composed out of these atomic services. For example, to compute a local spatial autocorrelation statistic, one would need not only the functionality to calculate the statistic and its significance, but there must be a mechanism to load the data, to identify, load and/or construct an appropriate spatial weights matrix and to visualize the results. Users or computer programs must be able to identify the proper code components and load and combine them in the right way. This is currently implemented in a prototype local spatial autocorrelation web application that runs on the Amazon Elastic Compute Cloud (EC2) service.

While a full technical discussion is beyond the scope of this essay, it may be

useful to point out a number of important challenges that remain to be addressed in order to move the current framework beyond the prototype stage. A major impediment consists of the computational bottlenecks encountered when moving data around on the internet and when carrying out spatial computations for very large data sets (e.g., the conditional permutations required to compute a pseudo-significance for the local spatial autocorrelation statistic). Especially when the intent is to deliver support for an interactive and highly visual decision support system, it is essential that all results be transmitted to the user in a span of a few seconds. In the current infrastructure, this is still a major challenge and further research is needed on the development of improved algorithms, efficient parallelization of the code and effective use of high performance computing infrastructure.

A second concern pertains to the description of the functionality required to allow automatic (i.e., by other programs) detection of the proper spatial analytical methods. This requires a form of semantic annotation to describe the range of spatial models and methods so that they can be detected by other programs (not humans). Commonly used ontology based approaches quickly break down and fail to deal with the almost combinatorial complexity of assumptions, spatial topologies and estimation methods. The web processing standards (WPS) developed by the OGC are still too generic to be able to cope effectively with the demands of a portfolio of analytical techniques. Initial efforts to develop "metadata" for spatial analytical methods show promise, but much remains to be done.

A final concern pertains to human capital. The democratization of GIS and spatial analysis spawned by web GIS and potentially by a geospatial cyberinfrastructure

implicitly assumes that the "user" will be able to correctly identify the technique appropriate to address the question at hand. Combined with the uncertainty inherent in data sources generated by volunteered geographic information, this constitutes a challenge for the educational community. It is of critical importance that the powerful geospatial tools encompassed in a cyberGIS infrastructure be accompanied by the proper guidance to ensure that the methods and data are used for the purposes for which they were designed. All too often, this remains an overlooked aspect.

5. Conclusions

The landscape of software solutions available to support spatial data analysis has changed dramatically since the early ventures in the 1980s and 1990s. Most technological barriers have been removed so that any new method can conceivably be implemented in a number of different open source toolboxes, integrated as a plug-in with popular GIS and delivered over the internet as a web application.

While tremendous progress has been made, a number of important challenges are worth mentioning. First, several computational roadblocks remain, both to be able to efficiently carry out spatial data analysis of large data sets on the desktop as well as to deliver effective analytical capability over the internet. The types of ever larger spatial and space-time data sets that are available to researchers require refined algorithms in addition to high-end hardware to allow computations to be carried out in a reasonable time. Some problems are still too large to be handled without the help of high performance computing. This creates challenges for an effective implementation of spatial decision support systems, where computations need to be delivered on the fly to allow scenario evaluation and other decision processes. Similarly, an effective

visualization of very large sets of data points remains a challenge.

Much work remains to be done to achieve true interoperability where machines are able to recognize and select the proper tools based on their metadata. As mentioned earlier, standards for effective metadata for spatial analytical methods and models have yet to be developed.

A less technological challenge pertains to the culture of research and its reward system, especially in the social sciences. Much of the design of a cyberinfrastructure is geared toward enabling and facilitating collaborative research, through so-called virtual communities of scholars. However, social science still very much follows a lone investigator paradigm. Also, the academic reward structure is not (yet) geared to evaluating collaborative efforts relative to individual ones, or the authorship of software code versus authorship of refereed articles.

Finally, an important challenge relates to education. The skills needed to effectively transfer methodological advances to software code require a combination of "geographical" insights and computer science. Very few academic programs (if any) prepare analysts with this combined skill set.

In this review essay, I have attempted to bring out some salient features that characterize the evolution of spatial data analytical software during the past decades. I hope it can provide an inspiration to a next generation of developers.

Acknowledgments

This work was supported in part by grants from the U.S. National Science Foundation (OCI-1047916) and the U.S. National Cancer Institute (1R01CA126858-01A1). The content is solely the responsibility of the author and does not necessarily represent the official views of the National Science Foundation, the National Cancer Institute, or the National Institutes of Health. Earlier versions were presented at a Fellows Colloquium, Research Triangle Institute International, Research Triangle Park, NC, at Geoinformatics 2010, Beijing, China, the NSF Teragrid Workshop on CyberGIS, Washington, DC, and the International Workshop on Cities, Globalization and Development, Toluca, Mexico. Many thanks to my colleagues Serge Rey, Myunghwa Hwang, Mark McCann and Julia Koschinsky for many stimulating conversations on this topic.

References

Abler, R. (1987). The National Science Foundation National Center for Geographic Information and Analysis. International Journal of Geographical Information Systems 1, 303-326.

ACLS (2006). Our Cultural Commonwealth. The Report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences. Washington, DC, American Council of Learned Societies.

Andrienko, G. and N. Andrienko (1999). Interactive maps for visual data exploration. International Journal of Geographical Information Science 13, 355-374.

Andrienko, G., N. Andrienko, P. Jankowski, D. Keim, M-J Kraak, A. MacEachren and S. Wrobel (2007). Geovisual analytics for spatial decision support: setting the research agenda. International Journal of Geographical Information Science 21,

839-857.

Andrienko, G., N. Andrienko, J. Dykes, M-J Kraak and H. Schumann (2010a). GeoVA(t) – Geospatial visual analytics: focus on time. International Journal of Geographical Information Science 24, 1453-1457.

Andrienko, G., N. Andrienko, U. Demsar, D. Dransch, J. Dykes, S. Fabrikant, M. Jern, M-J Kraak, H. Schumann and C. Tominski (2010b). Space, time and visual analytics. International Journal of Geographical Information Science 24, 1577-1600.

Andrienko, G., N. Andrienko, D. Keim, A. MacEachren and S. Wrobel (2011). Challenging problems of geospatial visual analytics. Journal of Visual Languages and Computing 22, 251-256.

Anselin, L. (1988). Spatial Econometrics, Methods and Models. Dordrecht, Kluwer Academic.

Anselin, L. (1989). Spatial Regression Analysis on the PC: Spatial Econometrics Using GAUSS. Department of Geography, University of California, Santa Barbara, CA.

Anselin, L. (1992). SpaceStatTutorial: A Workbook for Using SpaceStat in the Analysis of Spatial Data, Technical Sofware Series S-92-1. Santa Barbara, CA, National Center for Geographic Information and Analysis.

Anselin, L. (1995a). Local indicators of spatial association – LISA. Geographical Analysis 27, 93-115.

Anselin, L. (1995b). SpaceStat Version 1.80 User's Guide. Regional Research Institute, West Virginia University, Morgantown, WV.

Anselin, L. (2000). Computing environments for spatial data analysis. Journal of

Geographical Systems 2, 201-220.

Anselin, L. (2005). Spatial statistical modeling in a GIS environment. In D. Maguire, M. Batty and M. Goodchild (eds.) GIS, Spatial Analysis and Modeling, pp. 93-111. Redlands, CA, ESRI Press.

Anselin, L. (2010). Thirty years of spatial econometrics. Papers in Regional Science 89, 2-25.

Anselin, L. and S. Bao (1997). Exploratory spatial data analysis: linking SpaceStat and Arcview. In M. Fischer and A. Getis (eds.), Recent Developments in Spatial Analysis, pp. 35-59. Berlin, Springer-Verlag.

Anselin, L. and O. Smirnov (1997). SpaceStat Extension for ArcView 3.0. Regional Research Institute, West Virginia University, Morgantown, WV.

Anselin, L. and A. Getis (1992). Spatial statistical analysis and geographic information systems. The Annals of Regional Science 26, 19-33.

Anselin, L. and S. Hudak (1992). Spatial econometrics in practice: a review of software options. Regional Science and Urban Economics 22, 509-536.

Anselin, L., Y-W Kim and I. Syabri (2004). Web-based analytical tools for the exploration of spatial data. Journal of Geographical Systems 6, 197-218.

Anselin, L., I. Syabri and Y. Kho (2006). Geoda, an introduction to spatial data analysis. Geographical Analysis 38, 5-22.

Bailey, T. and A. Gatrell (1995). Interactive Spatial Data Analysis. New York, NY, Wiley.

Banerjee, S., B. Carlin and A. Gelfand (2004). Hierarchical Modeling and Analysis for Spatial Data. Boca Raton, FL, Chapman and Hall/CRC.

Bao, S. and D. Martin (1997). User's Reference for the S+ArcView Link. Seattle, WA, Mathsoft Inc.

Bao, S., L. Anselin, D. Martin and D. Stralberg (2000). Seamless integration of spatial statistics and GIS: the S-Plus for ArcView and the S+Grassland links. Journal of Geographical Systems 2, 287-306.

Batty, M., A. Hudson-Smith, R. Milton and A. Crooks (2010). Map mashups, Web 2.0 and the GIS revolution. Annals of GIS 16, 1-13.

Becker, R., W. Cleveland and A. Wilks (1987). Dynamic graphics for data analysis. Statistical Science 2, 355-395.

Bivand, R. (2006). Implementing spatial data analysis software in R. Geographical Analysis 38, 23-40.

Bivand, R. and A. Gebhardt (2000). Implementing functions for spatial statistical analysis using the R language. Journal of Geographical Systems 2, 307-317.

Bivand, R., E. Pebesma and V. Gomez-Rubio (2008). Applied Spatial Data Analysis with R. New York, NY, Springer.

Bradley, R. and J. Haslett (1992). High interaction diagnostics for geostatisticial models of spatial referenced data. The Statistician 41, 371-380.

Brunsdon, C. (1998). Exploratory spatial data analysis and local indicators of spatial association with XLISP-STAT. The Statistician 47, 471-484.

Buja, A., D. Cook and D. Swayne (1996). Interactive high dimensional data visualization. Journal of Computational and Graphical Statistics 5, 78-99.

Caldeweyher, D., J. Zhang and B. Pham (2006). OpenCIS – Open source GIS-based web community information system. International Journal of Geographical

Information Systems 20, 885-898.

Chen, J., A. MacEachren, and D. Guo (2008). Supporting the process of exploring and interpreting space-time, multivariate patterns: the Visual Inquiry Toolkit. Cartography and Geographic Information Science 35, 33-50.

Cleveland, W. and M. McGill (1988). Dynamic Graphics for Statistics. Pacific Grove, CA, Wadsworth.

Cressie, N. (1993). Statistics for Spatial Data. New York, Wiley.

Cressie, N. and C. Wikle (2011). Statistics for Spatio-Temporal Data. Hoboken, NJ, John Wiley.

DiBiase, D. (1990). Visualization in the earth sciences. Earth and Mineral Sciences Bulletin 59, 13-18.

Ding, Y. and S. Fotheringham (1992). The integration of spatial analysis and GIS. Computers, Environment and Urban Systems 16, 3-19.

Dragicevic, S. (2004). The potential of web-based GIS. Journal of Geographical Systems 6, 79-81.

Dragicevic, S., S. Li, M. Brovelli and B. Veenendaal (2011). Pervasive web mapping, geoprocessing and services. Transactions in GIS 15, 125-127.

Drukker, D., I. Prucha and R. Raciborski (2011). A command for estimating spatial-autoregressive models with spatial-autoregressive disturbances and additional endogenous variables. Working Paper, Department of Economics, University of Maryland. College Park, MD.

Dykes, J. (1997). Exploring spatial data representation with dynamic graphics. Computers and Geosciences 23, 345-370.

Dykes, J. (1998). Cartographic visualization: exploratory spatial data analysis with local indicators of spatial association using Tcl/Tk and cdv. The Statistician 47, 485-497.

Farley, J., W. Limp and J. Lockhart (1990). The archeologist's workbench: Integrating GIS, remote sensing, EDA and database management. In K. Allen, F. Green and E. Zubrow (eds.), Interpreting Space: GIS and Archeology, pp. 141-164. London, Taylor and Francis.

Fotheringham, S., C. Brunsdon and M. Charlton (2002). Geographically Weighted Regression. Chichester, John Wiley.

Gahegan, M., M. Takatsuka, M. Wheeler and F. Hardisty (2002). Introducing GeoVISTA Studio: An integrated suite of visualization and computational methods for exploration and knowledge construction in geography. Computers, Environment and Urban Systems 26, 267-292.

Gahegan, M., F. Hardisty, U. Demsar and M. Takatsuka (2008). GeoVISTA Studio: reusability by design. In B. Hall and M. Leahy (eds.), Open Source Approaches to Spatial Data Handling, pp. 201-220. Berlin, Springer-Verlag.

Goodchild, M. (1987). A spatial analytical perspective on geographical information systems. International Journal of Geographical Information Systems 1, 31-45.

Goodchild, M. (2007). Citizens as sensors: the world of volunteered geography. Geojournal 69, 211-221.

Goodchild, M. (2010). Whose hand on the tiller? Revisitiing "Spatial Statistical Analysis and GIS." In L. Anselin and S. Rey (eds.), Perspectives on Spatial Analysis, pp. 49-59. Heidelberg, Springer-Verlag.

Goodchild, M., R. Haining, S. Wise and others (1992). Integrating GIS and spatial analysis – problems and possibilities. International Journal of Geographical Information Systems 6, 31-45.

Goodchild, M., L. Anselin, R. Appelbaum and B. Harthorn (2000). Towards spatially integrated social science. International Regional Science Review 23, 139-159.

Goodchild, M., P. Fu and P. Rich (2007). Sharing geographic information: an assessment of the Geospatial One-Stop. Annals of the Association of American Geographers 97, 249-265.

Haining, R. (1989). Geography and spatial statistics: current positions, future developments. In B. Macmillan (ed.), Remodeling Geography, pp. 191-203. Oxford, Basil Blackwell.

Haining, R., J. Ma and S. Wise (1996). Design of a software system for interactive spatial statistical analysis linked to a GIS. Computational Statistics 11, 449-466.

Hall, B. and M. Leahy (2008). Open Source Approaches in Spatial Data Handling. Berlin, Springer-Verlag.

Hardisty, F. and A. Klippel (2010). Analysing spatio-temporal autocorrelation with LISTA-Viz. International Journal of Geographical Information Science 24, 1515-1526.

Hardisty, F. and A. Robinson (2011). The geoviz toolkit: using component-oriented coordination methods for geographic visualization and analysis. International Journal of Geographical Information Science 25, 191-210.

Harris, R., D. Grose, P. Longley, A. Singleton and C. Brunsdon (2010). Grid-enabling geographically weighted regression: a case study of participation in higher

education in England. Transactions in GIS 14, 43-61.

Haslett, J., G. Wills and A. Unwin (1990). SPIDER – An interactive statistical tool for the analysis of spatially distributed data. International Journal of Geographical Information Systems 4, 285-296.

Haslett, J., R. Bradley, P. Craig, A. Unwin and G. Wills (1991). Dynamic graphics for exploring spatial data with applications to locating global and local anomalies. The American Statistician 45, 234-242.

Ihaka, R. and R. Gentleman (1996) R: a language for data analysis and graphics. Journal of Computational and Graphical Statistics 5, 299-314.

Kelejian, H. and I. Prucha (2007). HAC estimation in a spatial framework. Journal of Econometrics 140, 131-154.

Kelejian, H. and I. Prucha (2010). Specification and estimation of spatial autoregressive models and heteroskedastic disturbances. Journal of Econometrics 157, 53-67.

Kielman, J., J. Thomas and R. May (2009). Foundations and frontiers in visual analytics. Information Visualization 8, 239-246.

Kraak M-J. and A. MacEachren (2005). Geovisualization and GIScience. Cartography and Geographic Information Science 32, 67-68.

Kulldorff, M. (2011) SaTScan Version 9.11. http://www.satscan.org.

Laurent, T., A. Ruiz-Gazen and C. Thomas-Agnan (2009). GeoXP: an R package for exploratory spatial data analysis. TSE Working Paper Series 99-099, Toulouse School of Economics. Toulouse, France.

Lee, L-F. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. Journal of Econometrics 137, 489-514.

LeSage, J. (1999). Spatial Econometrics. In The Web Book of Regional Science, Regional Research Institute. Morgantown, WV, West Virginia University.

LeSage, J. and K. Pace (2009). Introduction to Spatial Econometrics. Boca Raton, FL, CRC Press.

Levine, N. (2006). Crime mapping and the CrimeStat program. Geographical Analysis 38, 41-55.

Levine, N. (2010). CrimeStat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations. Washington, D.C., National Institute of Justice.

Li, S., S. Dragicevic and B. Veenendaal (2011a). Advances in Web-based GIS, Mapping Services and Applications. London, Taylor and Francis.

Li, Z., C. Yang, H. Wu, W. Li and L. Miao (2011b). An optimized framework for seamlessly integrating OGC web services to support geospatial sciences. International Journal of Geographical Information Sciences 25, 595-613.

Liu, X. and J. LeSage (2010). Arc_Mat: a Matlab-based spatial data analysis toolbox. Journal of Geographical Systems 12, 69-87.

Lunn, D., D. Spiegelhalter, A. Thomas and N. Best (2009). The BUGS project: evolution, critique and future directions. Statistics in Medicine 28, 3049-3082.

MacEachren, A. (1995). How Maps Work: Representation, Visualization and Design. New York, Guilford Press.

MacEachren, A. and M-J Kraak (1997). Exploratory cartographic visualization: advancing the agenda. Computers and Geosciences 23, 335-343.

MacEachren, A.M., Crawford, S., Mamata Akella and Lengerich, G., 2008. Design and Implementation of a Model, Web-based, GIS-Enabled Cancer Atlas. The

Cartographic Journal, 45: 246-260.

Mathsoft (1996). S+SpatialStats User Manual, Version 1.0. Seattle, WA, Mathsoft Inc.

Michaelis, C.D. and D.P. Ames (2009). Evaluation and implementation of the OGC web processing service for use in client-side GIS. GeoInformatica 13, 109-120.

Monmonier, M. (1989). Geographic brushing: enhancing exploratory analysis of the scatterplot matrix. Geographical Analysis 21, 81-84.

Neteler, M. and H. Mitasova (2008). Open Source GIS: a GRASS GIS approach. Berlin, Springer.

NSF (2003). Revolutionizing Science and Engineering Through Cyberinfrastructure. Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. Washington, DC, National Science Foundation.

Openshaw, S. (1990). Spatial analysis and geographical information systems: a review of progress and possibilities. In H. Scholten and J. Stillwell (eds.) Geographical Information Systems for Urban and Regional Planning, pp. 153-163. Dordrecht, Kluwer.

Ord, K. (1975). Estimation methods for models of spatial interaction. Journal of the American Statistical Association 70, 120-126.

Pace, K. and J. LeSage (2004). Chebyshev approximation of log-determinants of spatial weights matrices. Computational Statistics and Data Analysis 45, 179-196.

Pace, K. and J. LeSage (2009). A sampling approach to estimate the log determinant used in spatial likelihood problems. Journal of Geographical Systems 11, 209-225.

Peng, Z. and M-H Tsou (2003). Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Networks. New York, NY, Wiley.

Peng, Z.-R. and C. Zhang (2004). The roles of geography markup language (GML), scalable vector graphics (SVG) and web feature service (WFS) specifications in the development of Internet geographic information systems. Journal of Geographical Systems 6, 95-116.

Piras, G. (2010). sphet: spatial models with heteroskedastic innovations in R. Journal of Statistical Software 35, 1-21.

Plewe, B. (1997). GIS Online. Information Retrieval, Mapping and the Internet. Santa Fe, NM, OnWorld Press.

Raymond. E. (1999). The Cathedral and the Bazaar. Sebastopol, CA, O'Reilly.

Rey, S. (2009). Show me the code: spatial analysis and open source. Journal of Geographical Systems 11, 191-207.

Rey, S. and L. Anselin (2006). Recent advances in software for spatial analysis in the social sciences. Geographical Analysis 38, 1-4.

Rey, S. and M. Janikas (2006). STARS: Space-time analysis of regional systems. Geographical Analysis 38, 67-86.

Rhyne, T.-M., A. MacEachren, and J. Dykes (2006). Exploring geovisualization. IEEE Computer Graphics and Applications 26, 20-21.

Scharl, A. and K. Tochtermann (2007). The Geospatial Web: How Geobrowsers, Social Software and the Web 2.0 are Shaping the Network Society. London: Springer.

Smirnov, O. (2005). Computation of the information matrix for models with spatial interaction on a lattice. Computational and Graphical Statistics 14, 910-927.

Smirnov, O. and L. Anselin (2001). Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach.

Computational Statistics and Data Analysis 35, 301-319.

Smirnov, O. and L. Anselin (2009). A O(N) parallel method of computing the log-Jacobian of the variable transformation for models with spatial interaction on a lattice. Computational Statistics and Data Analysis 53, 2980-2988.

Steinger, S. and E. Bocher (2009). An overview of current free and open source desktop GIS developments. International Journal of Geographical Information Science 23, 1345-1370.

Stuetzle, W. (1987). Plot windows. Journal of the American Statistical Association 82, 466-475.

Symanzik, J., J. Majure and D. Cook (1994a). Dynamic graphics in a GIS: a bidirectional link between ArcView 2.0 and XGobi. Computing Science and Statistics 27, 299-303.

Symanzik, J., J. Majure, D. Cook and N. Cressie (1994b). Dynamic graphics in a GIS: a link between Arc/Info and XGobi. Computing Science and Statistics 26, 431-435.

Symanzik, J., D. Cook, N. Lewin-Koh, J. Majure and I. Megretskaia (2000). Linking ArcView and XGobi: Insight behind the front end. Journal of Computational and Graphical Statistics 9, 470-490.

Tait, M. (2005). Implementing geoportals: applications of distributed GIS. Computers, Environment and Urban Systems 29, 33-47.

Takahashi, K., M. Kulldorff, T. Tango and K. Yih (2008). A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring. International Journal of Health Geographics 7, 14.

Thomas, J. and K. Cook (2005). Illuminating the Path; The Research and Development

Agenda for Visual Analytics. Los Alamitos, CA, IEEE Computer Society Press.

Tiwari, C. and G. Rushton (2010). A spatial analysis system for integrating data, methods and models on environmental risks and health outcomes. Transactions in GIS 14, 177-195.

Tukey, J. (1977). Exploratory Data Analysis. Reading, MA, Addison Wesley.

Turton, I. (2008). Geotools. In B. Hall and M. Leahy (eds.), Open Source Approaches to Spatial Data Handling, pp. 153-170. Berlin, Springer-Verlag

Ungerer, M. and M. Goodchild (2002). Integrating spatial data analysis and GIS: a new implementation using the component object model (COM). International Journal of Geographical Information Science 16, 41-53.

Unwin, A. (1994). REGARDing geographic data. In P. Dirschedl and R. Osterman (eds.) Computational Statistics, pp. 345-354. Heidelberg, Physica Verlag.

Venables, W. and B. Ripley (1994). Modern Applied Statistics with S-Plus. New York, NY, Springer-Verlag.

Wang, S. (2010). A CyberGIS framework for the synthesis of cyberinfrastructure, GIS and spatial analysis. Annals of the Association of American Geographers 100, 535-557.

Wang, S. and M. Armstrong (2009). A theoretical approach to the use of cyberinfrastructure in geographical analysis. International Journal of Geographical Information Science 23, 169-193.

Wang, S. and M. Cowles and M. Armstrong (2008). Grid computing and spatial statistics: using the TeraGrid for Gi*(d) analysis. Concurrency and Computation: Practice and Experience 20, 1697-1720.

Wang, S. and Y. Liu (2009). TeraGrid GIScience Gateway: bridging cyberinfrastructure and GIScience. International Journal of Geographical Information Science 23, 631-656.

Wang, S., Y. Liu, N. Wilkins-Diehr and S. Martin (2009). SimpleGrid Toolkit: Enabling geosciences gateways to cyberinfrastructure. Computers and Geosciences 35, 2283-2294.

Wiegand, N., D. Kolas, and G. Berg-Cross (2010). Intersecting semantic web and geospatial technologies. Transactions in GIS 14, 93-95.

Wise, S., R. Haining and J. Ma (2001). Providing spatial statistical data analysis functionality for the GIS user: the SAGE project. International Journal of Geographical Information Science 15, 239-254.

Wright, D. and S. Wang (2011). The emergence of spatial cyberinfrastructure. Proceedings of the National Academy of Sciences 108, 5488-5491.

Yalta, A.T. and A.Y. Yalta (2010). Should economists use open source software for doing research. Computational Economics 35, 371-394.

Yan, J., M. Cowles, S. Wang and M. Armstrong (2007). Parallelizing MCMC for Bayesian spatiotemporal statistical models. Statistics and Computing 17, 323-335.

Yang, C., W. Li, J. Xie and B. Zhou (2008). Distributed geospatial information processing: sharing distributed geospatial resources to support Digital Earth. International Journal of Digital Earth 1, 259-278.

Yang, C. and R. Raskin (2009). Introduction to distributed geographic information processing research. International Journal of Geographical Information Science 23, 553-560.

Yang, C., R. Raskin, M. Goodchild and M. Gahegan (2010). Geospatial cyberinfrastructure: past, present and future. Computers, Environment and Urban Systems 34, 264-277.

Chaowei Yang, Michael Goodchild, Qunying Huang, Doug Nebert, Robert Raskin, Yan Xu, Myra Bambacus, Daniel Fray (2011). Spatial cloud computing: how can the geospatial sciences use and help shape cloud computing? International Journal of Digital Earth 4, 305-329.

Yue, P., L. Di, L. Sun, Q. Wang, J. Gong, J. Yuan and Z. Sun (2010). GeoPW: Laying blocks for the Geospatial Processing Web. Transactions in GIS 14, 755-772.

Zhang, Z. and D. Griffith (2000). Integrating GIS components and spatial statistical analysis in DBMSs. International Journal of Geographical Information Science 14, 543-566.