现代应用统计论文集

习童 主编

上海财经大学出版社

现代应用统计论文集

习　童　主编

上海财经大学出版社

图书在版编目(CIP)数据

现代应用统计论文集/习童主编. —上海：上海财经大学出版社，2009. 9

ISBN 978-7-5642-0615-4/F·0615

Ⅰ. 现… Ⅱ. 习… Ⅲ. 统计学—文集 Ⅳ. C8-53

中国版本图书馆 CIP 数据核字(2009)第 164303 号

□ 责任编辑 袁 敏 刘 兵
□ 封面设计 捷纳百川

XIANDAI YINGYONG TONGJI LUNWENJI

现 代 应 用 统 计 论 文 集

习 童 主编

上海财经大学出版社出版发行
（上海市武东路 321 号乙 邮编 200434）
网 址：http://www.sufep.com
电子邮箱：webmaster@sufep.com
全国新华书店经销
上海敬民实业有限公司长阳印刷厂印刷装订
2009 年 9 月第 1 版 2009 年 9 月第 1 次印刷

787mm×1092mm 1/16 22 印张 498 千字
定价：42.00 元

期待更好的发展(代序)

黄树颜

感谢同学们！25 年来，在全球很多地方，在很多领域，同学们都取得了非常好的成绩，我感到非常高兴。可以说，这是对我最大的安慰，也是对这些年来为这个专业辛勤劳动的同仁们最大的安慰！同时也使我想起了许多往事，这些往事又像是一个中国老知识分子一生的写照。

我 1920 年出生，抗战期间在西南联大求学，学习经济。当时经济问题人人关注，读经济最热门。专家、学者、政府工作人员、学生们差不多天天举行讨论会和写文章，十分热闹。但是我觉得洋洋大观的文章很多，可是往往流于空谈，说不清谁对谁错，谁比较符合实际！因为我认为他们缺乏翔实可靠的数据或者没有科学的推断为根据。我当时渴望有一个以统计数据为根据、以有力的经济理论为指导的方法来研究经济。

大学一毕业，恰逢抗战胜利，有机会参加全国统考，录取后赴美留学，在华盛顿大学攻读学位。在那里，我接触到计量经济这门学科，喜出望外，这正是我梦寐以求的经济学——实证而不是空谈！我先获得硕士学位，后继续攻读博士学位，希望学成回国，一展"宏图"。

1949 年，解放战争节节胜利，遂与进步学生从旧金山辗转回国，9 月中旬到达北京，由教育部派员接待，安排参加开国大典，并进华北大学政治研究所(后改为人民大学)学习。当时授课的，尤其是讲授马列课程的有很多著名学者，学习为期半年，受益匪浅，为此后工作及建立新的人生观打下了基础。因为工作需要，提前结业，先后在国家统计机构和基层工作。曾与苏联专家共事设计统计制度，也参加过重大普查。能参加统计工作，这应该算是我实现了部分愿望。不久，在周恩来总理落实知识分子政策中被调到了教育岗位，当时我很希望进一步实现我的最后愿望——让统计学与经济学相结合。

遗憾的是，此后的二十多年，我只有机会在院校教数理统计、经济统计、高等数学，甚至英语等课程，根本没有把统计学与经济学结合起来的机会！当时，受某些思潮的影响，用数学、数理统计的方法研究经济问题是禁区，要受到批判！所以我"海归"的最初三十年是"有希望而无机会"，"壮志"未酬，但也于心无愧。

直到改革开放，迎来了科学的春天。一个偶然的机会我被调到上海财经学院(后改为上海财经大学，简称"财大")。正好时任国家统计局局长李成瑞高瞻远瞩，

提出中国要搞统计教育改革，财大领导姚𢍰等同志有胆识、有勇气地设置了现代应用统计专业，并开设研究生班，培养能与国际接轨的高层次统计人才，充实国家统计局统计干部和财大师资力量。

按当时情况，除了领导支持以外，还必须辅之以外援。因此，我利用平时学术交流，以及老同学等关系，邀请到美国统计学会专家与国内外名教授无偿地结合国内情况反复研究了课程设置、教学计划、教材选用等问题，并低薪或无偿地聘请到国内外堪称一流的教授，如美国芝加哥大学刁锦寰教授、美国密歇根大学 Leslie Kish 教授、加拿大曼尼托巴大学陈乃久教授、中科院冯士庸教授、北大谢衷洁教授、武大张尧庭教授、华师大茆诗松教授等亲临上课，采用原版教材或专门讲义。本科生以数学、英语为重，课程设置大致相同，部分采用原版教材。

经过数年努力，边实践边改革，取得了一些成绩。诚如美籍华裔统计学界领军人物刁锦寰教授在他给世界银行的报告中说，这个尝试可以作为中国统计教育改革的一个"good example"，又如当时国家统计局一位司长说，财大分配来的学生知识结构新、功底扎实、作风踏实，将来一定可以挑重担。今天财大满园春色的这个大会，确实证明了他们的预言。

再令我高兴的是：在发展我们统计改革事业的同时，我们又在国内先行筹建了数量经济专业，现在已发展成为学士、硕士、博士等成系列的学科，可以说实现了我当时在西南联大读书时形成的一个愿望。

想到这些，我要深深地感谢改革开放，感谢国家统计局的领导，感谢财大给我这样的机会，感谢共创这些专业同仁们的辛勤劳动和智慧，特别是我们有志气、有作为的同学们。工厂出名靠产品，学校出名靠培养出对社会有贡献的人才。没有你们就没有我的今天。

总结自己走过的九十年，可分为三个阶段：第一个阶段是动荡的三十年，是磨剑；第二个阶段是有希望没有机会的三十年；第三个阶段是我焕发青春的三十年，既有希望又有机会。与你们大家在一起的三十年，是我最满意的三十年。

现在开始进入第四个三十年，是更加期待的三十年。期待同学们有更好的发展，期待学校有更好的发展，期待祖国有更好的发展。

（注：本文系根据潘建成同学 2009 年 6 月 26 日在作者上海市虹口区紫荆花园寓所所做的访谈整理而成。）

目　录

Spatial Data Analysis with Spatial Statistics and GIS

Shuming Bao[1] and Luc Anselin[2]

1. China Data Center, University of Michigan, 1007 E Huron St, Ann Arbor, MI 48104
Tel: (734)647 - 9610, Fax: (734)763 - 0335, E-mail: sbao@umich.edu

2. School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ 85287
Tel: (480) 965 - 7444, Fax: (480) 965 - 8313, E-mail: luc.anselin@asu.edu

Abstract: The extension of the functional capacity of geographic information systems (GIS) with tools for exploratory spatial data analysis (ESDA) has been an increasingly active area of research in recent years. In this paper, two operational implementations that link spatial analysis software with a GIS are considered more closely. They consist of a linkage between the *SpaceStat* software for spatial data analysis and the *ArcView* GIS (based on so-called loose coupling), and the *S - PLUS for ArcView* (based on so-called close coupling). The emphasis is on the implementation of methods of exploratory spatial data analysis to describe spatial distributions, visualize spatial patterns, and assess the presence of spatial association. Conceptual and technical issues related to the implementation of these approaches are addressed and some ideas are formulated on future directions for linking ESDA and GIS.

1 Introduction

In this paper, 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 [Anselin (1992, 1995a)] and the *ArcView* GIS [ESRI (1997)]. The other is designed as a seamless integration (close coupling) between the *S - PLUS* statistical computing environment [Mathsoft (1996a)] and the *ArcView* GIS. For each of these approaches we next briefly describe the overall architecture, linkage mechanism and operational implementation. We close with a comparison of the relative merits of these approaches and some thoughts on future developments.

2 Strategies in Linking *SpaceStat* and *S - PLUS* with *ArcView* GIS

As one of the most popular GIS software, *ArcView* is primarily geared to the manipulation of spatial vector data. Although *ArcView* doesn't offer much functionality of spatial statistics, the object oriented Avenue script language supported by *ArcView* allows external modules or programs to be integrated into *ArcView* environment by customizing *ArcView* GUI with user-developed programs. Both the *SpaceStat-ArcView* Link and *S - PLUS for ArcView* are implemented by customizing the *ArcView* user interface (Menu, Button, Tools) and adding functions written in Avenue and C programs, and are deployed as an extension to *ArcView* GIS.

Built in the *GAUSS* programming environment [Aptech (1995)], *SpaceStat* [Anselin (1992, 1995a)] is a DOS operating system based software package for the analysis of spatial data. The *SpaceStat* is written in *GAUSS* and distributed with a *GAUSS* Runtime Module. *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. The linkage between *ArcView* and *SpaceStat* is based on a loose coupling approach by means of a bi-directional data transfer with *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. The *SpaceStat-ArcView* Link is characterized by the following features: (1) a focus on exploratory spatial data analysis (ESDA) of lattice data; (2) a division of labor between the *ArcView* GIS used for the visualization of the statistical results and *SpaceStat* used for the statistical computation; and (3) an implementation targeted at PC platforms and windows environments. These functions fall into two categories: (1) data output to file formats compatible with *SpaceStat*; and (2) joining and mapping of *SpaceStat* output (report files) in a view window. The linkage is static and indirect in the sense that all the *SpaceStat* commands have to be issued within *SpaceStat* and cannot be called from within *ArcView* (and vice versa).

S - PLUS for ArcView is implemented primarily in the *ArcView* environment, by means of a set of Avenue programs that call special purpose functions included in a DLL (Dynamic Link Library) and that are associated with menu items [see Bao and Martin (1997)]. As a comprehensive statistical software with over 2 000 functions, *S - PLUS* [MathSoft (1997a)] provides powerful capabilities for graphical data analysis and statistical modeling. The added module *S+SpatialStats* [MathSoft (1996b)] provides analytical functionality to handle geostatistical data, lattice data, and point data, such as variogram, co-variogram, kriging, Moran's I statistic, Geary C, and the estimation of spatial regression models. The interface between *ArcView* and *S - PLUS* is based on a close coupling approach by means of a bi-directional linkage. The *ArcView* GIS serves as the visualization engine, while *S - PLUS* is used for spatial data analysis. The main objective of the interface is to provide a comprehensive and efficient tool for spatial data analysis that can be accessed from within a GIS environment. *S - PLUS for ArcView* is characterized by the following features: (1) a seamless integration of *S - PLUS* with *ArcView*; (2) access to the full range of *S - PLUS* functions; and (3) a graphical user interface that hides the full complexity of the linkage mechanisms.

S - PLUS for ArcView is implemented as an "extension" to the *ArcView* GIS software. The customary user interface is augmented with two new menus and user interaction is carried out by means of a set of dialogs, constructed using ESRI's ArcView Dialog Designer [ESRI (1997)]. Each dialog invokes a number of Avenue scripts and C programs contained in a DLL (Dynamic Link Library). Once the *S - PLUS* extension is loaded into *ArcView*, the *S - PLUS* window is launched and a connection between *S - PLUS* and *ArcView* is established. The linkage between *S - PLUS* and *ArcView* is dynamic and bi-directional. Data transfer occurs primarily between *ArcView* shapefiles and *S - PLUS* data frame objects. Within *ArcView*, selected fields and records are extracted from a shapefile and exported to *S - PLUS* as a data frame object. Conversely, data from an *S - PLUS* data frame object can be imported into *ArcView* as a generic table and joined with a selected theme's attribute table. Spatial weight matrices are constructed using scripts that exploit the geo-locational information from *ArcView* shapefiles and are saved in *S - PLUS* as spatial neighbor objects, which can then be used in spatial statistical analyses such as spatial autocorrelation, local spatial association, and spatial regression. Similarly, results are saved in *S - PLUS* as data frame objects that can be imported into *ArcView* for visualization. Summary reports from *S - PLUS* analyses are output to ASCII text files and displayed in Notepad, a standard text editor of Microsoft Windows.

3 Operational Issues on the *SpaceStat-ArcView* Link and *S - PLUS for ArcView*

3.1 Data Transfer and Conversion

Since *SpaceStat* uses the DOS version of *GAUSS* and *ArcView* runs under Windows, there is no direct mechanism to call internal *SpaceStat* functions from the *ArcView* environment. Although the MS windows platforms allow *SpaceStat* to run in a multi-tasking environment with *ArcView* in a separate window, no direct conversation can be established between the two programs. All data transfers are via intermediate ASCII (text) files with formats designed to be compatible with both *SpaceStat* and *ArcView*.

Data are passed between *SpaceStat* and *ArcView* using auxiliary files with standardized file names and data formats. Specifically, spatial information is moved from *ArcView* to *SpaceStat* for analysis, and location-specific results are passed back from *SpaceStat* to *ArcView* for visualization. Location-specific results include computed spatially transformed variables (such as spatially lagged variables to construct spatial bar charts or spatial pie charts), outliers (to be visualized in a box plot or box map), a Moran scatterplot, and local indicators of spatial association such as the Local Moran and Gi statistics [Anselin and Bao (1996, 1997)].

In *S-PLUS for ArcView*, data and commands are passed between the two environments using an automation technique. Data export from *ArcView* to *S-PLUS* is packaged into a SafeArray which can then be sent to *S-PLUS* via Automation. Once *S-PLUS* receives it, it writes it out to the current *S-PLUS* working directory. Specifically, attributes and spatial information, along with the *S-PLUS* commands for analysis, are moved from *ArcView* to *S-PLUS*, and location-specific results are passed back to *ArcView* for visualization. All data exported to *S-PLUS* are stored as *S-PLUS* objects. Conversely, data import from *S-PLUS* is exported to a text file which is then added into *ArcView* as a new table and linked with the current *ArcView* attribute table. Location-specific results, including estimated values from kriging, spatial regression, and local spatial statistics, can be integrated with *ArcView* tables. *S-PLUS* graphs are incorporated into layouts as external graphic files saved in PostScript format, for output with other *ArcView* elements in the usual fashion.

To export data from *ArcView* to *S-PLUS*, users can select a subset of records from the attribute table associated with the current theme or coverage. The selected fields and records are extracted from the shape files, and transferred to *S-PLUS* as a data frame object. To import data from *S-PLUS*, users can select one or more columns from a given *S-PLUS* data frame. The selected data are imported into *ArcView* as a new table. For image data, *S-PLUS for ArcView* allows users to manually export raster format grid files into *S-PLUS*. Finally, an *S-PLUS* data frame containing $X-Y$ coordinates may be imported into *ArcView* as a new point theme.

3.2 Spatial Data via Non-spatial Attribute Data

For spatial statistics, a fundamental element is the spatial weight, a measurement of spatial linkages or proximity of observations. The spatial weight matrices represent the strength of the potential interaction between locations. The spatial weight can be defined as a matrix, denoted by $W(n \times n)$.

A general spatial weight matrix can be defined by a symmetric binary contiguity matrix, which can be generated from topological information from geo-locational data by using adjacency or distance criteria:

Adjacency criterion:

$$w_{ij} = \begin{cases} 1, & \text{if location } j \text{ is adjacent to } i \\ 0, & \text{if location } j \text{ is not adjacent to } i \end{cases}$$

Distance criterion:

$$w_{ij}(d) = \begin{cases} 1, & \text{if location } j \text{ is within distance } d \text{ from } i \\ 0, & \text{otherwise} \end{cases}$$

WEIGHTA is a binary variable that represents a neighborhood relationship between locations. For ease of interpretation, another general spatial weight matrix is defined in row standardized form, in which the row elements sum to one. A number of procedures to construct spatial weight matrices using the topological information given by various GIS systems have been suggested [Anselin et al. (1992); Can (1992); Ding and Fotheringham (1992)].

Since the contiguity matrix cannot differentiate the strength of spatial linkages between adjacent locations, some more complex spatial weight matrices are proposed for more precise measurement of spatial linkages. Cliff and Ord (1981) suggested a combination of distance measure and the relative length of the border between spatial units. The resulting spatial weight matrix is asymmetric, which is defined as $w_{ij} = (d_{ij})^{-a}(\beta_{ij})^{b}$, with d_{ij} as the distance between location i and j, β_{ij} as the proportion of the interior boundary of location i which is in contact with location j, and a and b as parameters.

Similarly, Dacey (1968) defined the spatial weight matrix by taking into account the relative area of the spatial units: $w_{ij} = d_{ij}\alpha_i\beta_{ij}$, with d_{ij} as a binary contiguity factor, α_i as the share of unit i in the total area of all space in the study, and β_{ij} as the boundary measure used above. Bodson and Peeters (1975) introduced a general accessibility weight by combing the influence of several channels of communication between spatial units into a logical function: $w_{ij} = \sum_j k_j \{a/[1+b \cdot \exp(-c_j d_{ij})]\}$, with k_j as the relative importance of the means of communication j (such as roads, railways and other communication links), d_{ij} as the distance between unit i and j, and a, b and c_j are parameters, which need to be estimated.

The spatial weights can be defined by spatial neighborhood or spatial distance based on spatial information. Examples of spatial information are coordinates (such as the *X* and *Y* coordinates of the centroid of a polygon) and topological information on the spatial arrangement of selected points or areal units (such as spatial neighbor contiguity). Although the spatial information are usually not transparent to users, it can be extracted from the shape files by programs in Avenue scripts.

In *SpaceStat-ArcView*, a new function, programmed in Avenue, has been added to the Data Menu to allow users to add the *X* and *Y* coordinates into the attribute table. The units of *X* and *Y* coordinates depend on the current projection of the selected data layer. Those *X* and *Y* coordinates can then be exported along with other selected variables to an external ASCII data file in the format consistent with *SpaceStat*.

In Shape files, the topological relationship of polygons provides information for spatial neighborhood. Several added functions in the *SpaceStat-ArcView* Link have been provided for defining spatial weights, which include Rook Weights from Shape File and Queen Weights from Shape File. For point data, the spatial weights can be defined by using *X* and *Y* coordinates exported to the *SpaceStat* dataset. Based on those binary spatial weights created by the *SpaceStat-ArcView* Link, users can use *SpaceStat* to construct more complicated spatial weights such as standardization and manipulation of spatial weights matrices. For irregular and regular lattices, *SpaceStat* provided computations for roots and other characteristics of the weights matrix. With *X* and *Y* coordinates information, *SpaceStat* can calculate distance-based weights matrices.

Similar to *SpaceStat-ArcView* Link, *S-PLUS for ArcView* provides an option for users to include *X* and *Y* coordinates of points or centroids of polygons when they transfer attribute data to *S-PLUS* from *ArcView*. The data transferred to *S-PLUS* is stored as data frame objects. The *S+SpatialStats*, an added module to *S-PLUS*, provided analytical functions for geostatistical data, lattice data, and point data.

In *S-PLUS for ArcView*, users can use the Spatial Neighbor function to construct binary spatial neighbor objects for spatial statistics and modeling by means of Avenue's buffering and spatial query functions for *ArcView* shapefile data. In *S-PLUS/S+SpatialStats*, the spatial weights are stored in an object structure called spatial neighbor. Since *S-PLUS* doesn't have direct access to geographical data such as ARC/INFO files or shapefiles, users have to define spatial weights externally and saved the spatial neighborhood information in a text file, which can then be imported into *S-PLUS* and converted into a spatial neighbor object. Spatial Neighbor function in *S-PLUS for ArcView* provides an easy tool for constructing spatial neighbor objects. The spatial weights can be defined by using the topological information (adjacency criteria). Several options are provided in calculating the spatial weights, such as First Order Neighbor Weights, Adjusted First Order Neighbor Weights (a combination of adjacency and distance criteria), and Higher Order Neighbor Weights. The Spatial Neighbor function also provided several distance-based methods for both point and polygon shapefiles. The user can specify distance values, using border-to-border or centroid-to-centroid measurement options. The distance units are specified in the *ArcView* properties dialog. First Order Neighbor Weights constructs a binary spatial neighbor object based on the adjacency of spatial units. The element ($w[i, j]$) of the spatial weight matrix W is one if polygon j is adjacent to polygon i, and 0 otherwise. Adjusted First Order constructs a spatial neighbor weight not only on the topological relationship but also on the geo-spatial distance between spatial units. A spatial weight is first defined by using the adjacency criteria. Then, an average (centroid-to-centroid) distance between the neighbor polygons and the polygon (i) is calculated. Any polygon beyond the defined neighbor polygons of a polygon (i) will be included as a neighbor polygon if its spatial distance (centroid-to-centroid) to the polygon (i) is less than the average distance. The element ($w[i, j]$) of the weight matrix W is 1 if polygon j is adjacent to polygon i under the above criteria, and 0 otherwise. Higher Order Neighbor Weights are constructed by using a similar criteria in defining higher order neighbor weights. The second order spatial weight matrix is based on the first order spatial weight, and the nth order spatial weight matrix is based on the ($n-1$)th order spatial weight. The spatial weight element ($w[i, j]$) of the nth order weight matrix W is 1 if polygon j is adjacent to the neighbors of order $n-1$ of polygon i, and 0 otherwise. Those simple binary weights can then be transformed to a more complicated spatial weight by assigning different weights in *S-PLUS*. To enable *S-PLUS* objects to be linked to the *ArcView* attribute tables, an identity variable is necessary for each joined table.

3.3 Functions of Spatial Statistics

The current capabilities of spatial statistics in *SpaceStat* include descriptive statistics such as mean, variance, standard deviation, quartiles, interquartile distance, third and fourth moment, test for normality, and correlation coefficients; statistics for spatial autocorrelation and spatial associations such as join count statistics, Moran's I and Geary's C for global autocorrelation, Local Morans and Gearys (LISA), Getis-Ord G and Gi statistics (for local spatial associations); Hubert-Golledge QAP statistics (implemented for generic case, as well as for Moran, Geary and absolute differences); Wartenberg multivariate spatial autocorrelation; Least squares regression with diagnostics for spatial dependence (including

Moran's I, Lagrange Multiplier tests, Kelejian-Robinson test); robust least squares regression (Jackknife); maximum likelihood estimation of regression models with spatial autoregressive errors and spatial autoregressive dependent variable (with diagnostics); GMM estimation of models with spatial autoregressive dependent variable; bootstrap estimation of models with spatial autoregressive dependent variable; and two-step and ML estimation of heteroskedastic models (groupwise heteroskedasticity, random coefficients).

The *S - PLUS for ArcView* provided the following functions of spatial statistics for geostatistical data, lattice data (polygon data), and point data. The available functions for geostatistical data include contour plots, 3-D point clouds, variogram plots and boxplots, directional variograms and correlograms for exploring anisotropy, empirical variogram estimation including robust methods, variogram models including spherical and exponential, ordinary and universal kriging, kriging prediction at arbitrary locations with standard errors, parametric and nonparametric trend surfaces. The functions for point data include point maps that include region boundaries, spatial randomness tests, Ripley's K-functions, simulation of spatial random processes, and local intensity estimation. The functions for lattice data include "Binning" of high density data into a regular lattice of counts, Geary and Moran spatial autocorrelation coefficients, GLISA [Bao and Henry (1996)] for local spatial associations, spatial regression models including conditional and simultaneous autoregressive models, and nearest neighbor search. With those built menu functions in *S - PLUS for ArcView*, users can conduct various spatial analyses from *ArcView* environment without switching to *S - PLUS*, such as spatial autocorrelation, local spatial associations and spatial linear regressions.

3.4 Visualization of Spatial Data

One advantage of linking spatial statistics with GIS is that users can visualize spatial data in different approaches, which can be powerful in exploratory spatial data analyses. Those approaches include: visualize the spatial distribution of data in GIS maps before further statistical analyses and modeling; visualize the results from spatial statistics and models in GIS maps; visualize spatial data in various statistic graphics such as histogram, Moran I scatterplot, empirical variogram, which may reveal more inside nature of spatial data.

In *SpaceStat-ArcView* Link, the following functions implemented for exploratory spatial data analyses are organized into three groups: (1) visualization of the spatial distribution of the data; (2) visualization of spatial autocorrelation in attribute variables; and (3) visualization of local spatial association. These functions are each associated with specific output files generated by the corresponding *SpaceStat* commands [see Anselin (1994, 1995b, 1997) for more technical details].

The first group of functions for spatial data visualization are simple descriptive statistics: Histogram, Box Plot and Box Map. The Histogram is implemented as a standard bar chart for the current selected feature displayed in the View window, following an "Equal Interval" classification. The Box Plot is implemented as a quartile map using the data from the quartile report generated by *SpaceStat*. A "graphic" box plot is also added to the View, which include the upper quartile, lower quartile, outliers, median, and mean. The Box Map is a quartile map augmented with outlier indicators generated from the *SpaceStat* box map Report File. In Anselin and Smirnov (1997), these simple descriptions of spatial distributions are implemented as fully dynamically linked windows by means of external DLL functions instead of Avenue scripts, resulting in increased speed and flexibility. The second group is derived from spatial transformations in *SpaceStat*. The commands include the Spatial Lag Bar Chart and Spatial Lag Pie Chart. Using a Report File from *SpaceStat* that contains the spatial lags for the variables of interest, both functions create *ArcView* spot symbols for a graphic representing respectively a pie chart or bar chart for all selected polygons. The third group visualizes the results of local indicators for spatial autocorrelation computed in *SpaceStat*. These include a Moran Scatterplot and Map, LISA Local Moran Map and G-Stat Map. Each of these functions requires the input of a *SpaceStat* Report File with a fixed file name prefix (such as MS_, LM_, or GI_) followed by the name of the spatial weights file for which the spatial statistics were constructed. In addition to those functional menus, a number of tool buttons are implemented for the identification of a dynamic linkage between the maps, tables and charts in different application windows. Once those tools are activated, a rudimentary form of dynamic linking is established between the selected spatial units in different "views" of the data.

S - PLUS for ArcView provided a direct access to *S - PLUS* objects and statistic graphs. With the added menus to the *ArcView* GUI (Graphic Utility Interface), variables from *S - PLUS* data frame objects may be displayed in an *ArcView* map using the Color Classification and Spatial Bar/Pie Chart options. These functions are especially useful for comparing analytical results, such as fitted values and residuals from classical or spatial regression analyses. In addition, *S - PLUS* has a wide variety of editable graphics plot types. The rich graphical features of *S - PLUS* can be accessed with the Import Graph function. Users can easily generate the statistic graph by using the point-and-click interface. Those statistic graphs include empirical variogram, histograms, and various 2D and 3D plots such as Scatter-line plots (Smoothing Spline, Robust, Polynomial Fit, and Exponential Fit), QQ Quantile Plots (QQ Normal Plot w

Line, QQ Plot without Line), Box Plots, Pie Charts, Bar Plots, Scatter Plot Matrix, Regression, Coarse Surface, Grid Surface, image, and Contour Plot. Those *S - PLUS* graphs can be imported into an *ArcView* layout directly and combined with *ArcView* maps and charts for output.

3.5 Customization and Extendibility

Both the *SpaceStat-ArcView* Link and *S - PLUS for ArcView* are implemented primarily in the *ArcView* environment, by means of customized menus associated with Avenue programs and C programs. They are distributed as extensions to the *ArcView*. Several new menus will be added to the *ArcView* GUI after those extensions are loaded.

The *SpaceStat-ArcView*, two additional menus and a few extra buttons and tools have been added to the standard View window: a Data menu and a *SpaceStat* menu. In this link, the role of the Avenue scripts is limited to providing a shell for special-purpose functions included in a DLL (Dynamic Link Library); see Anselin and Smirnov (1997) for details.

The Data menu consists of nine commands divided into three categories: (1) the auxiliary manipulation of spatial information such as adding the X - Y centroid coordinates of polygons and constructing an indicator variable for selected locations; (2) the construction of spatial boundary files based on the information in an *ArcView* Shape file; (3) the data transfer between *ArcView* and *SpaceStat*, such as exporting selected attribute data from *ArcView* to *SpaceStat*, and importing and joining output from the *SpaceStat* report files into *ArcView*.

The *SpaceStat* menu consists of eleven commands divided into five categories: (1) Box Map and Percentile Map. Box Map creates a new View with a quartile map for a selected variable with the outliers highlighted (a box map). Percentile Map creates a new View with a percentile map for a selected variable. (2) Spatial Lag Bar Chart and Pie Chart. The Bar Chart creates a new View with a bar chart map showing the value of a selected variable and its spatial lag (can also be used for any spatial smoother computed in *SpaceStat*). Spatial Lag Pie Chart create a new View with a pie chart map showing the value of a selected variable and its spatial lag. (3) Spatial Smoother creates a new View with a quintile map for the spatially smoothed values of a selected variable. The smoother may be a spatial lag, window average, or any of the rate smoothers computed by *SpaceStat*. Moran Scatterplot Map creates a new View with a unique value map with four colors corresponding to the four quadrants of the Moran Scatterplot of a selected variable. (4) LISA Local Moran Map creates a new View with a unique value map for those locations with a significant Local Moran statistic. Moran Significance Map creates a new View with a combination of a Moran Scatterplot Map and a Local Moran map, showing the quadrant of the Moran Scatterplot only for those locations with a significant Local Moran statistic. G-Stat Map is similar to LISA Local Moran Map but using the Gi or Gi* statistic. (5) Residual Map creates a new View with a standard deviational map for the residuals of any spatial regression in *SpaceStat*. Predicted Map creates a new View with a bar chart map showing the observed and predicted values for any spatial regression in *SpaceStat*.

The *S-PLUS for ArcView* has two added menus: *S-PLUS* Menu and Spatial Stats Menu. The *S-PLUS* menu consists of eleven functions divided into five categories: (1) data transfer between *ArcView* and *S-PLUS*; (2) auxiliary manipulation of spatial information; (3) spatial data visualization; (4) linear regression; and (5) executing *S-PLUS* commands from *ArcView*. The Spatial Statistics menu contains four functions: (1) Spatial Neighbor; (2) Spatial Autocorrelation; (3) Spatial Association; and (4) Spatial Regression [See Bao and Martin (1997) and MathSoft (1998) for technical details].

The Spatial Statistics menus provide direct access to *S-PLUS/S+SpatialStats* functions such as spatial autocorrelation (Moran's I and Geary's C), spatial association (Generalized Local Indicators of Spatial Association, GLISA), and spatial linear regression. Spatial association options include General Local Moran and Local Geary by Bao and Henry (1996), which are derived from the LISA statistic by Anselin (1995). Spatial linear regression include three types of spatial error models [Cressie (1993)]: SAR - *Simultaneous Autoregressive Model* [Whittle (1954)], CAR - *Conditional Autoregressive Model* [Bartlett (1971); Besag (1974)], and MA - *Moving Average Model* [Cliff and Ord (1981)]. The variables can be selected from either an *ArcView* theme table or an *S-PLUS* data frame object, but a spatial neighbor object must have been pre-defined and be consistent with the selected variables for those spatial statistics. The summarized results are output to text files and the estimates are saved in *S-PLUS* objects that can then be joined with an *ArcView* theme table for map visualization. The Linear Regression function allows users to build a regression equation using variables from either the current *ArcView* theme or an *S-PLUS* data frame. A summary report of the regression is saved in a text file, and predicted values and residuals are saved in an *S-PLUS* data frame object, which can easily be joined to the current theme table.

Finally, the Execute *S-PLUS* Commands feature provides ready access to all *S-PLUS* commands from the *ArcView* environment. All *S-PLUS* objects are listed on this dialog window, and users can type in *S-PLUS* commands on the command line, or make use of several auxiliary button functions.

For users to customize their own GUI menus, *S-PLUS for ArcView*