

# UC Santa Barbara

## NCGIA Technical Reports

### Title

The Integration of Spatial Analysis and GIS: The Development of the Statcas Module for ARC/INFO (91-5)

### Permalink

<https://escholarship.org/uc/item/6q49093t>

### Authors

Ding, Yuemin  
Fotheringham, A. Stewart

### Publication Date

1991-04-01

**National Center for Geographic Information and Analysis**

**The Integration of Spatial  
Analysis and GIS: The  
Development of the Statcas  
Module for ARC/INFO**

by

Yuemin Ding and A. Stewart Fotheringham

**National Center for Geographic Information & Analysis  
Department of Geography  
State University of New York at Buffalo  
Buffalo, NY 14261  
USA**

National Center for Geographic Information & Analysis/NCGIA  
**Technical Paper 91-5**  
April 1991

## **THE INTEGRATION OF SPATIAL ANALYSIS AND GIS:**

### **THE DEVELOPMENT OF THE STACAS MODULE FOR ARC/INFO**

#### **ABSTRACT**

It seems widely expected that future GISs will have increased analytical capabilities that will take them beyond being efficient display and database management devices. Several attempts have already been made to link existing analytical software to various GISs. However, a problem with all of these attempts is that the user is forced to switch back and forth between the GIS operating environment and the analytical software. In this paper we present a statistical analysis package, STACAS, that runs totally within the operating environment of a GIS and utilizes a command structure that makes running the package transparent to the user of the GIS.

# THE INTEGRATION OF SPATIAL ANALYSIS AND GIS:

## THE DEVELOPMENT OF THE STATCAS MODULE FOR ARC/INFO

### 1. INTRODUCTION

The research areas of spatial analysis and GIS have generally developed quite independently of one another. The volumes of research on spatial analysis prior to the evolution and popularization of GIS technology clearly demonstrate that spatial analysis, and more generally, geographical analysis, can be undertaken without the aid of a GIS<sup>1</sup>. It is equally clear that GISs have proliferated as display devices without claiming any spatial analytical capabilities other than spatial overlay and perhaps some basic descriptive statistics related to area and distance computations. Rhind's comment, quoted by Densham and Goodchild (1989), is particularly potent: "Virtually all GIS developments thus far have resulted in 'data retrieval and sifting' engines" (Rhind, 1988).

Despite their independent histories, there are several advantages to integrating spatial analytical capabilities within a GIS framework. From a GIS perspective, there is an increasing demand for systems that "do something" other than display maps and organize data. It would be useful to some users to have the capability of analyzing data once they have been displayed. For instance, it might be useful to know something about the statistical relationship between the spatial distribution of welfare recipients and the distribution of welfare offices after displaying the two data sets on one map. Similarly, it would be useful to have some knowledge of the statistical impacts of changing the units for which data are displayed (Fotheringham and Wong, 1991). From the spatial analytical perspective, there are advantages to linking statistical methods to the database and display capabilities of a GIS. In this way the GIS can act as an "enabling technology"; that is, whilst the GIS may not be necessary for the analysis of spatial data, it can facilitate such analysis. It is also possible that the display and data organizational capabilities of the GIS could even produce insights into the analysis that would otherwise be missed. This might be especially the case if increasing access to highly disaggregate data sets is possible (Abler, 1987).

The aim of this paper is to introduce a means of integrating spatial analysis with GIS technology. We take as our example of spatial analysis the calculation of spatial autocorrelation and spatial association statistics and our example of GIS technology, ARC/INFO. Whilst still relatively undeveloped, the integration of these two areas is by no means new. Horn (1988) combined the cartographic and data handling functions of a GIS with other algorithms and interactive techniques for solving spatial planning problems. Young et al (1988) designed a link between ARC/INFO and a decision support system by INFO programming for the forest management. To avoid the limitations of the INFO programming language, Kehris (1990) designed a direct way to access ARC/INFO data files with FORTRAN routines and established an effective link between ARCM4FO and the GLIM statistical package. However, in each of these linkages there is a severe problem in that by not being able to run the spatial analytical routines in the ARC environment, the user has to change the package or system each time between calculation and display. Here we describe a means of overcoming this problem by integrating statistical routines into the ARC Macro Language (AML).

The purpose of this paper is thus to describe a SpAtial AutoCorrelation and ASsociation analysis ( STACAS ) module which is operated as a set of ARC/INFO functions. The spatial autocorrelation and spatial association statistics are first briefly described followed by a discussion of the characteristics and limits of ARC/INFO for the integration of spatial analytical routines. The structure and design of the STACAS module in ARC/INFO along with a description of its main functions is then given. The paper is concluded with a demonstration of the package using population data from the 1990 census of the People's Republic of China and some suggestions for future developments.

### 2. SPATIAL AUTOCORRELATION AND SPATIAL ASSOCIATION

Spatial analysis deals with two quite distinct types of information ( Goodchild, 1986 ). One concerns the attributes of spatial objects, which include measures such as area, population, rainfall, or soil type, etc. The other concerns locational information about the spatial objects which generally are described by means of their positions on map or by geographic coordinate systems.

The spatial objects concerned in most analyses are polygons which correspond to measurement zones, statistical reporting areas such as census tracts and school districts, and points which correspond to sampling points. For some types of spatial analysis it is common to represent polygons by points (a geographic center or a weighted mean center, for example) although this can lead to the introduction of considerable error.

---

<sup>1</sup> The term "spatial analysis", perhaps incorrectly, has become synonymous to many researchers with "spatial statistical analysis" and we confine our views in this paper to this usage. The term "geographical analysis" represents the broader application of quantitative techniques, including mathematical modeling, to spatial problems.

Spatial autocorrelation is concerned with the degree to which objects or activities are similar to other objects or activities located nearby. In contrast to other types of spatial statistical analysis, such as point pattern analysis for example, spatial autocorrelation deals simultaneously with both locational and attribute information. If objects which are similar in location also tend to be similar in attributes, then the pattern as a whole is said to show positive spatial autocorrelation. Conversely, negative spatial autocorrelation exists when objects which are close together in space tend to be more dissimilar in attributes than objects which are further apart. Zero autocorrelation occurs when the distribution of the attributes is independent of the distribution of locations. Excellent reviews of spatial autocorrelation and its role in spatial analysis are given by Goodchild (1986), Odland (1987) and Griffith (1987).

Spatial association statistics measure the concentration of an attribute over space. While they are constructed in a very similar way to spatial autocorrelation measures, they offer the twin advantages of being able to differentiate spatial patterns caused by clusters of low values as opposed to clusters of high values, and they can be disaggregated by polygon or point to provide much more detailed information. This latter property makes them amenable to visual display, an example of which is provided by the demonstration of the STACAS module below. Further information on spatial association statistics is provided below and also by Getis (1990).

### 2.1 Measurement of Spatial Proximity

Both spatial autocorrelation and spatial association statistics examine the relationship of an attribute value in one polygon or for one point with the values for proximal polygons or points. It is thus necessary in the computation of both sets of statistics to define and measure what is meant by "proximal". The spatial proximity measure of n objects in spatial autocorrelation and spatial association is usually described by a weight matrix which is defined by a given spatial relationship as follows

$$W = [ w_{ij} ]_{n \times n}$$

where

$$w_{ij} = \begin{cases} 1 & \text{if polygon } i \text{ and polygon } j \text{ are adjacent;} \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

or

$$w_{ij} = \begin{cases} 1 & \text{if point } j \text{ is within distance } d \text{ from point } i; \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

and

$$w_{ii} = 0. \quad (3)$$

Alternatively, a continuous measure of proximity can be defined in terms of some function of the distance between points and a hybrid measure exists whereby proximity is defined in terms of a function of the length of the common boundary between polygons with non-adjacent polygons having a weight of zero. However, in this study we will restrict the analysis to the more common binary weight matrix described by equations (1) to (3).

### 2.2 Measurement of Spatial Autocorrelation

The most useful measure of spatial autocorrelation is the Moran coefficient, I. The attribute similarity measure between two objects used by the Moran coefficient can be stated as follows:

$$c_{ij} = (x_i - x_{\text{mean}}) (x_j - x_{\text{mean}}) \quad (4)$$

where  $x_i$  denotes the value of the attribute for object  $i$ , and  $x_{\text{mean}}$  denotes the mean of the attribute variable for all  $n$  objects. The similarity measure  $c_{ij}$  when weighted by the spatial proximity measure  $w_{ij}$  and summed over all  $i$  and  $j$  thus measures the

covariance between the value of an attribute at location i and its value at all other locations. Dividing this measure by a standardizing factor that constrains the value of the coefficient to lie between +1 and -1 yields the Moran coefficient:

$$I = \frac{\sum_i \sum_j (w_{ij} c_{ij})}{s^2 \sum_i \sum_j w_{ij}} \quad (5)$$

where  $s^2$  denotes the sample variance of the attribute being measured:

$$s^2 = \frac{\sum_i (x_i - x_{mean})^2}{n} \quad (6)$$

The coefficient is positive when nearby areas or points tend to be similar in attributes, negative when they tend to be dissimilar, and approximately zero when attribute values are arranged randomly and independently in space (Goodchild, 1986).

The simplest and most straightforward null hypothesis on which to test the significance of the Moran coefficient assumes the spatial autocorrelation in the population from which the sample is drawn to be zero. Two assumptions about the sample can be made: one is that the sample values are drawn from a normally distributed population; the other is that they represent one random arrangement of attribute values from all the possible arrangements that could occur (Cliff and Ord, 1981, Goodchild 1986). The expected value and variance of the Moran coefficient for samples of size n under the two assumptions can be determined as follows:

$$E_N(I) = -1 / (n-1) \quad (7)$$

$$E_R(I) = -1 / (n-1) \quad (8)$$

$$\text{Var}_N(I) = \frac{(n^2 W_1 - n W_2 + 3 W_0^2)}{W_0^2 (n^2 - 1)} - E_N(I)^2 \quad (9)$$

$$\begin{aligned} \text{Var}_R(I) = & \frac{n[(n^2 - 3n + 3)W_1 - nW_2 + 3W_0^2]}{(n-1)^3 W_0^2} \\ & - \frac{K_2[(n^2 - n)W_1 - 2nW_2 + 6W_0^2]}{(n-1)^3 W_0^2} \\ & - E_R(I)^2 \end{aligned} \quad (10)$$

where

n is the number of areas or points to be analyzed;

$W_0$  is the total of the weight matrix,  $\sum_i \sum_j w_{ij}$ ;

$$W_1 \text{ is } \sum_i \sum_j (w_{ij} + w_{ji})^2 / 2$$

$$W_2 \text{ is } \sum_i (w_{i.} + w_{.i})^2$$

$w_{i.}$  and  $w_{.i}$  are the row and column totals of the weight matrix,  $\sum_j w_{ij}$  and  $\sum_j w_{ji}$ , respectively.

$K_2$  is the sample kurtosis coefficient

$$K_2 = n \sum_i (x_i - x_{\text{mean}})^4 / (\sum_i (x_i - x_{\text{mean}})^2)^2$$

$n^{(b)}$  denotes the product  $n(n-1)(n-2)\dots(n-b+1)$ .

Thus, the test statistic for both cases is

$$z = (I - E(I)) / \sqrt{\text{Var}(I)} \quad (11)$$

and a normal curve can be used to test the significance of the observed I statistic.

### 2.3 Measurement of Spatial Association

The formal measurement of spatial association, through a G statistic, is more recent than that of spatial autocorrelation and has only recently been brought to the attention of geographers (Getis 1990). Unlike Moran's I which measures the correlation between attribute values and location, the G statistic measures the **concentration** of a spatially distributed attribute variable.

G statistics are based on the weight matrix  $W(d)$  which is determined by a given distance radius in equations (2) and (3). All regions (polygons) to be analyzed are represented as points by their centers. The G statistic has the advantage over the Moran coefficient in that it can be disaggregated by point so that a set of  $G_i(d)$  statistics can be obtained each of which measures the degree of association between weighted point  $i$  and all other weighted points within a radius of distance  $d$  from the point  $i$ . This allows the testing of hypotheses about the clustering of attribute values around each location by calculating:

$$G_i(d) = \frac{\sum_{j \neq i} w_{ij}(d) x_j}{\sum_j x_j} \quad (12)$$

$i = 1, 2, \dots, n$

If large values of the attribute are clustered close to  $i$ ,  $G_i$  will be large. In order to test hypotheses on  $G_i(d)$ , the null hypothesis is that there are no differences (and thus no spatial association) among the  $x_j$  within  $d$  of the point  $i$ .

The expected values and variances of  $G_i(d)$   $i = 1, 2, \dots, n$  can be defined as follows (Getis 1990):

$$E(G_i(d)) = W_i / (n-1) \quad (13)$$

where

$$W_i = \sum_j w_{ij}(d) \quad (14)$$

and

$$\text{Var}(G_i(d)) = \frac{W_i}{(n-1)^2 (n-2)} \left\{ 1 + \frac{(n-1-W_i) (Y_2)}{Y_1^2} \right\} \quad (15)$$

where

$$Y_1 = \sum_j x_j / (n-1)$$

$$Y_2 = \sum_j x_j^2 / (n-1)$$

Thus the z value for each point ( or each polygon ) can be calculated as follows

$$z_i = [G_i(d) - E(G_i(d))] / \sqrt{E(\text{Var}(G_i(d)))} \quad (16)$$

A significant large positive z implies that large values of the attribute are spatially associated with point i or polygon i whereas a significant negative z means that small values of  $x_j$  are spatially associated with point i or polygon i. One advantage of the disaggregation of the G statistic is that it can be mapped within ARC/INFO to show the levels of spatial association across the study area.

### 3. THE POTENTIAL FOR INTEGRATING SPATIAL ANALYSIS AND ARC/INFO

Software presently exists independent of any GIS for certain types of spatial analysis, in particular for the measurement of spatial autocorrelation (Goodchild, 1986, Griffith, 1989). However, from the above brief descriptions of spatial autocorrelation and spatial association analysis, there are three immediate advantages of being able to integrate this type of analysis within a GIS.

The first concerns identifying the spatial relationships between objects which is necessary for building the spatial weight matrix  $W$  to measure the spatial proximity. The weight matrix can be derived from two kinds of relationships between spatial objects: one describes the adjacency relationship between two regions (polygons); the other describes the range of a distance radius from a point or the center of a polygon. It is obviously time-consuming to create the weight matrix manually for a large area but it can be created with little effort within a GIS that has already stored the locational and topological relationships between polygons or points.

In the calculation of spatial autocorrelation and spatial association indices, it is necessary to link the attribute data with the locational information. Hence there is a need for an efficient data model or database to represent, store, manipulate different kinds of



attribute data with a good link to their spatial locations. Much more thought has generally been given to this task within a GIS environment than in a non-GIS environment.

For many spatial statistics, especially those that can be disaggregated by location, such as the spatial association statistic described above, it would be extremely useful to be able to display the results in map form. Linking spatial analysis routines to the powerful graphics capabilities of GISs would make the visual presentation of results much easier especially if the linkage can be achieved without the user having to move from one environment to another.

ARC/INFO is perhaps the most popular commercial GIS used to automate, manipulate, and display geographic data in digital form ( ESRI 1987 ). It has many impressive features for these tasks such as a sophisticated input subsystem for digitizing, editing, and reformatting geographic data; a powerful output subsystem for constructing impressive maps and producing reports; a useful array of spatial operations for topological overlay, buffer creation, and spatial query. However, its analytical capabilities are extremely limited. There is a very restricted repertoire of spatial modelling routines available such as NETWORK and TIN and a few simple statistics can be provided. There is also a problem in combining existing software with ARC/INFO in that there is no direct link to transfer the locational and attribute data and their topological relationships between ARC/INFO and other packages.

One solution to the problem of accessing the potentially powerful capabilities of ARC/INFO to spatial analysis, which we describe below, is to utilize the fourth generation of the ARC Macro Language (AML) which as well as allowing access to INFO, provides programming facilities and the ability to run routines without leaving the ARC environment. We now provide an example of the use of AML to facilitate spatial analysis on ARC/INFO.

#### **4. THE STACAS MODULE IN ARC/INFO**

##### **4.1 The General Objectives**

In designing a new module for spatial analysis on ARC/INFO, we set ourselves the following objectives:

1. To integrate the stored spatial topological information of objects and the spatial operators in ARC/INFO;
2. To develop the necessary routines for calculating the measures of spatial autocorrelation and spatial association described above;
3. To build a direct and simple link to transfer the topological and attribute data between ARC/INFO and the developed programs;
4. To design an efficient method for the data file management;
5. To develop a function to display and print the spatial graphic results of the analysis directly (without explicit accessing of ARCPLOT);
6. To design a friendly interface for the spatial analysis module and to make all the commands in the module similar to those in ARC/INFO and which can directly deal with coverages, INFO files and data files.

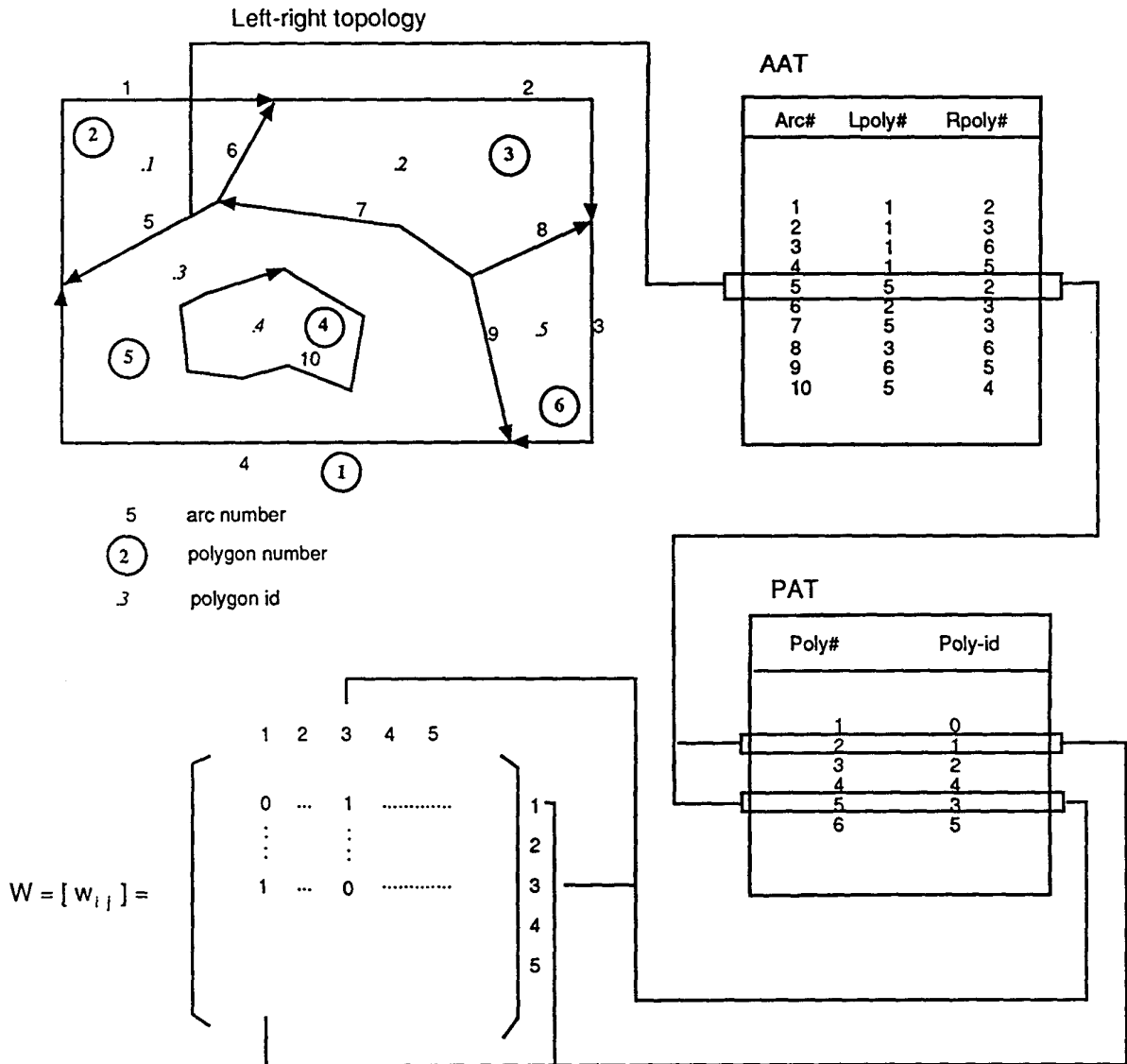
##### **4.2 The Use of Topological Data and ARC Functions for Spatial Analysis**

One of the useful features of ARC/INFO's topological data model is being able to building the weight matrix from the stored adjacency relationships of polygons. Since each arc in ARC/INFO has direction, the Arc Attribute Table ( AAT ) in INFO maintains a list of the polygons on the left and right sides of each arc (ESRI 1990). Thus any polygons sharing a common arc are adjacent. According to the left-right fist in AAT and the relationship between polygon numbers ( which is given by ARC program and shown in both AAT and PAT ) and user specified polygon IDs in Polygon Attribute Table ( PAT ), it is easy to create the weight matrix by checking each arc and its left-right topology, and assigning I to the corresponding row and column in the matrix (Figure 1).

Using a combination of ARC commands, it is also possible to obtain the distance matrix  $D = [ d_{ij} ]_{n \times n}$  which includes the distance between any two points or centers of any two polygons in the study area. The combination of the commands is as follows

```
arc: BUILD <coverage> POINT  
arc: POINTDISTANCE <coverage> <coverage> INFODIS.FILE
```

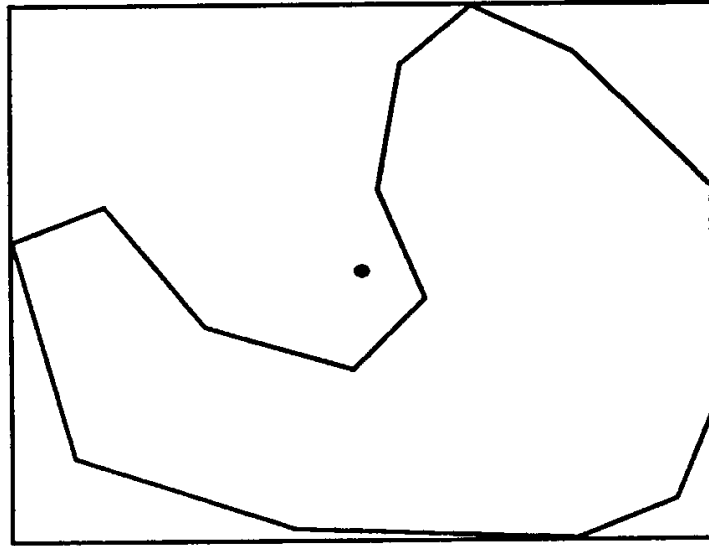
The first command creates a point coverage or a center coverage of polygons. The second calculates the distance between any two points in the coverage and then stores the distance data in a user-specified INFO file. It should be noted that the processing coverage should have centers or label points for polygons in the coverage otherwise both commands will not give the expected results.



**Fig. 1 Adjacent Weight Matrix and Left-right Topology**

Once the distance matrix is obtained, it is a straightforward procedure to create the weight matrix for a given radius  $d$  as follows:

$$w_{ij} = \begin{cases} 1 & \text{if } d_{ij} \leq d; \\ 0 & \text{otherwise.} \end{cases} \quad (17)$$



**Fig. 2 The Boundary Box Center of a Concave Polygon**

The centers of polygons can be created by either digitizing or using the ARC command CREATELABEL. The latter will create the center of polygon boundary box which may be located at outside of some extremely concave polygons (Figure 2).

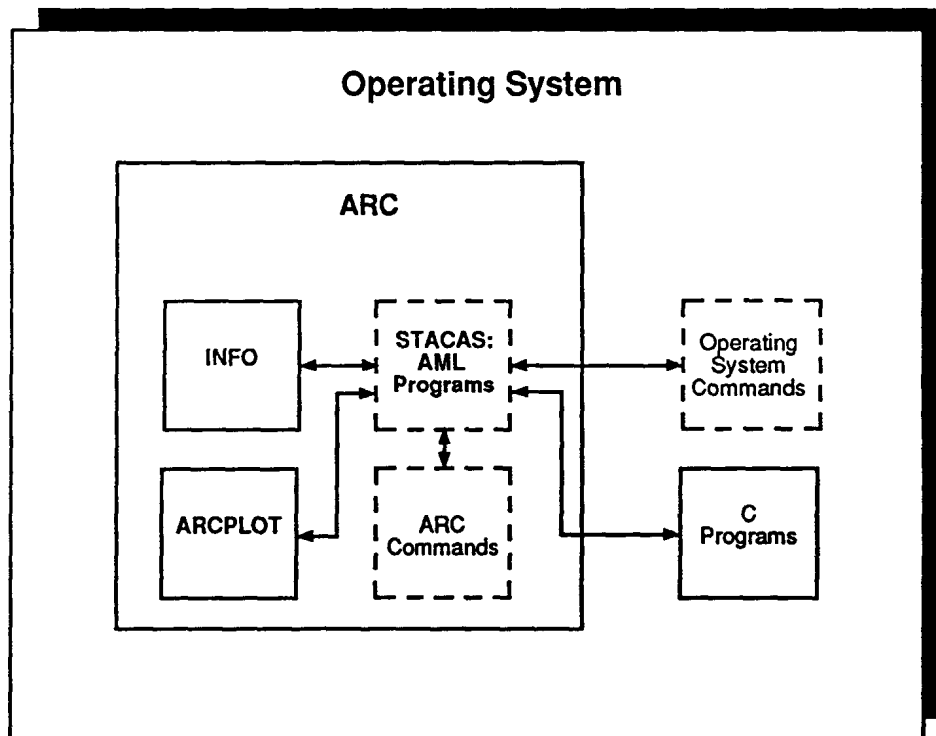
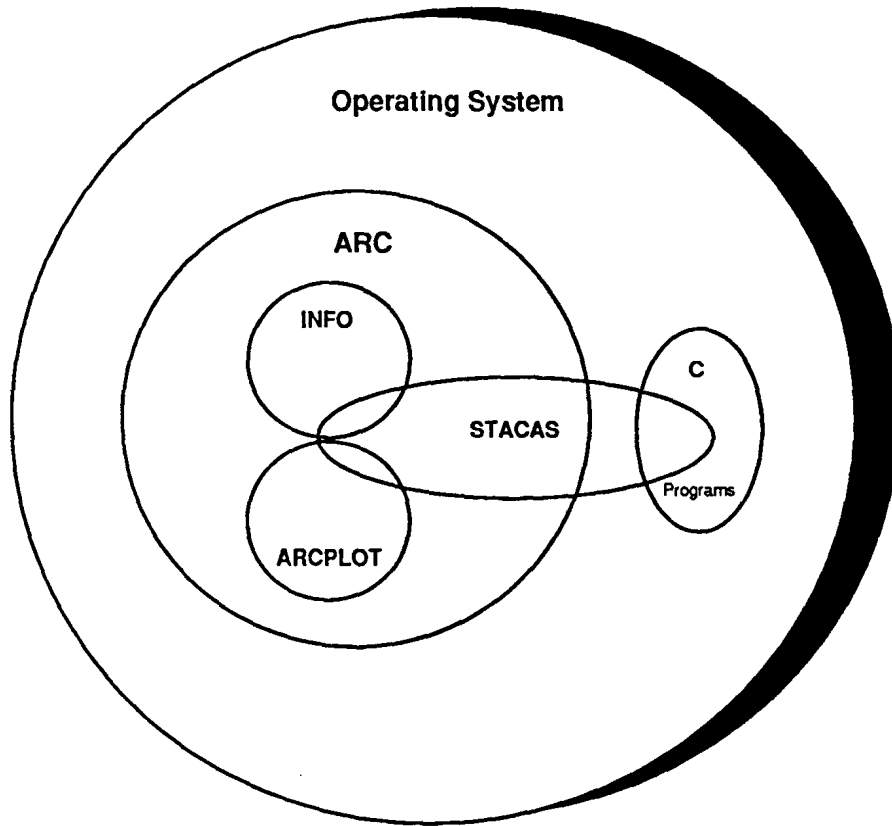
It is not difficult to create centers of polygons based on the coordinates of their boundary arcs, although there remains a definitional question about what kinds of centers make the most sense for the analysis of spatial autocorrelation and spatial association.

#### **4.3 The STACAS Module**

The heart of the STACAS module consists of several programs written in C for calculating the Moran coefficient and associated moments under both the normality and randomization assumptions,  $G_i$  statistics and associated z-scores, and for transferring the necessary data into and out of ARC/INFO files. While the module is presently restricted to these functions, it is written in such a way that it is a simple matter for the user to add new routines and models. The relationship between these programs and the ARC/INFO environment through STACAS is shown in Figure 3.

Because the topological and attribute data all are stored and managed in the INFO database, a direct and simple way has been built into STACAS to transmit data between the INFO database and the spatial analytical routines by AML. An AML directive enables the user to access INFO and use INFO commands to deal with INFO files and data transmission (ESRI 1989). Since most spatial operations are executed by ARC commands, only a few data transmissions are needed to complete the spatial autocorrelation and association analysis.

Since there is a need to copy and transfer coverages and data files during the processing, any processing of commands in STACAS will create a temporary files or coverages when they are needed and will delete all of them when processing finished in the same way that the ARC commands work. This reduces the duplication of data files and save memory space.



**Fig. 3 Relationship between STACAS and Other Components**

### 4.3.1 STACAS Commands

Every command in STACAS is an AML program which is composed of some combination of the following:

**AML directives**

**AML functions**

**ARC commands**

**ARC PLOT commands**

**INFO commands**

**Operating system commands**

**C programs**

The AML directive, &ATOOL, makes all AML programs be the user developed ARC commands. In each AML program, a necessary interface is designed for showing the usage of the command and checking the legal status of coverage and data file to be processed in order to avoid incorrect execution.

There are nine commands in STACAS and their operation is outlined below. All commands in STACAS are run in the ARC environment. To start STACAS the user simply types

**&run stacas**

in the ARC environment at which time all the STACAS commands are made available while retaining the ability to use any of the regular ARC/INFO commands. The basic command structure of STACAS in Figure 4.

A description of each of the commands shown in Figure 4 is now given.

**CNTDISTANCE <coverage> <dist.rile>**

Calculates distances between centers of polygons in the given coverage.

<coverage> -- name of input coverage to be processed.

<dist.file> -- name of file for storing the distance matrix.

**DISPLAYGZ <coverage> <gz.file>**

Displays the color map of z values of Gi statistics directly.

<coverage> -- coverage name of the study area to be displayed for the spatially distributed z values.

<gz.file> -- name of input file to store the Gi statistics after GSTATISTIC is run.

**GSTATISTIC <coverage> <ITEM> <wd.file> <gz.file>**

Calculates the Gi statistics and their testing values of the given item (attribute) in the given coverage.

<coverage> -- name of coverage containing the item (attribute) to be analyzed.

<ITEM> -- the specific name of the item (attribute) in the INFO PAT file in the given coverage. All letters must be in upper case.

<wd.file> -- name of the input file where the weight matrix is stored. This weight matrix is determined by a distance radius from the given coverage after WDISTANCE is executed.

<gz..file> -- name of the output file for storing the Gi statistics and their testing values.

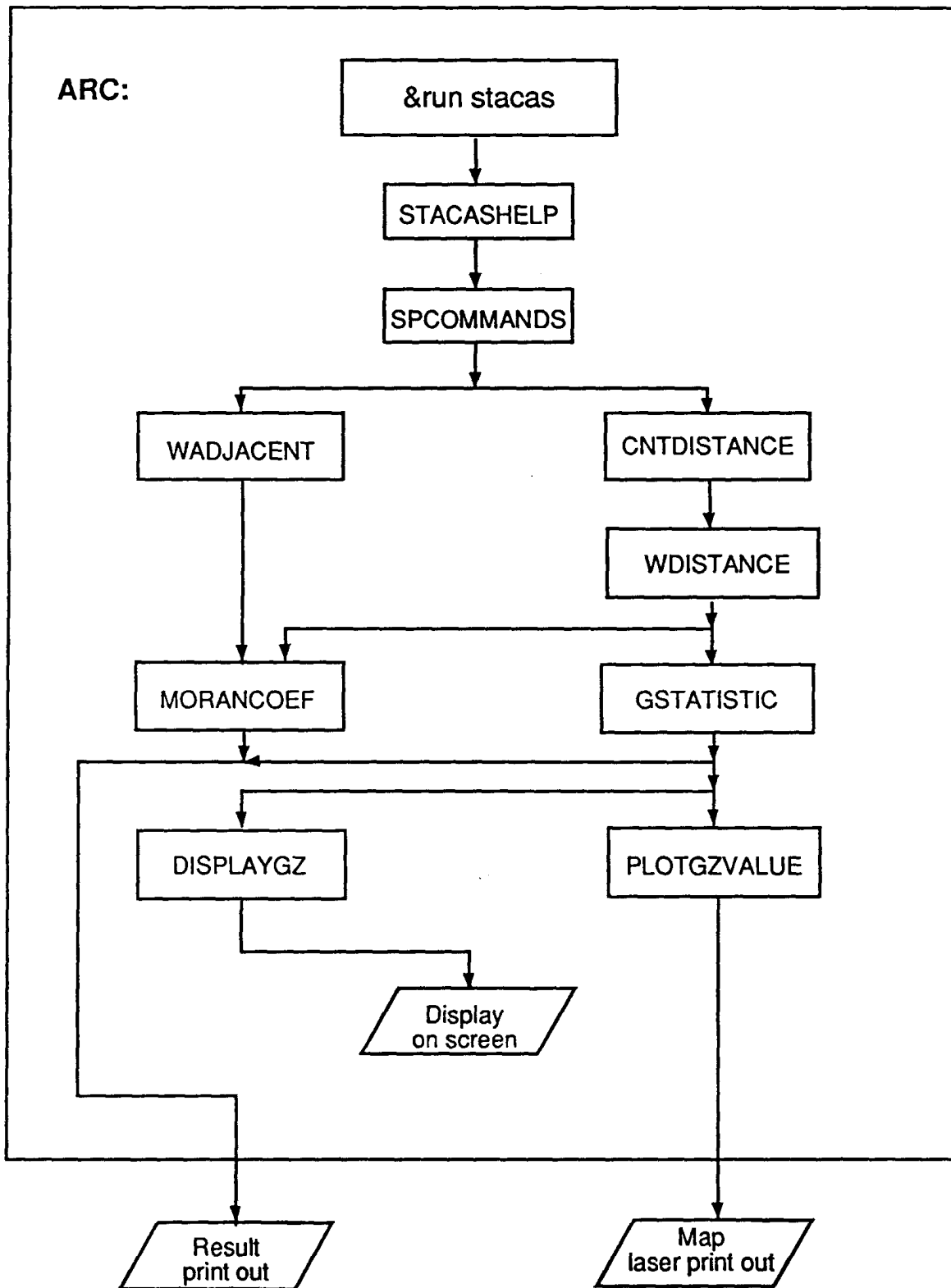


Fig. 4 The Process Diagram of STACAS in ARC/INFO

### **MORANCOEF <coverage> <ITEM> <w.file> <mc.file>**

Calculates the Moran coefficient and associated moments of the given item (attribute) in the given coverage.

<coverage> -- name of coverage containing the item (attribute) to be analyzed.

<ITEM> -- the specific name of the item (attribute) in the INFO PAT file in the given coverage. All letters tyedin upper case.

<w.file> -- name of the input file in which the weight matrix is stored. The weight matrix exists following the either of the commands WADJACENT or WDISTANCE.

<mc.file> -- name of the output file for storing the results.

### **PLOTGZVALUE <coverage> <gz.rile> <map.pos>**

Creates a postscript file for laser printing the map of z value from  $G_i$  statistics directly.

<coverage> -- coverage name of study area to be plotted for the spatially distributed z values.

<gz.file> -- name of the input file containing the results of running the GSTATISTIC command.

<map.pos> -- name of the output postscript file of map of z values which can be laser printed directly.

### **SPCOMMANDS**

To display the list of all commands and their functions in the STACAS on the screen.

### **STACASHELP**

Help system for running STACAS.

### **WADJACENT <coverage> <wm.file>**

Creates the weight matrix by the topological adjacent relationship in the given coverage.

<coverage> -- coverage name of study area contained the polygons to be analyzed.

<wm.file> -- name of output file for storing the result of weight matrix.

### **WDISTANCE <dist.file> <wd.file>**

Creates the weight matrix by the given distance matrix and a distance radius.

<dist.file> -- name of the input file containing the distance matrix after the command CNTDISTANCE.

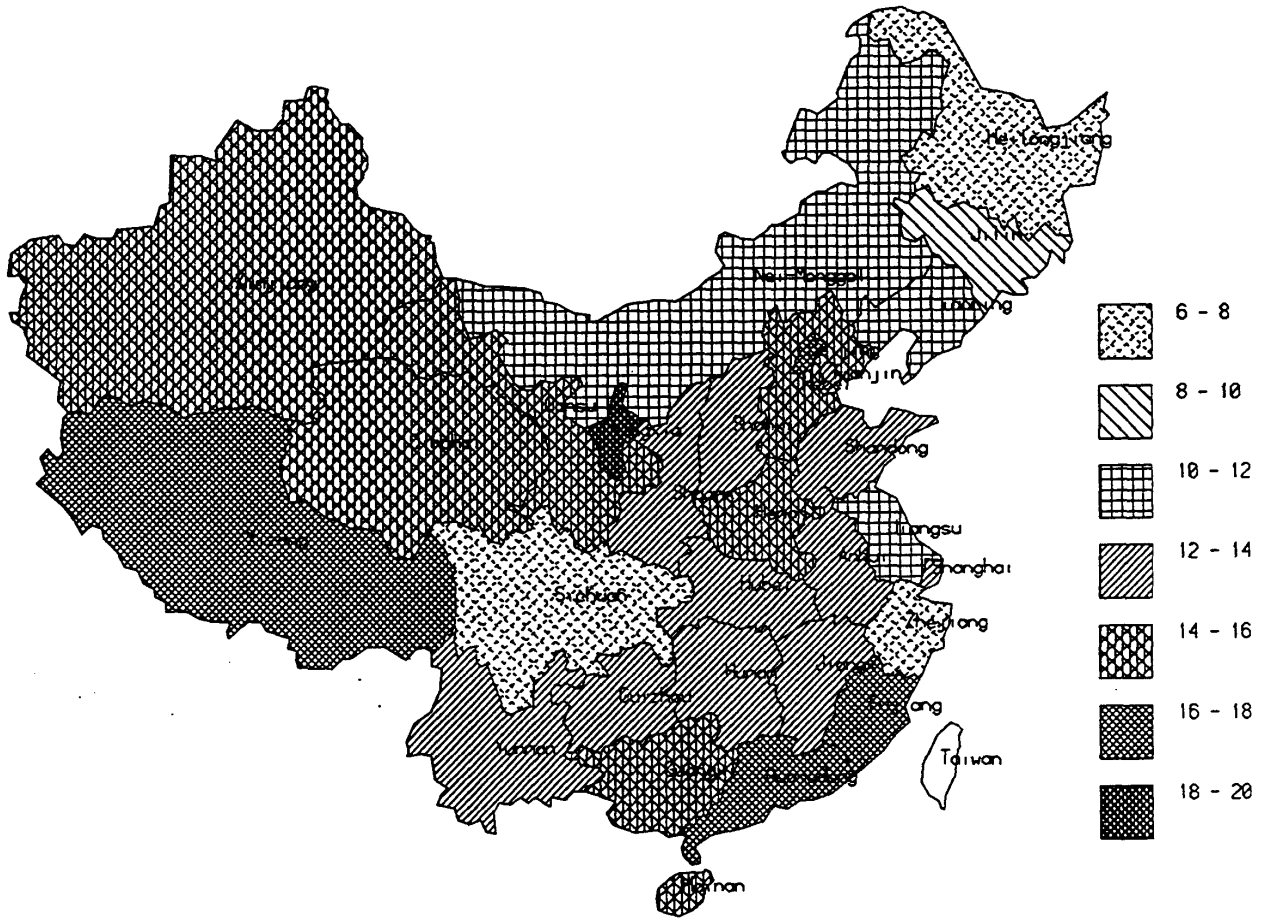
<wd.file> -- name of the output file for storing the results.

## **5. AN EXAMPLE OF RUNNING STACAS WITH POPULATION DATA**

To demonstrate the type of output generated by STACAS, population growth rates (1982-1990) for the 30 provinces of the People's Republic of China (PRC) were analyzed (State Bureau of Statistics, P.R.China 1990). A map of these growth rates generated by the ARCPLOT command is presented in Figure 5.

The tabular output from the commands MORANCOEF and GSTATISTIC for two different definitions of proximal polygons (distance = 1.5 units and distance = 2.5 units) are presented in Tables 1 and 2, respectively. The spatial distributions of the z-scores associated with the  $G_i$  values for the two distance measures plotted using the PLOTGZVALUE command are presented in Figures 6 and 7.

Fig 5 Population Growth Rates in China (1982-1990)



To appreciate the effect of varying the distance within which zones are declared to be neighbors, we present the connectivity matrices for the China map when  $d=1.5$  and  $d=2.5$  in Tables 3 and 4 respectively.

Under both definitions of proximity, the Moran coefficient is significantly positive (at the 95 % confidence level) under both the normality and randomization assumptions indicating a trend in the data whereby a high growth in one zone is associated with high values in neighboring zones and a low value in one zone is associated with low values in neighboring zones. However, the findings of the G statistic are much more sensitive to the definition of proximity: whereas a reasonable proportion of the values are significant when a restricted definition of proximity is employed ( $d=1.5$ ), none are significant when a broader definition of proximity is used ( $d=2.5$ ). The G values that are significantly different from zero are all negative (see Figure 6) and are just significant indicating a pattern dominated by a clustering of medium to low values rather than a pattern dominated by a clustering of high values (Getis, 1990). The medium to low growth rates are primarily clustered around the northeastern and east central provinces.



TABLE 1 MORAN'S I AND  $G_i$  STATISTICS AND THEIR Z VALUES ( $d=1.5$ )

atb#	Moran I	E(I)	VARN(I)	VARR(I)	ZNORMAL	ZRANDOM
1	0.125	-0.034	0.005	0.005	2.319	2.228

ID#	$G_i(d)$	E( $G_i$ )	VAR( $G_i$ )	ZVALUE
1	0.065	0.069	0.000	-0.208
2	0.157	0.207	0.000	-2.688
3	0.000	0.000	0.000	0.000
4	0.111	0.138	0.000	-1.464
5	0.208	0.241	0.000	-1.796
6	0.178	0.172	0.000	0.307
7	0.342	0.379	0.000	-2.012
8	0.307	0.345	0.000	-2.035
9	0.375	0.414	0.001	-2.067
10	0.103	0.103	0.000	-0.028
11	0.340	0.379	0.000	-2.099
12	0.386	0.414	0.001	-1.476
13	0.270	0.276	0.000	-0.300
14	0.028	0.034	0.000	-0.357
15	0.394	0.448	0.001	-2.895
16	0.440	0.483	0.001	-2.281
17	0.370	0.414	0.001	-2.360
18	0.419	0.448	0.001	-1.558
19	0.258	0.241	0.000	0.870
20	0.573	0.552	0.001	1.134
21	0.322	0.310	0.000	0.623
22	0.370	0.345	0.000	1.373
23	0.440	0.414	0.001	1.389
24	0.394	0.379	0.000	0.764
25	0.117	0.103	0.000	0.753
26	0.274	0.241	0.000	1.767
27	0.327	0.310	0.000	0.907
28	0.229	0.207	0.000	1.177
29	0.175	0.172	0.000	0.131
30	0.272	0.241	0.000	1.645

TABLE 2 MORAN'S I AND  $G_i$  STATISTICS AND THEIR Z VALUES (d=2.5)

atb#	Moran I	E(I)	VARN(I)	VARR(I)	ZNORMAL	ZRANDOM
1	0.044	-0.034	0.001	0.001	2.399	2.336

ID#	$G_i(d)$	E( $G_i$ )	VAR( $G_i$ )	ZVALUE
1	0.196	0.241	0.000	-1.692
2	0.523	0.552	0.001	-1.055
3	0.063	0.069	0.000	-0.213
4	0.309	0.345	0.000	-1.332
5	0.477	0.517	0.001	-1.504
6	0.522	0.552	0.001	-1.115
7	0.696	0.724	0.001	-1.038
8	0.617	0.655	0.001	-1.399
9	0.713	0.759	0.001	-1.709
10	0.328	0.310	0.000	0.637
11	0.655	0.690	0.001	-1.291
12	0.796	0.828	0.001	-1.174
13	0.589	0.621	0.001	-1.170
14	0.180	0.172	0.000	0.265
15	0.819	0.828	0.001	-0.303
16	0.776	0.793	0.001	-0.632
17	0.741	0.759	0.001	-0.637
18	0.778	0.793	0.001	-0.577
19	0.634	0.621	0.001	0.504
20	0.817	0.828	0.001	-0.393
21	0.698	0.724	0.001	-0.978
22	0.636	0.655	0.001	-0.715
23	0.728	0.724	0.001	0.152
24	0.755	0.759	0.001	-0.131
25	0.409	0.379	0.000	1.111
26	0.722	0.690	0.001	1.190
27	0.624	0.621	0.001	0.139
28	0.558	0.517	0.001	1.506
29	0.466	0.448	0.001	0.652
30	0.591	0.552	0.001	1.454

Fig 6 Spatial Association of Population Growth in China (d=1.5)

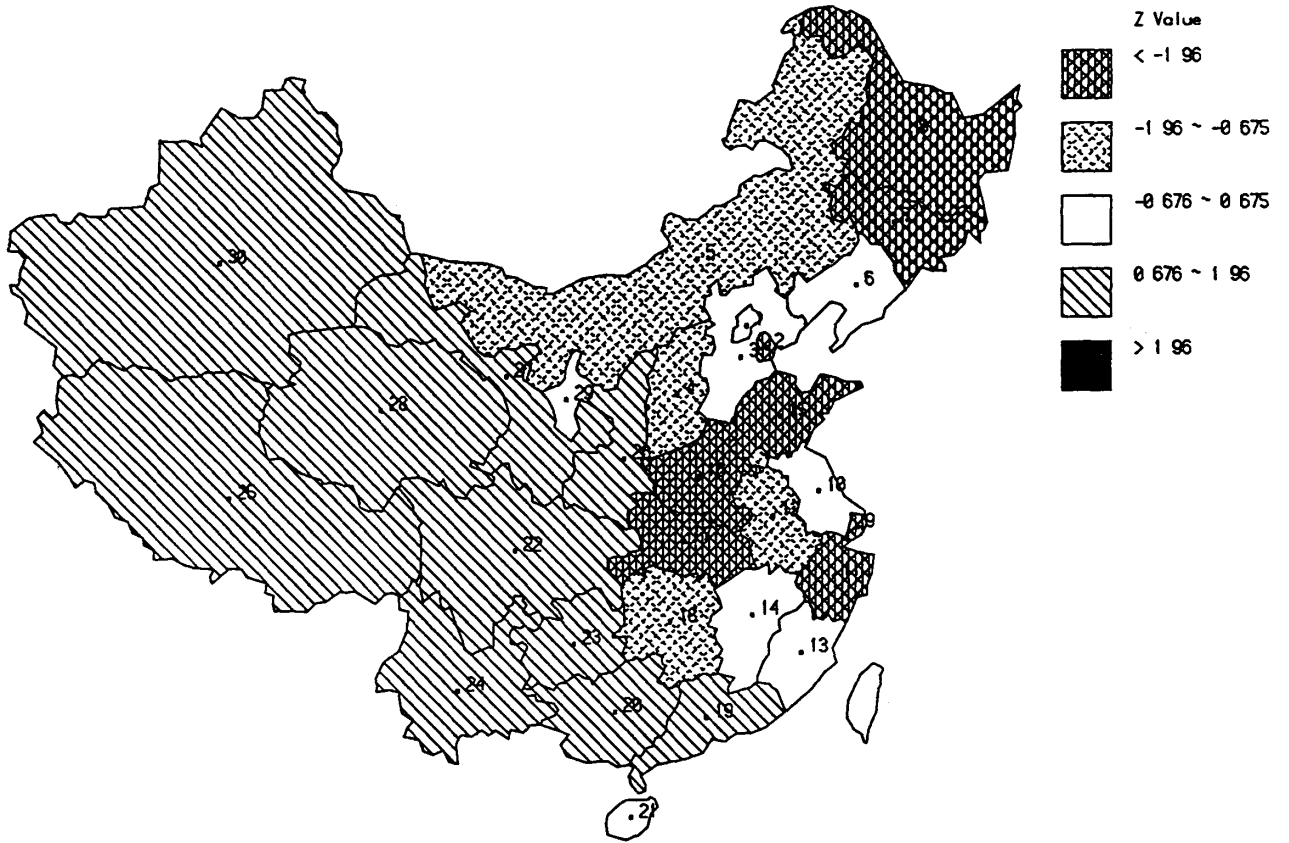
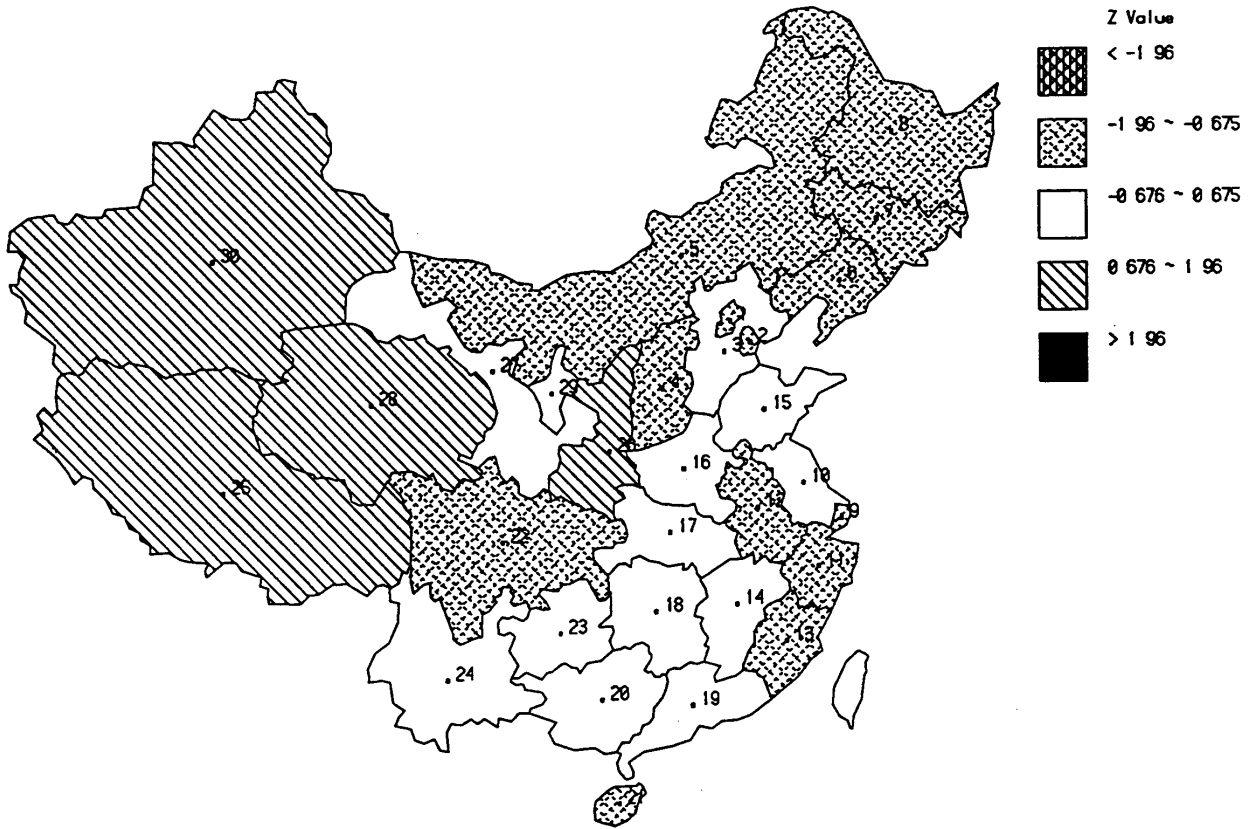


Fig 7 Spatial Association of Population Growth in China (d=2.5)







## 6. SUMMARY AND CONCLUSIONS

The STACAS module that has been described here and which is available from the authors has the following attributes:

1. It makes the spatial autocorrelation and association analysis feasible in ARC/INFO;
2. It can deal with coverages, INFO files, items in INFO files, and data files directly, all of which facilitates spatial analysis within the GIS environment. It also can analyze any attribute within a given coverage;
3. It has a user-friendly interface using the same syntax as regular ARC commands;
4. It has an on-screen help system which is easy to access;
5. The results of a spatial analytical procedure can be mapped and displayed easily;
6. The user can run the commands of both ARC and STACAS in the same environment.

The development of the STACAS module demonstrates that it is possible to run complex spatial algorithms under ARC/INFO. In a more general sense, the ideas in this paper demonstrate the potential for GISs for uses other than storing, displaying and simple querying of data. While the STACAS module will undergo continuous refinements (for example new routines can be added and a menu-driven system installed), it points the way for a major expansion of the use of GIS technology. In a smaller way it also points out the advantages to spatial analysts of linking routines into the tremendous display and data manipulation facilities of a GIS.

## REFERENCES

- Abler, R. F. 1987. The National Science Foundation National Center for Geographic Information and Analysis, *International Journal of Geographical Information Systems*, 1: 303-326.
- Cliff, A. D. 1975. *Elements of Spatial Structure: A Quantitative Approach*, Cambridge University Press, London, UK.
- Cliff, A. D. and J. K. Ord, 1981 *Spatial Processes: Models and Applications*, Pion: London
- Densham, P.J. and M.F. Goodchild 1989. *Spatial Decision Support Systems: A Research Agenda*, GIS/LIS'89 Proceedings, Orlando, Florida.
- ESRI, 1987. *ARCIINFO User Guide, Vol.1: Geographic Information System Software*. Environmental Systems Research Institute, Redlands, CA.
- ESRI, 1989. *AML User Guide: ARC Macro Language and User Interface Tools*. Environmental Systems Research Institute, Redlands, CA.
- ESRI, 1990. *Understanding GIS: The ARC/INFO Methods*. Environmental Systems Research Institute, Redlands, CA.
- Fotheringham, A.S. and D.W.S. Wong, 1991 *The Modifiable Areal Unit Problem in Multivariate Statistical Analysis*, *Environment and Planning A*, at press.
- Getis, Arthur, 1990. *The Analysis of Spatial Association by Use of Distance Statistics*, The Annual Meeting of the Association of American Geographers, Toronto, Canada, April 19-21.
- Goodchild, M. F., 1986. *Spatial Autocorrelation*, CATMOG 47, Geobooks: Norwich UK.
- Griffith, D. A., 1987. *Spatial autocorrelation: A Primer* AAG Resource Publications in Geography, Washington DC.
- Griffith, D.A., 1989. *Spatial Regression Analysis on the PC: Spatial Statistics Using Minitab*, IMAge Discussion Paper #1, University of Michigan, Ann Arbor, MN.
- Horn, M. et al. 1988. *Design of Integrated Systems for Spatial Planning Tasks*, Proceedings of the third International Symposium on Spatial Data Handling August 17-19, Sydney, Australia.

Kehris, E. 1990. A Geographical Modelling Environment Built Around ARC/INFO, Research Report No 13, North West Regional Research Lab, Lancaster University, UK.

Odland, J. 1987 Spatial Autocorrelation Sage: Beverly Hills, CA

Rhind, D. 1988. A GIS Research Agenda, International Journal of Geographic Information Systems, 2: 23-28.

State Bureau of Statistics, P. R. China, 1990. The 1990 Census Data (11), People's Daily, Overseas Edition, Nov. 7th.

Young, D. L. et al. 1988. Integrating GIS with a Decision Support System for Forest Management, Proceedings of the Eighth Annual ESRI User Conference, Mar. 21-25, Palm Springs, CA.