Use of Space-Filling Curves in Generating a National Rural Sampling Frame for HIV/AIDS Research*

Nina Siu-Ngan Lam and Kam-biu Liu
Louisiana State University

The classical geographic research problem of regionalization and resource allocation is most commonly tackled by means of location-allocation methods. This paper introduces the spatial-order method as an alternative for creating regions or clusters. The spatial-order method utilizes space-filling curves, also known as Peano curves, to determine the nearness or spatial order of areal units, such as counties. Given a capacity constraint, the areal units are grouped consecutively according to their spatial order values. We applied the method to create clusters of rural counties for a national sampling survey of HIV/AIDS patients in the United States. Using the criteria that each cluster had approximately 50 new AIDS cases in 1991–1993 and that contiguity of areal units was maximized, 226 clusters were created from the 1,853 rural counties or health districts. The rural clusters generated by this method have been adopted as the national rural sampling frame in the HIV Cost and Services Utilization Study (HCSUS) being undertaken by RAND. In addition to its simplicity and fast computational speed, the spatial-order method produces satisfactory results. With minor modifications, this method can be an efficient alternative to the location-allocation method for solving a wide variety of locational problems, such as routing, political districting, and facilities location and allocation. This paper also demonstrates how a classical geographic research methodology, with the enhancement of GIS, can contribute to the multidisciplinary study of a pressing societal problem in our nation. Key Words: **clustering, HIV/AIDS, HCSUS, location-allocation models, Peano curves, space-filling heuristics, spatial-order method, fractal geometry, GIS.**

Introduction

Partitioning space into regions is a central theme in geographic research. Some of the common applications are political districting, school zoning, market boundaries delineation, and facility location and allocation. A conventional approach in handling this research problem is the location-allocation method, which delineates regions by optimizing some objective functions, such as minimizing the total travel distance, so that the location of central facilities as well as the allocation of the demand to the facilities can be determined simultaneously. The location-allocation method has been widely used in geographic research with many extensions and variations to solve a variety of locational problems (Cooper 1963; Goodchild and Massam 1969; Scott 1970; Rushton et al. 1973; Ghosh and Rushton 1987). However, the computational complexity involved in the location-allocation method may inhibit its use, especially when the research problem at hand has a short lead time and involves large data sets. In this paper, we introduce the spatial-order method, a method that utilizes space-filling curves, also known as Peano curves, as an alternative for creating regions and clusters (Bartholdi and Platzman 1988; Platzman and Bartholdi 1989). To illustrate the application of this method, we describe its use in developing a national rural sampling frame for the HIV Cost and Services Utilization Study (HCSUS).

The HCSUS is a five-year, large-scale, multidisciplinary study being undertaken by RAND (Shapiro and Bozzette 1994). The main goal of the study is to collect national data to develop national estimates of lifetime direct and indirect costs of HIV treatment and to answer a broad range of questions on HIV cost and service utilization, such as how the

*The authors would like to acknowledge assistance from the state health agencies of the 50 states, the rural research team of the HIV Cost and Services Utilization Study (HCSUS), and RAND. This research is funded by RAND as a component of the HCSUS (Grant no. 95-10). Previous support from the State of Louisiana Board of Regents (Grant no. LEQSF(1990–92)-RD-A-05 and LEQSF(1991–92)ENH-29) is also acknowledged. Rajabhushanam Cherukuri, Wei Zhao, Ke Xiao, and Kui Xu helped in keeping the project on time. The Cartographic Section of the Department of Geography and Anthropology assisted in preparing some figures. This paper could not have been completed without the computing facilities provided by the Computer Mapping Science Laboratory of the Department of Geography and Anthropology at LSU.

Professional Geographer, 48(3) 1996, pages 321–332 © Copyright 1996 by Association of American Geographers.
Initial submission, September 1995; revised submissions, January 1996, February 1996; final acceptance, February 1996.
Published by Blackwell Publishers, 238 Main Street, Cambridge, MA 02142, and 108 Cowley Road, Oxford, OX4 1JF, UK.

use, costs, access, and quality of medical and nonmedical services vary according to locale, health care delivery systems, and patient clinical and demographic characteristics. These data will be collected from a national probability sample of persons with HIV/AIDS and their providers through personal questionnaire interviews. Approximately 3,200 adults with HIV from urban areas and 500 adults from rural areas will be selected for interview. Additionally, about 400 children with HIV will also be sampled for interview. This paper focuses on the rural component of the study.

In order to develop national estimates for rural America, where a rapid increase in AIDS incidence in recent years has been documented (Gardner et al. 1989; Verghese et al. 1989; Gould 1993; Lam and Liu 1994; Lam et al. 1996), a sampling strategy that covers all rural counties had to be constructed. Similar to its urban component, the rural study used a three-stage probability sampling design. In the first stage, a sample of approximately 25 rural counties or clusters of rural counties was selected using the probability proportional to size (PPS) method of random sampling (Andersen et al. 1979). In the second stage, primary care providers from each cluster were sampled. About 20 HIV-infected persons from these providers per cluster were then selected in the third stage. The key characteristic of a probability sample is that each patient in the probability sampling frame has a non-zero probability of entering the sample. The reciprocal of each patient's sampling probability could then be used as his/her weight in the statistical analysis of the data collected for the patient.

Since some rural counties may not have a sufficient number of HIV-infected persons to sample from, it was necessary to group these counties together to create a viable sample. AIDS caseload data by county were used as a proxy, since national HIV data do not exist. Unfortunately, AIDS caseload data at the county level are not available from a single source such as the Centers for Disease Control and Prevention; they must be acquired from individual state health agencies. Thus, developing a national rural sampling frame for sampling HIV-infected persons and providers required two steps. First, county-level AIDS caseload data were collected directly from the health agencies of the 48 conterminous states for the period January 1991 to December 1993. The second step was to perform clustering of rural counties on this newly collected AIDS caseload data.

The spatial-order method was applied to generate the rural clusters. The spatial-order method is based on procedures available from the Arc/Info GIS software, which is also the GIS software base for managing, analyzing, and mapping the present AIDS database (Environmental Systems Research Institute 1994). As shown below, the spatial-order method is fast, simple, and produces satisfactory results. The method is especially useful for the analysis of large data sets such as the national AIDS data used in this study and for studies that have a short lead time. The same methodology can be applied to many other locational problems involving routing and districting (Bartholdi et al. 1983; Bartholdi and Platzman 1988). The clusters generated by this method have subsequently been adopted by the HCSUS to serve as the national rural sampling frame for HIV/AIDS survey. In the following, we describe the procedures involved in the spatial-order method and present the national rural sampling frame. Problems related to the clustering of large-scale data with real constraints are highlighted, along with some suggestions for improving the method for future applications.

Data

Following the same data collection method described in Lam and Liu (1994), new AIDS caseload data by county for the period January 1991 to December 1993 were recently collected by contacting individual state health agencies. Table 1 lists the states, their data-reporting units, numbers of urban and rural counties, and respective AIDS cases for 1991–1993. Of the 50 states, 41 of them report AIDS caseload data by county, and 8 by health districts. Alaska reports AIDS data by municipality, and Vermont is the only state that does not provide any data below the state level. The time period 1991–1993 is used because it constitutes the fourth three-year period in our existing AIDS database (1982–1984, 1985–1987, 1988–1990, and 1991–1993). We aggregated the yearly caseload data into three-year intervals in order to minimize year-to-year

Table 1 *AIDS Cases in 1991–1993 by Urban/Rural County by State*

State	Unit	# Units			AIDS Cases		
		MSA	Non-MSA	Total	MSA	Non-MSA	Total
AL	C	21	46	67	1265	267	1532
AZ	C	5	10	15	1804	71	1875
AR	C	11	64	75	572	291	863
CA	C	34	24	58	35396	302	35698
CO	C	10	53	63	1962	93	2055
CT	C	6	2	8	3000	70	3070
DE	C	2	1	3	508	96	604
DC	C	1	0	1	480	0	480
FL	C	34	33	67	20507	744	21251
GA	C	41	118	159	***	***	5652
ID	D	2	7	9	***	***	99
IL	C	27	75	102	6143	375	6518
IN	C	35	57	92	1394	204	1598
IA	C	10	89	99	903	475	1378
KS	C	9	96	105	517	121	638
KY	D	22	15	37	527	203	730
LA	C	23	41	64	2776	399	3175
ME	C	3	13	16	128	140	268
MD	C	15	9	24	4477	105	4582
MA	C	11	3	14	4476	26	4502
MI	C	25	58	83	1392	103	1495
MN	C	18	89	87	970	66	1036
MS	D	7	9	16	444	409	853
MO	C	22	93	115	2682	279	2961
MT	C	2	55	57	30	63	93
NE	D	6	6	12	234	48	282
NV	C	4	3	17	1110	19	1129
NH	C	3	7	10	150	68	218
NJ	C	21	0	21	9358	0	9358
NM	C	6	27	33	408	79	487
NY	C	37	25	62	33681	581	34262
NC	C	34	66	100	1969	614	2583
ND	D	4	5	9	***	***	14
OH	C	39	49	88	2858	222	3080
OK	C	14	63	77	855	196	1051
OR	C	9	27	36	1212	124	1336
PA	C	33	34	67	5798	267	6065
RI	C	4	1	5	544	42	586
SC	C	16	30	46	1757	579	2336
SD	D	4	3	7	19	22	41
TN	D	26	8	34	1793	272	2065
TX	C	59	195	254	12941	502	13443
UT	C	4	25	29	489	38	527
VT	C	3	11	14	***	***	66
VA	C	62	74	136	2715	326	3041
WA	C	12	27	39	2491	153	2644
WV	C	12	43	55	141	96	237
WI	C	20	52	72	896	156	1052
WY	C	2	21	23	14	42	56
TOTAL		829	1853	2682	178778	10187	188965

C = county
D = district
**** Indicates that AIDS cases were suppressed to maintain confidentiality.*

fluctuation and to make identification of dominant trends and patterns easier (Lam and Liu 1994). This newly acquired data will enable detailed analysis of the spatial and temporal trends of AIDS on a national basis since the beginning of the epidemic.

We adopted the Census Bureau classification and designated counties that belong to an MSA (Metropolitan Statistical Area) as urban and non-MSA counties as rural (U.S. Bureau of the Census 1986, 625–30). A binary value (0 for urban, 1 for rural) indicating whether or not the county is an MSA was assigned to each county, in addition to its FIPS code, county name, 1990 population, and area in square miles. These data come from the *County and City Data Book* (U.S. Bureau of the Census 1994).

States releasing data by health district present a problem, because health districts are aggregates of counties, and they typically contain both MSA and non-MSA counties. In an effort to include the rural counties of these states in the sample and at the same time maintain confidentiality about AIDS caseloads, we asked those states using health districts to provide AIDS caseload data for only the MSA counties in each district, so that we can subtract the MSA counties from the district total and keep the remaining, non-MSA, counties of the health district in the database for clustering. Kentucky, Mississippi, Nebraska, and Tennessee provided the MSA counties data. With assistance from the U.S. Health Resources Services Administration (HRSA), we managed to obtain county data from Georgia and Vermont and MSA county data from Idaho and North Dakota, but only under certain constraints. The data obtained through HRSA, including the counties of Georgia and Vermont and the MSA counties of Idaho and North Dakota, can only be used in constructing the rural sampling frame but results derived from these data cannot be published. Therefore, in this paper, results based on these data are deleted accordingly in order to maintain confidentiality. South Dakota did not provide the requested data. MSA county data for South Dakota were subsequently interpolated based on the trends in the neighboring states.

Additional steps were also taken to ensure that the map boundaries corresponded with the data items. Especially for health districts, the various data items by county must be aggregated to conform to the health district boundaries, with respective MSA counties subtracted from the total. Moreover, because a county or health district may have more than one polygon (e.g., islands, lakes, holes created by subtracting MSA counties from the health districts), it is imperative to change the data items of multiple records into 0's so that they are not counted more than once. Otherwise, the statistics and subsequent clustering would be incorrect. This last step could be easily overlooked, and extra verification steps must be taken to ensure accuracy. The Arc/Info GIS software was employed to edit and manipulate the data, generate clusters, and display the output.

There are 2,682 counties/districts in the database, of which 1,853 are rural. The number of new AIDS cases during these three years, 1991–1993, for all counties is 188,965, of which 10,189 cases occurred in rural areas (i.e., non-MSA counties), about 5.4% of the total.

Method

In developing clusters, the criteria for grouping counties must first be established. Our criteria for developing a rural sampling frame were that: (1) only rural counties, that is, non-MSA counties, were included in the population for clustering; (2) each cluster had approximately 50 AIDS cases for 1991–1993 (which in reality should have had more than 50 patients to sample from, as some HIV patients had not developed AIDS in that time period); and (3) contiguity for each cluster was maximized. Moreover, clusters should be of similar geographic size. Although not included as a prime criterion, the cluster size was important to the logistics of the field interview, as field interviewers would have to travel farther for larger clusters and therefore would cost more time and money. As will be shown later, these criteria were difficult to maintain simultaneously.

Two approaches can be used in solving this clustering problem. The classical approach is to employ the location-allocation method. The location-allocation method refers to a group of models that are used to determine the best or optimal location for one or more facilities so that the service or good is accessible to the population in the most efficient manner. The method optimizes efficiency by simultaneously determining the location of the facilities (i.e., supply) and allocation of people (i.e., demand) to the facilities, hence the term location-allocation. Literature on the location-allocation method is voluminous, as the method has many variations and extensions and has been applied to a variety of locational problems (Ghosh and Rushton 1987; Pirkul and Schilling 1988; Domich et al. 1991). The location-allocation problem is mathematically well-defined, but an exact solution to the problem is almost impossible. Therefore, heuristic algorithms are used to solve such problems. In fact, most of the literature in location-allocation has been on the development of efficient heuristics (Teitz and Bart 1968; Goodchild and Noronha 1983; Densham and Rushton 1992 a,b). For problems involving large data sets, the heuristic solution requires intensive computation with a series of iterations, and computational time increases dramatically with the size of the data set. As for all heuristics, optimal solutions cannot be guaranteed. Horn (1995) reports that in a 1991 study, for example, the application of a branch-and-bound algorithm, a type of location-allocation procedure, to a problem with 139 zones required several days' computation on a Unix workstation. Although significant improvements to some location-allocation algorithms have been made so that their computational time is greatly reduced (Densham and Rushton 1992 a,b), the computational complexity involved in the location-allocation method as well as hardware and software limitations may inhibit its use in large-scale studies that also have a short lead time, such as the present AIDS study.

An alternative that is relatively unknown to geographers is to use space-filling curves, also known as Peano curves, to determine the nearness or spatial order of polygons. The polygons are grouped according to their spatial-order values to form clusters. Space-filling curves, discovered by the mathematician Giuseppe Peano in 1890, are curves that traverse space in a continuous and recursive manner (Fig. 1). Thus, they can be used as a transform between a line and n-dimensional space. Each locational element in the data space is visited once by the unbroken curve, and points close to each other in the curve are close to each other in space (Peano 1973; Peuquet 1984; Bartholdi and Platzman 1988). Space-filling curves have gained increased attention recently because they are considered part of the family of fractal curves (Mandelbrot 1983; Goodchild and Mark 1987). As summarized in Bartholdi and Platzman (1988), this method exploits geometry without the need to compute distance measures repeatedly, therefore it is extremely fast, simple, and easy to program. The data requirement is minimal, as the distances between points are ignored. Finally, the method is robust so that addition or deletion of a point would seldom affect the solution. Bartholdi and Platzman (1988) also present a number of algorithms (one of which is implemented in Arc/Info) and provide an excellent review of the applications of this method, such as the well-known traveling salesman problem, zip code assignment, military target attack sequencing, spatial database organization, and partitioning space into clusters.

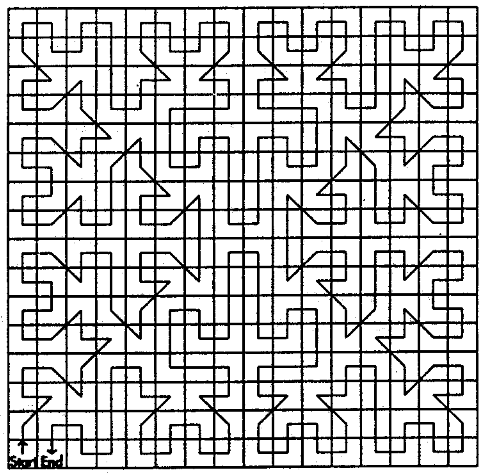

Figure 1: *The exact path of the Peano curve used in Arc/Info as shown on a hypothetical 16 × 16 grid.*

The drawback of the spatial-order method is that it is not very accurate. Platzman and Bartholdi (1989) estimate that the method sometimes produces routes that are 25% longer than optimum. Moreover, the quality of the solution depends heavily on the type of Peano curves used, and theoretical guidelines for selecting a Peano curve for a better solution are still missing.

Both the location-allocation and the spatial-order methods are available in the Arc/Info Version 7.0.3 GIS software. The spatial-order method was used because of its speed, simplicity, and fairly accurate results as reported in the literature. This is important to our present application because of the large data set involved and the limited time frame to complete the study.[1]

In the spatial-order method, the x,y coordinates of the centroids of each county are converted into a single number between 0 and 1 following the path of a Peano curve. The exact path of the Peano curve used in the software is shown on a hypothetical 16 × 16 grid (Fig. 1). The spatial-order value for each county is determined by the total length of the path, which is a factor of how many polygons are involved in the study area. For example, the first polygon might have a value of 0.0025, the second 0.0050; the value keeps increasing until it reaches a value of close to 1.0 at the end of the path. Counties that are close together will have similar spatial-order value. There are some exceptions in that polygons that are close in the xy-space may have quite different spatial-order values, depending on where the curve starts and ends. However, polygons that have similar spatial-order values are always close in xy-space, and polygons that are far apart do not have similar spatial-order values (Environmental Systems Research Institute 1994).

Once the spatial-order value of each polygon is determined, the grouping of counties can be achieved by using the command "collocate" in Arc/Info. Given a capacity of 50 AIDS cases, "collocate" will assign a class number for the rural counties in ascending spatial-order value until the capacity value is exceeded. When the capacity value is exceeded, the class number is increased by one until all the rural counties have been visited, thus resulting in a number of classes of rural counties that will satisfy the criteria specified in our sampling frame. The computer time needed to run these two procedures in a SUN Sparc workstation is negligible.

Results and Discussion

Figure 2 shows the rural counties or districts included in the clustering process. It is clear from Figure 2 that perfect contiguity is impossible to maintain because the rural counties and districts are not completely contiguous. The spatial-order method created 226 clusters. Figures 3–5 map the clusters of rural counties with their cluster number. Table 2 shows the frequency distributions of the clusters according to AIDS caseload and geographical size.

Evaluation of the method and its clustering results must be based on the prime purpose of generating the clusters, which is to serve as a national sampling frame for HIV/AIDS patient survey. As such, the performance of the method can be evaluated on the basis of two

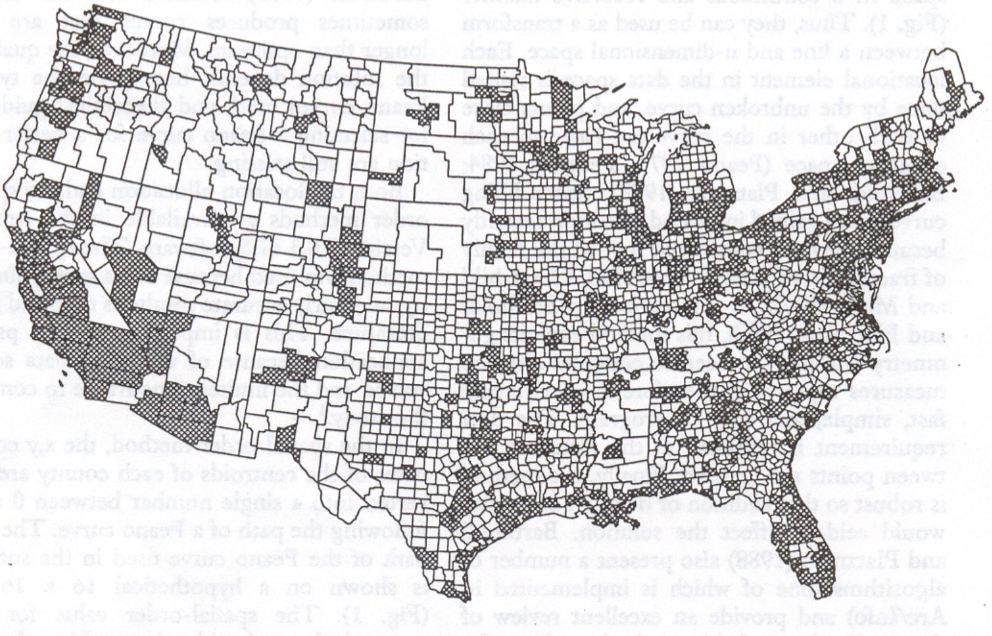

Figure 2: *MSA (shaded) and non-MSA (not shaded) counties and health districts used in the study.*

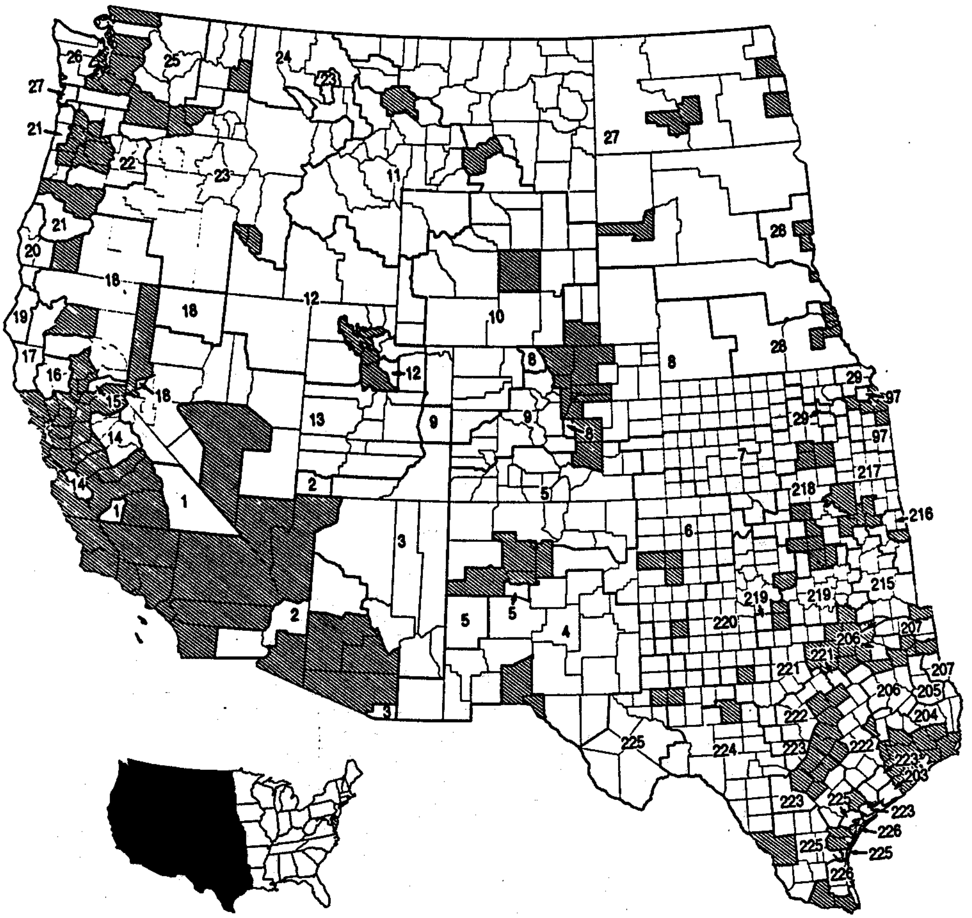

Figure 3: *Clusters in the western United States. Shaded areas are MSA counties or districts. Cluster boundaries for Idaho and North Dakota remain in the map because these clusters include a large number of counties outside these two states so that the inclusion of their cluster boundaries would not affect the confidentiality requirement.*

Table 2 *Frequency Distribution of AIDS Caseloads and Size of Clusters*

Caseloads	# Clusters	%	Square Miles	# Clusters	%
≤10	3	1.3	≤10,000	185	82.0
20	8	4.5	20,000	19	8.4
30	10	4.4	30,000	2	0.9
40	46	20.4	40,000	7	3.1
50	147	65.0	50,000	3	1.3
60	2	0.9	60,000	0	0.0
70	3	1.3	70,000	3	1.3
80	2	0.9	80,000	3	1.3
90	1	0.4	90,000	1	0.4
100	0	0.0	100,000	2	0.9
>100	4	1.8	>100,000	1	0.4
Total	226	100.0	Total	226	100.0

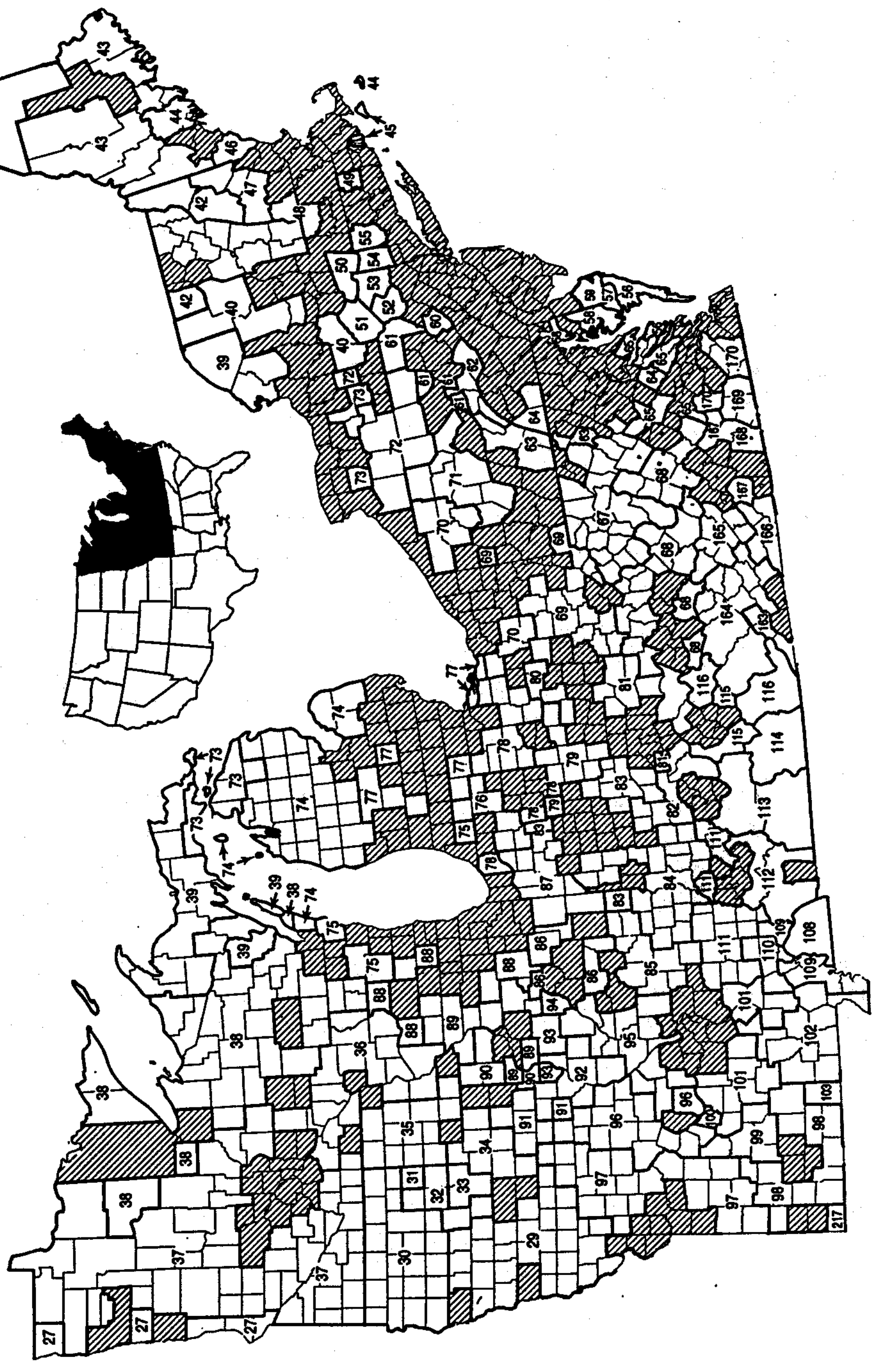

Figure 4: *Clusters in the northeastern United States. Cluster boundaries for Vermont are omitted due to confidentiality requirements.*

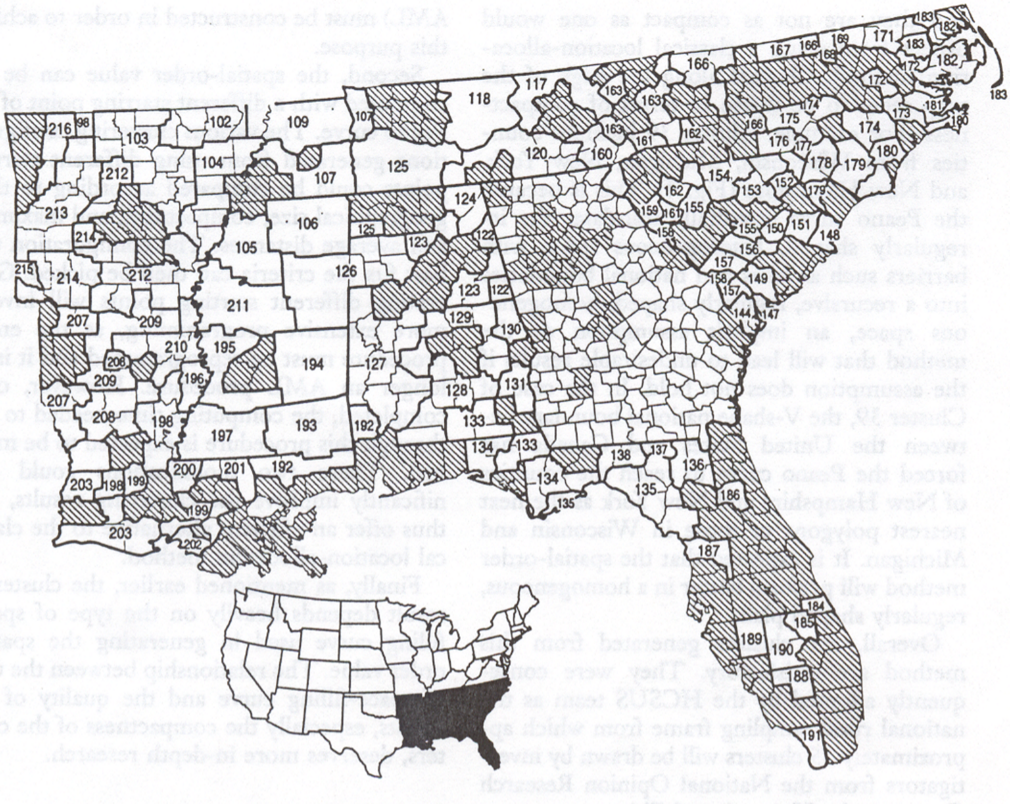

Figure 5: *Clusters in the southeastern United States. Cluster boundaries for Georgia are omitted due to confidentiality requirements.*

factors: the computational aspect of the algorithm and the quality of the results in accord with the criteria set for clustering. As for the computational factor, the spatial-order method is no doubt very fast and easy to understand. The latter is also important to the present study, as the procedures, advantages, and disadvantages of the method can be easily grasped and evaluated by other HCSUS researchers on the team. It was also found in our experiment that minor modifications of the database (e.g., corrections made to the database due to some miscodings) altered the clustering results only in local areas. In this sense, the method is quite robust.

The most important criterion in terms of the quality of the clusters is to create clusters that have similar AIDS caseloads. The results from the spatial-order method are considered acceptable, with about 85% of the clusters having AIDS cases between 40–60 (Table 2). Clusters that have high numbers of AIDS cases are generally a single county or district that by itself has a high AIDS caseload. These clusters are also typically of small geographical size. Cluster size varies, but 82% of the clusters have an area less than 10,000 square miles. The small clusters are generally found in the more populated regions such as the east and southeast. Large clusters are formed owing to the small number of AIDS cases in these counties, and they occur mostly in the west and midwest.

The clustering results show that some of the criteria set above are difficult to meet simultaneously. Perfect contiguity of clusters is impossible to maintain due to the inherent noncontiguous configuration of the rural counties. Also, when the capacity of 50 AIDS cases is imposed, geographically large clusters are inevitably formed in sparsely populated rural areas, owing to their small number of AIDS cases. The clusters are also irregularly shaped and they are not as compact as one would expect from using a classical location-allocation method. Clusters along the edge of the map seem to be worse in terms of compactness. For example, Cluster 39 includes counties from Wisconsin, Michigan, New York, and New Hampshire (Fig. 4). This is because the Peano curve essentially translates the irregularly shaped, heterogeneous space with barriers such as lakes and national boundaries into a recursive, regularly shaped, homogeneous space, an implicit assumption of the method that will lead to undesirable results if the assumption does not hold. In the case of Cluster 39, the V-shape national boundary between the United States and Canada has forced the Peano curve to reach the counties of New Hampshire and New York as the next nearest polygons to those in Wisconsin and Michigan. It is expected that the spatial-order method will perform better in a homogeneous, regularly shaped plain.

Overall, the clusters generated from this method are satisfactory. They were consequently adopted by the HCSUS team as the national rural sampling frame from which approximately 25 clusters will be drawn by investigators from the National Opinion Research Center at the University of Chicago.

Our clustering results show that a reasonable solution can be achieved quickly by the spatial-order method. Three specific modifications or extensions of the method are suggested here to make future applications more flexible and accurate. First, although the "collocate" procedure utilizes a capacity constraint, the procedure may be too stringent in that clusters not meeting the capacity constraint may occur more frequently than necessary. For example, consider two consecutive polygons in terms of their spatial order; the first one has 30 AIDS cases and the second has 25 cases. Instead of forming into one cluster that has a sum of 55, the "collocate" command will form two different clusters, since the addition of the second exceeds the capacity value of 50. The "collocate" procedure can be modified by specifying a tolerance of 40–60 AIDS cases instead of a stringent 50 cases as the cutoff point. This more flexible criterion will allow more clusters meeting the capacity requirement. However, special computer program utilizing the software's macro language (called AML) must be constructed in order to achieve this purpose.

Second, the spatial-order value can be recomputed with a different starting point of the Peano curve. The various clustering configurations generated from using different starting points could be compared according to their geographical size, compactness, and maximum and average distances. The configuration that best fits the criteria can then be picked. Generating different starting points will involve more extensive programming, as the entire procedure must be reprogrammed and it is no longer an AML procedure. However, once completed, the computing time needed to run through this procedure is expected to be minimal. These two modifications could significantly improve the clustering results, and thus offer an efficient alternative to the classical location-allocation method.

Finally, as mentioned earlier, the clustering result depends heavily on the type of space-filling curve used in generating the spatial-order value. The relationship between the type of space-filling curve and the quality of the results, especially the compactness of the clusters, deserves more in-depth research.

Conclusion

The driving force behind the development of a sampling frame for the HCSUS is the quest for unbiased national estimates on HIV/AIDS cost and utilization. A major characteristic of the HCSUS is to overcome the drawback of most previous studies on the HIV/AIDS epidemic in which only subpopulations were included, making inference to the national level impossible. Without a national sampling frame that can cover all parts of the rural United States, it is impossible to select a representative national sample. The generation of rural clusters to serve as a national rural sampling frame is thus an attempt to compromise between sampling accuracy and efficiency, where every rural county in the conterminous states has a chance to be included in the sample. In this sense, this paper serves as a medium to bring to the attention of the research community the importance of research design, sampling strategies, and statistical inference. It also demonstrates how a classical geographic research methodology, with enhancement from GIS, can contribute to a large-scale national study on a pressing problem in our society.

We have demonstrated in this paper the use of the spatial-order method as an efficient and practical alternative in generating clusters of rural counties for the entire nation for HIV/AIDS research. The rural clusters generated from the spatial-order method provide the national rural sampling frame for the HCSUS for further sampling. The spatial-order method is a useful alternative to the more conventional location-allocation method that has been more commonly used in locational research. We have also shown that a classical locational problem such as the present study can be solved effectively in a GIS setting, and that GIS can be applied efficiently to the study of a pressing societal issue such as AIDS. The present spatial-order method can be improved by specifying a more flexible criterion value (such as a range of cases instead of a fixed number) for cluster formation, by introducing different starting points of the Peano curve, and by considering different curve forms. The main advantage of this method is that it is fast, simple, and has the flexibility of including a capacity constraint; thus it is especially useful for the analysis of large, national data sets and for studies that have a short lead time. The spatial-order method is applicable to a wide range of locational problems, such as bus routing, political districting, health delivery system zoning, and facility location and allocation. ■

Note

[1]We also applied the location-allocation method to the present problem using the minimum distance criterion model in Arc/Info. The heuristic used is the Global/Regional Interchange Algorithm (GRIA) proposed by Densham and Rushton (1992 a,b) and implemented as default in Arc/Info. The objective function is to find the best locations of centers (rural counties) so that nearby rural counties are allocated to their closest centers. Unfortunately, none of the six models available in Arc/Info provides an option of enforcing a capacity constraint (i.e., 50 AIDS cases), and as such, the resultant clusters have varying capacities. This proves to be a major problem for the present purpose of creating clusters of similar numbers of AIDS cases. Consequently, the HCSUS group adopted the clusters generated by the spatial-order method. It should be noted that in our experiment, each run of the location-allocation method with approximately 2,000 polygons (1,853 rural counties or districts plus duplicate polygons) took approximately 12–24 hours on a multiuser SUN Sparc20 workstation, depending on the time of the day the job was submitted. While computational speed is not the prime reason for not selecting the location-allocation method and the speed could possibly be improved through system reconfiguration, alternative procedures that can generate clusters faster with acceptable results need to be explored.

Literature Cited

Andersen, R., J. Kasper, and M. Frankel, eds. 1979. *Total Survey Error: Applications to Improve Health Surveys.* San Francisco: Jossey-Bass.

Bartholdi, J. J., III, and L. K. Platzman. 1988. Heuristics based on spacefilling curves for combinatorial problems in Euclidean space. *Management Science* 34(3):291–305.

Bartholdi, J. J., III, L. K. Platzman, R. L. Collins, and W. H. Warden, III. 1983. A minimal technology routing system for meals on wheels. *Interface* 13(3):1–8.

Cooper, L. 1963. Location-allocation problems. *Operation Research* 11(3):331–43.

Densham, P. J., and G. Rushton. 1992a. Strategies for solving large location-allocation problems by heuristic methods. *Environment and Planning A* 24:289–304.

———. 1992b. A more efficient heuristic for solving large *P*-median problems. *Papers in Regional Science* 71(3):307–29.

Domich, P. D., K. L. Hoffman, R. H. F. Jackson, and M. A. McClain. 1991. Locating tax facilities: A graphics-based microcomputer optimization model. *Management Science* 37(8):960–79.

Environmental Systems Research Institute. 1994. Arc/Info Version 7.03 online Help Menu. Redlands, CA: ESRI.

Gardner, L. I., J. F. Brundage, D. S. Burke, J. G. McNeil, R. Visintine, and R. N. Miller. 1989. Evidence for spread of the human immunodeficiency virus epidemic into low prevalence areas of the United States. *Journal of Acquired Immune Deficiency Syndrome* 2:521–32.

Ghosh, A., and G. Rushton. 1987. *Spatial Analysis and Location-Allocation Models.* New York: Van Nostrand Reinhold.

Goodchild, M. F., and D. M. Mark. 1987. The fractal nature of geographic phenomena. *Annals of the Association of American Geographers* 77(2):265–78.

Goodchild, M. F., and B. H. Massam. 1969. Some least-cost models of spatial administrative systems in Southern Ontario. *Geografiska Annaler* 52B:86–94.

Goodchild, M. F., and V. Noronha. 1983. *Location-allocation for small computers.* Monograph No. 8.

Iowa City: Department of Geography, University of Iowa.

Gould, P. 1993. *The Slow Plague: A Geography of the AIDS Pandemic.* Cambridge, MA: Blackwell Publishers.

Horn, M. E. T. 1995. Solution techniques for large regional partitioning problems. *Geographical Analysis* 27(3):230–48.

Lam, N. S.-N., and K-b. Liu. 1994. Spread of AIDS in rural America, 1982–1990. *Journal of Acquired Immune Deficiency Syndrome* 7(5):485–90.

Lam, N. S.-N., M. Fan, and K-b. Liu. 1996. Spatial-temporal spread of the AIDS epidemic: A correlogram analysis of four regions of the United States. *Geographical Analysis* 28(2):93–107.

Mandelbrot, B. B. 1983. *The Fractal Geometry of Nature.* San Francisco: W. H. Freeman and Co.

Peano, G. 1973. *Selected Works*, ed. H. C. Kennedy. Toronto: University of Toronto Press.

Peuquet, D. J. 1984. A conceptual framework and comparison of spatial data models. *Cartographica* 21(4):66–113.

Pirkul, H., and D. A. Schilling. 1988. The siting of emergency service facilities with workload capacities and backup service. *Management Science* 34(7):896–908.

Platzman, L. K., and J. J. Bartholdi, III. 1989. Spacefilling curves and the planar travelling salesman problem. *Journal of the Association for Computing Machinery* 36(4):719–37.

Rushton, G., M. F. Goodchild, and L. M. Ostresh, eds. 1973. *Computer Programs for Location-Allocation Problems.* Monograph No. 6. Iowa City: Department of Geography, University of Iowa.

Scott, A. J. 1970. Location-allocation systems: A review. *Geographical Analysis* 2:95–119.

Shapiro, M., and S. Bozzette. 1994. *HIV Cost and Services Utilization Study (HCSUS).* Abstract of the HCSCUS proposal. Santa Monica, CA: RAND.

Teitz, M. B., and P. Bart. 1968. Heuristic methods for estimating the generalized vertex median of a weighted graph. *Operations Research* 16:955–61.

U.S. Bureau of the Census. 1986. *State and Metropolitan Area Data Book: 1986.* Washington, DC: U.S. Government Printing Office.

———. 1994. *County and City Data Book: 1994.* Washington, DC: U.S. Government Printing Office.

Verghese, A., S. L. Berk, and P. Sarubbi. 1989. Urbs in rure. Human immunodeficiency virus infection in rural Tennessee. *Journal of Infectious Disease* 160:1,051–55.

NINA SIU-NGAN LAM is Professor of Geography in the Department of Geography and Anthropology and Adjunct Professor in the Department of Epidemiology and Community Health at Louisiana State University, Baton Rouge, LA 70803. Her current research interests include AIDS modeling, fractals and scale in global change studies, and use of GIS in medical geography and environmental modeling.

KAM-BIU LIU is Professor of Geography in the Department of Geography and Anthropology at Louisiana State University, Baton Rouge, LA 70803. His research interests are in Quaternary paleoecology, paleoclimatology, global change, biogeography, and medical geography.