



THESIS

This is to certify that the

thesis entitled

Areal Data Reaggregation: A Comparison of Two Methods

presented by

Gustave William Rylander

has been accepted towards fulfillment of the requirements for

Master of Artsdegree in Geography

Ly M. Olson Major professor

Date_____10/28/86

O-7639

MSU is an Affirmative Action/Equal Opportunity Institution



RETURNING MATERIALS: Place in book drop to remove this checkout from your record. <u>FINES</u> will be charged if book is returned after the date stamped below.

AREAL DATA REAGGREGATION: A COMPARISON OF TWO METHODS

By

Gustave William Rylander

A THESIS

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

MASTER OF ARTS

Department of Geography

ABSTRACT AREAL DATA REAGGREGATION: A COMPARISON OF TWO METHODS

By

Gustave William Rylander

Two methods for reaggregating data between sets of areal units, "source units" and "target units", are compared. With the overlay method, the value for each source unit is allocated among the target units in proportion to its area of overlap with each target unit. The pycnophylactic method first interpolates an intermediate surface from the source units and then aggregates it into the target units. A previous case study with census tracts and planning districts in London, Ontario, revealed no significant differences in their performance. Two factors that may lead to better performance by one method or the other, however, are the size of the units and their delineation criteria. The performance of the methods in estimating populations of the 48 contiguous states from those of 181 economic areas is compared with their performance in the previous study. Despite the fact that pycnophylactic interpolation resulted in a closer approximation to the actual surface, the difference in the performance of the methods was not significant. A difference was observed, however, between the relative success of the methods in the two case studies.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Judy Olson, and my other committee members, Bruce Pigozzi and John Hunter, for their valuable comments, criticisms, and support during the development of this thesis. My department chairman, Gary Manson, was also very patient and supportive, and I extend my thanks to him. Cheers and the best of luck to three friends and fellow graduate students: Ann Goulette, Charlie Johnston, and Kim Medley. Bruce, Murray, and Matt provided me with a comfortable home and plenty of pizza during my last month at Michigan State. Finally, my parents, Gus and Nancy Rylander, offered their constant support, and I want to express my gratitude and love to them.

TABLE OF CONTENTS

List of Tables	i
List of Figures	ii
Chapter I. Introduction	1
Chapter II. Previous Research on Areal Data Reaggregation	4
A. Development of	-
Areal Data Reaggregation Methods	4
1. The Overlay Method Method	
3. The Pycnophylactic Method	8
B. Variations on the Methods	11
C. Empirical Comparison of the	
Areal Data Reaggregation Methods	13
D. Applications and Adaptations of the Methods	15
E. Summary	16
Chapter III Becertah Wypetherer and Objectiver	17
A. Effect of Size of Source Units on	1/
Areal Data Reaggregation	17
B. Effect of Source Unit Delineation Criteria	- /
on Areal Data Reaggregation	18
C. Research Objectives	18
-	
Chapter IV. Research Methodology	23
A. Data Sources and Compilation	23
B. Rasterization Methods	
and the Loss of County Data	24
C. Implementation of the	25
Rieal Data Reaggregation Methods	20
Chapter V Pesults	27
A. Convergence of the Pvcnophylactic	
Interpolation Process	27
B. Goodness-of-Fit of the Estimating Surfaces	28
C. Target Unit Estimates and Errors	33
Chapter VI. Summary and Conclusions	45
Chapter VII. Recommendations for Further Study	48
Bibliography	5Ø

LIST OF TABLES

2.1.	Population Data for Planning Districts	14
2.2.	Summary Statistics for Actual and Estimated Values	
	for Planning Districts	15
3.1.	Source Unit Delineation Criteria	
	A. Census Tracts: London, Ontario	21
	B. Economic Areas: United States	21
5.1.	Population Data for States	36
5.2.	Summary Statistics for Actual and Estimated Values	
	for States	38
5.3.	Relationship of Success of Reaggregation Methods to	
	Source Units Used	44

.

LIST OF FIGURES

.

2.1.	Interpolating a Smooth, Pycnophylactic Surface:	10
~ ~	Refeate of Volume of Volumbaring Units on	12
2.2.	Effects of values of Neighboring Units on	
	Pychophylactic Surface Configuration	
	a. Trough	12
	D. Slope	12
•	c. Peak	.12
3.1.	Source Units and Target Units	
	a. Source Units: Economic Areas	2Ø
	b. Target Units: 48 Contiguous States	2Ø
	c. Source Units and Target Units Superimposed	2Ø
3.2.	Idealized Surface Configuration	
	a. Within an Economic Area	21
	b. Within a Census Tract	21
5.1.	Convergence of the Pycnophylactic Interpolation	
	Process	
	a. Change in Source Unit Population	
	Due to Volume-Adjustment	зø
	b. Change in Estimating Surface	50
	Between Successive Smoothings	20
5 2	Error in the Estimating Surfaces	20
5.2.	Error in the Estimating Surfaces	שכ
5.3.	Error in the Estimating Surfaces	~ 1
	a. Positive Errors: Economic Areas	31
	D. Negative Errors: Economic Areas	32
	c. Positive Errors: Pycnophylactic Surface	33
	d. Negative Errors: Pycnophylactic Surface	34
5.4.	Target Unit Population Density	
	a. Population Density, 1980: Actual	38
	b. Population Density, 1980: Overlay Method	39
	c. Population Density, 1980: Pycnophylactic Method.	39
5.5.	Target Unit Error	
	a. Positive Error: Overlay Method	40
	b. Negative Error: Overlay Method	40
	c. Positive Error: Pycnophylactic Method	41
	d. Negative Error: Pycnophylactic Method	41

Geographers often use data that are associated with areal units. The boundaries of different sets of overlapping units are usually delineated by different institutions for their own purposes and do not always coincide. A result of this non-coincidence is that the sets of data that are available for a region may not be compatible. This is a problem for the researcher who needs to relate the data from different sets of units either statistically or visually on maps. For example, the hypothesis that income is geographically related to political behavior is difficult to test empirically when the boundaries for census tracts do not respect those of political wards. Assuming that a census of one of the sets of units is not feasible, the solution is to estimate the data for one set, the target units, from those of the other set, the source units. The term areal data reaggregation is used here to refer to the entire family of methods by which values for target units are estimated from source units.

Previous research (Lam, 1980) has shown that two of the available methods of areal data reaggregation, the <u>overlay</u> and <u>pycnophylactic</u> methods, are clearly better than the others. Overlay reaggregates the data directly from the source units into the target units, while the pycnophylactic method first interpolates an intermediate, continuous surface from the source units and then reaggregates it into the target units. The two methods, then, aggregate

different <u>estimating surfaces</u>, approximations of the actual surface, into the target units. The estimating surface represents an assumption about the configuration of the actual surface.

In a previous study (Lam, 1980) the overlay and the pycnophylactic methods performed nearly identically, and it is not known whether their relative accuracy and precision are stable when units with different spatial characteristics are used. Information about the actual surface is important to the selection of an appropriate data reaggregation method because the configuration of the actual surface affects the validity of the assumptions implied by the use of the different estimating surfaces.

This study isolates two of the factors that may affect the performance of the methods, the <u>size</u> (or aggregation level) of the areal units and the <u>criteria used to</u> <u>delineate their boundaries</u>. The size of the source units may affect the likelihood that the actual distribution is homogeneous within their boundaries. Larger units may be less likely to contain homogeneous distributions. The delineation criteria often provide more detailed information about the configuration of the actual surface, which strongly affects the accuracy of the estimates. In other words, they suggest the kind of estimating surface that may yield good estimates.

The performance of the two reaggregation methods is compared using the results from an earlier study along with parallel information for a new set of units that is clearly

different in these two factors. Although a single comparison cannot greatly improve the confidence with which we choose a method, this study may suggest some characteristics of size and delineation criteria to consider beforehand. CHAPTER II. PREVIOUS RESEARCH ON AREAL DATA REAGGREGATION

Research on the areal data reaggregation problem has yielded three basic reaggregation methods: overlay, contour reaggregation, and the pycnophylactic method. Systematic, if limited, comparisons of their performance have also been made.

A. Development of the Areal Data Reaggregation Methods1. The Overlay Method

Markoff and Shapiro (1973) demonstrated and empirically tested the simplest method of areal data linkage, the overlay method. This method reaggregates data directly from the source units into the target units, taking each target unit value as a weighted sum of the source unit values. The weights are the areas of overlap between each target unit and the source units. The contribution of each source unit to a target unit, then, is proportional to its area of overlap. Different formulas for overlay are necessary depending on whether the source unit data are absolute numbers or percentages.

For absolute numbers,

$$\hat{V}_{i} = \sum_{j} \frac{A_{ij}}{A_{j}} * V_{j}$$
where \hat{V}_{i} is the estimated value for target i
$$V_{j}$$
 is the actual value for source j
$$A_{ij}^{i}$$
 is the area of overlap between i and j
$$A_{j}^{i}$$
 is the area of j.

For percentages,

$$\hat{v}_{i} = \sum_{j} \frac{A_{ij}}{A_{i}} \cdot v_{j}$$

where A_i is the area of i.

Using data for 18th Century French <u>generalites</u> and <u>departments</u>, Markoff and Shapiro estimated the populations of each of these sets of areal units from the other. They made their area measurements with a grid overlay. Using equation 1), the estimates had a correlation of .53 when estimating data for <u>departements</u> from those of <u>generalites</u> and .96 when the direction was reversed. The higher correlation was obtained when estimating the values for the relatively larger <u>generalites</u>. This improvement is expected, because the <u>departements</u> provide a more detailed representation of the actual population density surface. In general, for any areal data reaggregation method, better estimates will be obtained when data are reaggregated from smaller units to relatively larger units.

Errors in the overlay method occur because of intra-unit deviations from the single density value assigned to each of the source units. In other words, the implicit assumption of the overlay method is spatially homogeneous density within the source units. This has been called the <u>choropleth assumption</u>, after the mapping method commonly used for areal data.

Crackel (1975) presented modifications of the overlay formulas for special cases in which the area of overlap is

either less than or greater than the area of the target unit. This can happen when the target unit is not completely covered by source units or when more than one set of source units (overlapping one another) are used to estimate the value of a target unit. He suggested multiplying the weighted sum used in formulas 1 and 2 by the ratio of the target area to the total area of overlap. This correction reduces the estimated value when the target area is smaller than the area of overlap and increases it when the target area is greater, and it should should often reduce the severity of errors.

2. The Contour Reaggregation Method

Ford (1976) contributed an alternative to the overlay method. The contour reaggregation (CR) method involves interpolating values for a regular grid from those given for control points located at the centers-of-gravity of the source units and reaggregating the interpolated values into the target units. The primary difference between CR and overlay, then, is that CR transforms the source units into a smooth, disaggregated surface before aggregating it into the target units. This intermediate step is intended to improve the representation of the actual surface by introducing a degree of spatial autocorrelation into the estimating surface. The value of a spatially-distributed variable at a particular location is usually positively related to nearby values, but the aggregation of the surface into source units hides some of this information. Interpolation can recover some of the lost detail.

Ford's interpolation method was similar to the one used by the SURFACEII program. He fit a second-order trend surface equation to the control points using weighted least squares (weight = $1/d^2$) of the nearest 8 control points. Then the equation was used to predict the values at the nodes of the regular grid. He estimated rent values for 10 postal zones from those of 15 census tracts in Dade County, Florida. Although his sample size was too small to produce a reliable measure of statistical association, he reported a relative root mean squared error (RRMSE) of .255 between the actual and estimated values. RRMSE is a root mean squared error (RMSE) standardized about the mean of the estimated values:

RRMSE =
$$\frac{RMSE}{\overline{E_i}}$$
 = $\sqrt{\frac{1}{N}\sum_{i}^{N} (O_i - E_i)^2} / \frac{1}{N}\sum_{i}^{N} E_i$ 3)

where Ei and Oi are the expected and observed values.

Ford suggested that methods yielding an RRMSE of around .10 could be considered fairly accurate.

Although CR may have a theoretical advantage over the overlay method because of its incorporation of spatial autocorrelation into the estimating surface, it also has a critical disadvantage: interpolation from control points generally does not preserve the total value, or volume, of each of the source units. The common interpolation methods often preserve the values at the control points, but these values are inconsequential when <u>areas</u> are the units of interpolation. If one assumes that the source unit data

were collected without error, then CR adds detail (however realistic) to the surface at the expense of accuracy. The next section traces the progression of research that proved that this trade-off is unnecessary.

3. The Pycnophylactic Method

The problem left for subsequent research into areal data reaggregation was to incorporate smoothness into the estimating surface while preserving the volume of the source units.

The solution to this problem had its origins in work by Boneva, Kendall, and Stefanoff (1971). They introduced a technique for interpolating yearly birth rates from data given in five-year intervals. Aggregated, temporal data of this sort can be displayed graphically as a histogram, with the total births for a five-year period given by the area of an individual bar. Histogram bars, however, hide some of the serial autocorrelation usually exhibited by data collected over time. The authors' objective, then, was to replace the discrete bars with a continuous curve while preserving the total area under the curve as a whole and within five-year intervals. The curve was generated as a mathematical spline, which is essentially a series of simple curves (or functions) pieced together end-to-end. The resulting curve, called a histospline, adds hypothetical detail to the intervals without changing the accuracy of the original, aggregated data.

Tobler and Lau (1978) adapted the technique for the two-dimensional case, i.e., for bivariate histograms. A

bivariate histogram has data associated with rectangular cells instead of intervals on a line. The bars of a univariate histogram are replaced by columns, with the density and total value for each cell represented by its column's height and volume, respectively. Tobler and Lau demonstrated a geographic application of this method using population data compiled by 0.25 square mile grid cells for Ann Arbor, Michigan. When the surface interpolated from these grid cells is threaded with isolines, an <u>isopleth map</u> is produced. The primary motivation for their research, in fact, was the improvement of isopleth mapping methods, not areal data reaggregation.

Tobler (1979) developed the technique further by adapting it to operate on irregular geographic units. He accomplished this by disaggregating the units into a set of small grid cells. In computer terminology, he used a raster data structure. This is the digital counterpart of the physical grid overlay that Markoff and Shapiro used to make their area measurements. Tobler used the term pycnophylactic (volume-preserving) to refer to the new interpolation method. He also provided a useful analogy to enhance its intuitive appeal. He suggested thinking of the original stepped surface as a clay model. If the pycnophylactic interpolation method were to operate physically, it would sculpt the clay such that the surface would become as smooth as possible without moving any clay between neighboring columns.

Tobler's study, then, was the final step in maintaining

source unit accuracy (like overlay) while interpolating realistic detail into the surface (like CR), and the Pycnophylactic method has obvious appeal as a method for areal data reaggregation.

Algorithms for pycnophylactic interpolation are described in Tobler (1979) and Lam (1980, 1983). A simple raster-based algorithm, based on the one used by Lam (1980, 1983), is described below because an understanding of the interpolation process is necessary in order to understand the pycnophylactic estimating surface.

The pycnophylactic interpolation algorithm is shown graphically in figure 2.1. First, the grid cell values used to identify the source units are replaced with the corresponding density values. The surface is then smoothed using an arbitrary smoothing operator (e.g., each grid cell is assigned the mean of its four non-diagonal neighbors). Finally, the total values within the source unit boundaries are recovered by multiplying each grid cell by the ratio of the original source unit value to the new (smoothed) value. This process is repeated an arbitrary number of times. Since the smoothing routine reduces differences in neighboring grid cells, it operates from the edges of the source units (i.e., from the "cliffs" between neighboring units) inward. The monotonic reduction of differences also implies that the process converges; that is, the changes in the surface become smaller with each iteration.

Another characteristic of this interpolation algorithm is that the configuration of the interpolated surface

within source units depends entirely on the densities of neighboring units. Figure 2.2 shows how the locations of maxima and minima within a unit are affected by neighboring units. If all of the neighbors have higher densities, a "trough" will be interpolated within the unit (fig. 2.2a). If the neighbors have a combination of higher and lower densities on opposite sides then the surface will slope from one side of the unit to the other (fig. 2.2b). Finally, a "peak" is interpolated if the neighboring units all have lower densities (fig. 2.2c).

B. Variations on the Methods

The research up to and including Tobler (1979), then, had produced three basic methods for areal data reaggregation: overlay, CR, and pycnophylactic. Variations of all three methods are possible. Overlay can be approximated by using unweighted averages of the overlapping source units or simply the value of the source unit with the largest overlap area. CR can use any of the multitude of methods for interpolating from control points. And the pycnophylactic method can use different smoothing routines and boundary conditions. Overlay uses the source units themselves as the estimating surface, while CR and the pycnophylactic method use interpolated surfaces. The estimating surfaces for overlay and pycnophylactic both preserve the total volume within the source units.



				ERATION 1	
Unit	Area	Initial Pop.	Initial Density	Pop.	Ratio (initial pop. / pop.)
1	6	30	5.0	32,83	0,914
2	13	78	6.0	79.00	0,987
3	15	105	7.0	103,75	1.012
4	6	48	8.0	45.42	1.057

Figure 2.1. Interpolating a Smooth, Pycnophylactic Surface: Iteration 1



Figure 2.2. Effects of Values of Neighboring Units on Pycnophylactic Surface Configuration

C. Empirical Comparison of Areal Data Reaggregation Methods

Although the performance of overlay and CR had been assessed independently (Markoff and Shapiro, Ford), no comparisons of the three methods were made until 1980. Lam (1980) did not use the same interpolation algorithm for CR as Ford but took the values at grid cell nodes as distance-weighted averages of the control points. The weight she used, $2^{-d/B}$, allows arbitrary specification of the value of B. Higher B values result in more weight allocated to distant points and, thus, a smoother surface. To give CR the benefit of the doubt relative to the volume-preserving methods, Lam chose a value of B that minimized the difference between the CR surface and the actual one (\emptyset .15 for her data set). She estimated population data for 21 planning districts from 51 census tracts in London, Ontario, and her results are listed in Table 2.1. The summary statistics in Table 2.2 were derived from these results.

The correlation coefficients from these experiments are indicators of the <u>precision</u> of the estimation method, which is clearly higher for the volume-preserving methods than for the "optimal" CR method. The <u>accuracy</u> of the estimates is indicated by the regression and RMSE statistics. Perfect accuracy would result in a slope of 1, an intercept of Ø, and an RMSE of Ø. Table 2.2 thus also indicates the generally higher accuracy of the volume-preserving methods. Finally, Lam observed that the highest errors for the

1980)
Lam,
(Source:
Districts
Planning
for
Data
Population
2.1.
Table

tual)	с С С	- 0 <u>.5</u> 19	-0.059	1.439	-0.419	-0.275	-0.297	-0.700	-0.256	-0.428	0.434	0.216	5.234	Ø.663	0.337	89.479	0.469	-0.517	-0.530	0.193	-0.449	0.730
S (& AC	Pycnc	-0.414	0.071	1.541	0.102	0.042	-0.109	0.065	0.119	-0.066	-0.120	-0.041	-0.779	-0.031	-0.083	0.440	0.598	-0.097	-0.170	-0.015	-0.195	-0.051
Erroi	Overlay	-0.127	0.056	1.016	0.063	0.047	0.005	0.005	0.002	-0.055	-0.123	-0.035	-0.038	-0.036	-0.045	2.009	0.477	-0.011	-0.136	0.052	-0.114	Ø.255
	S	2804.9	-660.1	3948.7	-4082.2	-3514.4	-1965.6	-10873.6	-4432.5	-7338.8	4851.9	1270.4	4155.6	6996.0	5978.2	7784.7	894.8	-3329.3	-10258.5	2976.4	-6426.3	13430.6
Errors	Pycno	-2239.3	7.98.7	4231.2	6 •966	535.8	-721.6	1004.8	2058.0	-1131.1	-1346.7	-240.2	-618.5	-328.2	-1468.5	38.3	1140.5	-626.8	-3297.6	-235.1	-2786.2	-941.1
	Overlay	-685.6	629.8	2790.2	613.6	594.8	30.8	84.3	29.0	-943.2	-1377.2	-205.2	-29.8	-381.7	-789.7	174.8	908.9	-71.2	-2622.4	810.9	-1633.3	4682.9
	CR	8211.9	10579.9	6693.7	5659.8	9249.6	4659.4	4653.4	12900.5	9790.2	16041.9	7151.4	4949.6	17550.0	23724.2	7871.7	2800.8	3113.7	9093.5	18430.4	7.0687	31818.6
stimates	Pycno	3167.7	12038.7	6976.2	10738.9	13299.8	5903.4	16531.8	19391.0	15997.9	9843.3	5640.8	175.5	10225.8	16277.5	125.3	3046.5	5816.2	16054.4	15218.9	11530.8	17446.9
т	Overlay	4721.4	11869.8	5535.2	10355.6	13358.8	6655.8	15611.3	17362.0	16185.8	9812.8	5675.8	764.2	10172.3	16956.3	261.8	2814.9	6371.8	16729.6	16264.9	12683.7	23070.9
	Actual	5407	11240	2745	9742	12764	6625	15527	17333	17129	11190	5881	794	10554	17746	87	1906	6443	19352	15454	14317	18388
	Distr	-	2	ო	4	S	9	7	œ	6	10	H	12	13	14	15	16	17	18	19	20	21

Est. Method	RMSE (RRMSE)	(R ²)	Slope (S.E.)	Intercept (S.E.)
Overlay	1479.756	.97Ø	0.974 +	151.18Ø
Pycno.	1657.200	.96Ø	1.015 + (.055)	97.621
CR (w = $2^{-d/.15}$)	(1102) 6109.640 (.576)	•51Ø •26Ø)	(.005) Ø.494 + (.159)	(708.243) 5267.386 + (2032.109)
+ Significant	at 95% cor	nfidence	level	

Table 2.2. Summary Statistics for Actual and Estimated Values for Planning Districts (Source: Lam, 1980)

planning districts occurred in suburban areas. This is to be expected because a target unit in a suburban area often overlaps one or more predominantly <u>urban</u> source units. The high density of these source units must be partially allocated to the target unit, resulting in frequent overestimation.

D. Applications and Adaptations of the Methods

Most other articles addressing the areal data reaggregation problem either directly or indirectly (Goodchild and Lam, 1980; Lam, 1985; Wallin, 1984) were generally applications rather than evaluations of the methods. Clarke (1984) presented an interesting adaptation of the pycnophylactic method. He generated a pycnophylactic surface exhibiting periodicities, implementing it using two-dimensional fourier series instead of simple smoothing. A fourier series is a summation of a set of sine and cosine waves of various wavelengths, amplitudes, and phase angles. One possible application of a periodic surface is to model geographic central places. By changing the number of harmonics in the series, he was able to simulate surfaces ranging from the stepped surface represented by the source units to a smooth surface similar to one interpolated using inverse distance-weighting. Between these extremes, irregular periodicities were generated in the surface. Clarke was able to incorporate all of these diverse hypotheses about the actual surface while maintaining the pycnophylactic property.

E. Summary

The research into the areal data reaggregation problem, then, has yielded three basic methods: overlay, contour reaggregation, and the pycnophylactic method. Markoff and Shapiro confirmed that better estimates will generally be achieved when the source units are small relative to the target units. In addition, Lam (1980) showed that the volume-preserving methods, overlay and the pycnophylactic method, clearly perform better than those that do not preserve volume. However, no difference in performance between the two volume-preserving methods has been demonstrated. This is surprising considering the difference in their estimating surfaces.

CHAPTER III. RESEARCH HYPOTHESES AND OBJECTIVES

Although Lam (1980) found no significant difference between the overlay and pycnophylactic data reaggregation methods for her study area, there are characteristics of the areal units and the actual distribution that may be useful in discriminating between the performance of the two methods. In particular, the absolute size of the areal units and the criteria used in delineating them could affect the relative performance of the methods. Variations in these two factors can change the validity of the choropleth and smoothness assumptions implied by the use of the overlay and pycnophylactic methods, respectively.

A. Effect of Source Unit Size on Data Reaggregation

It is well established that reaggregation from smaller to larger units will generally produce better estimates using any of the reaggregation methods. However, the <u>absolute</u> size of the source units may also play a role. Units of larger size may have less "opportunity" to capture areas of spatial homogeneity and thereby satisfy the choropleth assumption. This idea was introduced by Coulson (1978) in his paper on the potential for variation in areal units. Potential for variation is an index that Coulson devised for assessing the likelihood, or opportunity, for a single areal unit to contain a homogeneous distribution. It is based on the hypothesis that larger and less compact areal units are less likely to be homogeneous. The

empirical evidence in support of this hypothesis is scanty, but the idea has intuitive appeal. Since the overlay method is based on the choropleth assumption, larger source units may result in poorer performance for this method.

B. Effects of Source Unit Delineation Criteria on Data Reaggregation

The criteria, if any, used to delineate the source units can give clues to the configuration of the actual surface. There are other, preferred methods for obtaining information on actual surface configuration, namely: 1) smaller source units and 2) data for intra-unit features such as the locations and populations of cities. However, this kind of information can be difficult to obtain. Often, verbal descriptions of the delineation criteria are the best information at our disposal about the intra-unit configurations. Our ability to model this configuration for a study area, i.e., to generate an accurate estimating surface, could strongly affect the accuracy of the reaggregated data.

C. Research Objectives

The present study replicates Lam's experimental methodology using units that are significantly different in size and delineation criteria. It uses source units that are much larger than Lam's census tracts and that are delineated in a manner that should favor the pycnophylactic

method. In other words, an attempt is made to ascertain whether source unit size and delination will lead to a significant difference between them. The performances of the two volume-preserving data reaggregation methods are compared using the 48 contiguous states as target units and 181 economic areas, delineated by the U.S. Bureau of Economic Analysis, as source units (fig. 3.1).

The economic areas completely cover the area of the United States. The sets of delineation criteria for Lam's census tracts and the economic areas are listed in Table 3.1. These criteria give some indication of the configuration of the population density surface within the source units. They can be used to hypothesize idealized, intra-unit population distributions for each set of units, and a comparison of these idealized distributions with the estimating surfaces used by the data reaggregation suggests which method will perform better.

Of the two sets of criteria, the one for economic areas reveals more about the actual surface configuration. The idea that each area contains a predominant node or central place (criterion B2) surrounded by tributary counties suggest an idealized distribution that is bivariate normal. A cross-section of this distribution is shown in figure 3.2a. This idealized economic area distribution corresponds closely to the "peak" that is interpolated using the pycnophylactic method when a source unit is surrounded by neighbors with lower densities (fig. 2.2c), i.e., when it is a local maximum. Thus, the



Figure 3.1. Source Units and Target Units

Table 3.1. Source Unit Delineation Criteria

- A. Census Tracts in London, Ontario (used by Lam)
 - Boundaries follow permanent, easily identified features such as roads and streams.
 - 2. Total population between 2500 and 8000 for each tract, except within central business district.
 - 3. Socio-economic homogeneity within each tract.
 - 4. Compact shape.
- B. Economic Areas in the United States (used in present study)
 - 1. Each area disaggregates into counties.
 - 2. An economic node, usually an SMSA, within each area.
 - 3. Some areas contain smaller SMSA's as secondary nodes.
 - Cities with population above 25000 function as nodes in regions where no SMSA's occur.



a) Within an Economic Area b) Within a Census Tract

Figure 3.2 Idealized Surface Configurations pycnophylactic method should produce better estimates for states intersecting economic areas that are local maxima and worse estimates when the economic areas are local minima. The <u>difference</u> in the errors associated with each method will generally be greatest in the locally high-valued economic areas, where the pycnophylactic method performs best. Based on the size and delineation criteria for economic areas, then, it is anticipated that the pycnophylactic method will yield superior estimates.

Although the criteria for census tracts indicate nothing specific about the actual surface configuration,

some information can be hypothesized. Permanent linear features (criterion Al) often separate land uses that are associated with different population densities. That is, they may represent discontinuities in the population surface. Socio-economic homogeneity (criterion A3) within a tract may imply some degree of discontinuity in the population surface. For instance, regions of lower socio-economic status are often associated with higher population densities. Finally, if we accept Coulson's hypothesis, compact shape (criterion A4) reduces the potential for variation of the census tracts. Thus, there is some justification for proposing a model for the census tract distributions similar to the stepped surface shown in cross-section in figure 3.2b. This model resembles the estimating surface associated with the overlay method, the source units themselves. The overlay method did, in fact, have a higher percentage of superior estimates for Lam's planning districts (71%).

To summarize, the objective of this research is to determine whether there are significant differences in the performance of the volume-preserving methods of areal data reaggregation. It is argued that the absolute size and delineation criteria for source units can be important discriminating variables, and the units selected for this study are markedly different in these two factors than units used previously. The use of economic areas, rather than census tracts, as source units may favor of the pycnophylactic method.

CHAPTER IV. RESEARCH METHODOLOGY

The general research strategy has precedent in the earlier studies. A set of small units that aggregate into both the source and target units is used as an approximation of the actual density surface. The data for these small units are aggregated to give "actual" values for the source and target units. The data reaggregation methods are applied to the source units to yield estimated values for the target units. These estimates are then compared with the actual target unit values.

A. Data Sources and Compilation

U.S. counties in 1980 were used as the small units because they aggregate into both economic areas and states. The digitized map data for 3073 counties were obtained from a file distributed by the U.S. Bureau of the Census. The county population data were obtained from the <u>County and</u> City Data Book, 1980 (U.S. Bureau of the Census).

The organization of the data was then transformed from the original polygon, or vector, structure to a raster structure using the ERDAS 400 digitizing system. A grid cell approximately equal in size to Baltimore City, Maryland and Washington, D.C was selected (167.3 sq. mi.). The original vector data had been digitized from a base map on an Alber's Equal Area projection.

The rasterization process resulted in the elimination of 433 counties from the file. This loss was probably an

artifact of the rasterization algorithm and the file sequence of the digitized county data. A more detailed description of these two factors is provided here to account for the loss of data.

B. Rasterization Methods and the Loss of County Data

The vector-to-raster conversion process often operates sequentially, processing one feature (in this case, a polygon) at a time. In other words, it assigns a cluster of grid cells to each county in the file as it comes to it. Depending on the criteria used for assignment, small counties processed earlier in the file can be lost because their grid cells are assigned to subsequent counties in the file. Some common assignment criteria are: 1) assign the grid cell to the county with the largest overlap area, 2) assign the grid cell to the county that overlaps at the center point of the cell, and 3) assign the grid cell to the last county in the file that intersects it, regardless of the area of overlap. The third criterion is the crudest and would probably result in the loss of more counties than the other two, but any of these methods can result in some loss. Information on the algorithm used by the ERDAS system is proprietary.

The counties in the file were generally organized alphabetically within states that were also organized alphabetically. A consequence of this organization was that the loss of counties was biased toward those with names that are low in the alphabet because they were

processed earlier. Although many counties were lost, none of the states or economic areas were unrepresented; and because the surface without the lost counties became the surface of reference for all further analyses, the data loss should not affect the conclusions reached about the reaggregation methods.

C. Implementation of the Areal Data Reaggregation Methods

Once the population and map data were consolidated and aggregated into states and economic areas, the two reaggregation methods were applied to the economic areas. With raster data, the overlay method simply required aggregating all of the grid cell values for economic areas into states. The pycnophylactic interpolation process described in figure 2.1 was applied to the economic areas and allowed to iterate twelve times. The changes in individual cell values and in total populations for economic areas were closely monitored during this process. It was found that the rate of change for both had nearly decreased to zero after twelve iterations. Additional iterations probably would not have transformed the surface significantly. The final pycnophylactic surface was then aggregated into states to produce population estimates.

Some of the details of pycnophylactic interpolation are not shown in figure 2.1 and deserve mention. First, the smoothing routine assigned each grid cell the mean value of its non-diagonal neighbors, regardless of how many there

were (zero to four). If a cell had no non-diagonal neighbors, then its value remained constant; and if a cell had only one neighbor, then it was assigned the neighbor's value. Second, cells at the boundary of the study area were not affected by their neighbors outside of the study area, which were assigned a constant value of zero. This means that there were "cliffs" at the edges of the study area even after interpolation. This boundary condition is more realistic along the coasts of the United States than along its Canadian and Mexican borders.

Some of the primary deficiencies of the research design were the approximation of the actual surface with county-level data, the loss of counties due to rasterization, the somewhat arbitrary selection of twelve iterations for the pycnophylactic interpolation process, and the unrealistic boundary conditions. It is felt that none of these weaknesses was important enough to affect the results substantially.

The results of these data reaggregation experiments can be divided into three parts. First, since the number of iterations used in the pycnophylactic algorithm directly affects the target unit estimates, information on the convergence of the interpolation process is valuable. Second, the "goodness-of-fit" of the estimating surface (economic areas for overlay, the interpolated surface for the pycnophylactic method) to the actual one, may be the most important factor affecting the overall performance of the reaggregation methods and needs examination. And third, the target unit estimates themselves constitute the primary results. The overall performance of the methods is compared with their performance in Lam's study.

A. Convergence of the Pycnophylactic Interpolation Process

The interpolated surface was reaggregated into the target units after twelve iterations. The decision to stop at this point was arbitrary, but it can be supported by examining the change in the estimating surface during the interpolation process. Two statistics indicative of this change were monitored during this process. These statistics were 1) the RMSE between the values for each grid cell after subsequent smoothings, and 2) the RMSE between total target unit values before and after adjustment. The following formula was used:

$$R^{MSE_{j}} = \sqrt{\frac{1}{n} \sum_{i}^{n} (O_{i,j} - O_{i,j-1})^{2}}$$

4)

where, for the second RMSE, ^O_{i,j} the the value for target unit i after adjustment, ^O_{i,j-1} is the value for target unit i before adjustment, n is the number of target units.

These RMSE's indicate the overall effectiveness of the smoothing routine and the volume-adjustment routine, respectively. The progression of their values is displayed in figure 5.1.

Since the first RMSE was computed over 18060 grid cells, as opposed to 48 states, its value changed more smoothly. The rate of change of both curves had declined almost to zero by twelve iterations, which can be interpreted to mean that the smoothing and adjustment routines had lost most of their effectiveness. Additional iterations, then, probably would not have changed the surface significantly.

B. Goodness-of-Fit of the Estimating Surfaces

The "optimal" estimating surface would minimize the difference between itself and the actual surface, and the trend of this difference during the pycnophylactic interpolation process indicates whether or not the process is improving the surface. RMSE can be used again to summarize this error in the interpolated surface.

Figure 5.2 shows the succession of RMSE values between the interpolated and actual (county) values for all grid cells. Iteration Ø is the pre-interpolated (economic area) surface, and iterations 1 through 12 are the increasingly smooth, pycnophylactic surfaces. The error monotonically decreases and is similar to the convergence of the pycnophylactic process represented in figure 5.1. The pycnophylactic surface after 12 iterations, then, was generally a better approximation than any of the previous ones, including the economic area surface used to derive overlay estimates. This is an important result, because it suggests that the target estimates using the pycnophylactic method should be an improvement over those using overlay. But positive and negative errors in the estimating surface can offset one another when they are aggregated into the target units, so the aggregation can hide the improvement.

The goodness-of-fit can also be viewed spatially using maps of the cell-by-cell difference between the estimated surfaces and the actual one. The maps displayed in figure 5.3 correspond to iterations Ø (5.3a and 5.3b) and 12 (5.3c and 5.3d) in figure 5.2. They show the error in the estimating surfaces used by the overlay and pycnophylactic methods. For visual clarity, the values are classed and separated into their positive and negative components. The two estimating surfaces must be very similar judging by their similar error distributions. The postive errors are extensive and the negative errors are intensive and concentrated in urban areas. Neither method, then, is able



Figure 5.1. Convergence of the Pycnophylactic Interpolation Process



Figure 5.2. Error in the Estimating Surface



Figure 5.3a. Error in the Estimating Surface, Positive Error: Economic Areas



Figure 5.3b. Error in the Estimating Surface, Negative Error: Economic Areas



Figure 5.3c. Error in the Estimating Surface, Positive Error: Pycnophylactic Surface



Figure 5.3d. Error in the Estimating Surface, Negative Error: Pycnophylactic Surface

to capture the sudden "peaks" in the distribution at the urban areas. The largest positive errors occur in suburban areas.

C. Target Unit Estimates and Errors

Tables 5.1 and 5.2 show the actual and estimated population data for states and the summary statistics compiled from them. These data are also shown spatially in figures 5.4a-5.4c (actual and estimated population densities) and 5.5a-5.5d (positive and negative errors for each method). In general, the results appear to reinforce Lam's findings. There is little difference in the performance of the volume-preserving methods, and both yielded fairly accurate and precise estimates.

Paired-sample t-tests revealed no significant difference between the overlay and pycnophylactic estimates or between each set of estimates and the actual values at the 95% confidence level. Four states (Arizona, California, Maine, and Rhode Island) had perfect estimates using both methods because their boundaries coincided perfectly with those of the economic areas that overlapped them. The data for these states were not included in the summary statistics.

The correlation/regression and RMSE statistics also indicate good overall performance for both methods. The correlations show the high precision of the estimates. The regression slopes are close to 1, but the intercepts, like Lam's, are not significantly different from zero. The

Table 5.1. Population Data for States

(% Act.)	(Pycno	0.0	5.2	0.0	-1.7	-37.5	451.0	-0.3	4.0	-5.3	-16.1	46.7	-5.6	-12.9	18.1	-6.7	0.0	-26.8	-11.8	6.1	-5.3	14.3	7.5	-0.3	-27.4
Error	Overlay	0.0	9.8	0.0	-2.6	-47.4	606.5	-0.8	7.5	-2.3	-18.7	48.1	-5.2	-11.9	30.3	-8.7	0.0	-28.0	-18.8	5.8	-7-9	16.3	6.5	0.6	-24.9
ror	Pycno	0.02 0.0	103661.7	0.0	-40815.1	-1163878.9	441997.4	-24246.9	141979.9	-45348.0	-1765469.1	2000679.5	-141178.3	-293639.4	483378.6	-193393.9	0.0	-902531.5	-605668.7	488643.7	-192698.7	337248.9	344852.1	-2588.5	-365631.9
E	Overlay	0.00 2.005102-	194983.1	0.0	-63043.4	-1473195.8	594351.4	-70245.9	264188.5	-19847.0	-2053077.9	2060715.6	-130430.2	-271475.9	810010.3	-251722.2	0.0	-942298.8	-964945.5	464651.0	-288021.0	384722.9	301017.4	4634.1	-333363.2
mate	Pycno	2718215.0	2103431.7	23497744.0	2345376.9	1943697.1	540001.4	9024080.1	3683469.9	818326.0	9213565.9	628229 0. 5	2374371.7	1991134.6	3158613.6	2711502.1	1025003.0	2467505.5	4524781.3	8508254.7	3453559.3	2697786.9	4948249.1	784101.5	970780.1
Esti	Overlay	2718215.0	2194753.1	23497744.0	2323148.6	1634380.2	692355.4	8978081.1	3805678.5	843827.0	8925957.1	6342326.6	2385119.8	2013298.1	3485245.3	2653173.8	1025003.0	2427738.2	4165504.5	8484262.0	3358237 .0	2745260.9	4904414.4	791324.1	1003048.8
	Actual	2718215	1999770	23497744	2386192	3107576	98004	9048327	3541490	863674	10979035	4281611	2515550	2284774	2675235	2904896	1025003	3370037	5130450	8019611	3646258	2360538	4603397	786690	1336412
	State	AR A	AK	IJ	8	ธ	昭	Ŀ	G	8	Ц	NI	IA	S	КY	P	¥	₽	AA	IW	¥	SM	£	Ę	Ð

Table 5.1. (cont'd)

(\$ Act.)	Pycno 9.4	103.0	42.4	28.4	-0.8	-2.8	-11.1	-14.8	6.5	4.4	-21.5	0.0	9.4	63.7	-17.8	-2.8	-23.1	26.3	12.4	-1.8	83.0	6.3	90.3
Error (Overlay -10.3	149.3	23.5	28.4	6.9	-2.6	-1.9	-14.7	6.4	4.2	-15.2	0.0	6. 6	65.2	-20.0	-3.2	-17.1	76.7	19.1	-2.3	74.2	5.9	96.6
TOT	Pycno 3067_0	903653.2	2334526.7	364682.9	-78184.9	-126421.5	-71840.9	-1436050.8	177852.8	112087.0	-2222120.8	0.0	279736.8	406599.4	-698242.8	-392496.1	-336335.3	134443.6	401940.6	-76329.0	1307875.2	216105.4	423783.2
	<u>Overlay</u> -82471.4	1310866.1	1297354.2	365629.9	94961.7	-113814.7	-12086.7	-1420068.0	176285.6	106565.1	-1567770.1	0.0	197334.0	416131.7	-786096.8	-448242.2	-249022.9	392271.1	620494.3	-93621.7	1169922.1	202304.4	453620.1
imate	Pycno 803560 0	1781379.2	7844710.7	1649977.9	9964181.1	4327472.5	573669.1	8255760.2	2924116.8	2676981.0	8118636.2	746049.0	3248311.8	1045223.4	3226214.2	13590547.9	1119784.7	645899.6	3646138.6	4055827.0	2883916.2	3662822.4	893340.2
Est	Overlay 718021.6	2188592.1	6807538.2	1650924.9	10137327.7	4340079.3	633423.3	8271743.0	2922549.6	2671459.1	8772986.9	746049.0	3165909.0	1054755.7	3138360.2	13534801.8	1207097.1	903727.1	3864692.3	4038534.3	2745963.1	3649021.4	923177.1
	Actual 800493	877726	5510184	1285295	10042366	4453894	645510	9691811	2746264	2564894	10340757	746049	2968575	638624	3924457	13983044	1456120	511456	3244198	4132156	1576041	3446717	469557
	State	HN	2	Æ	λŇ	Ŋ	Ð	Ю	¥	¥	PA	RI	g	ß	z	ž	5	5	R A	MA	M	SM	λM

Est. Method	RMSE (RRMSE)	(R2)	Slope (S.E.)	Intercept (S.E.)	
Overlay	773094.8	.985	.946 +	214946.655	
_	(.182)	(.971)	(.024)	(141770.957)	
Pycno.	757856.9	.984	.950 +	201315.690	
-	(.189)	(.968)	(.Ø25)	(147942.336)	
+ Significa	nt at 95%	confide	nce level		

Table 5.2. Summary Statistics for Actual and Estimated Values for

.

Figure 5.4a. Population Density, 1980: Actual

States

Figure 5.4b. Target Unit Population Density, Population Density, 1980: Overlay Method

Figure 5.4c. Target Unit Population Density, Population Density,1980: Pycnophylactic Method

Figure 5.5a. Target Unit Error, Positive Error: Overlay Method

Figure 5.5b. Target Unit Error, Negative Error: Overlay Method

Figure 5.5c. Target Unit Error, Positive Error: Pycnophylactic Method

Figure 5.5d. Target Unit Error Negative Error: Pycnophylactic Method

RMSEs are very similar, but the RRMSEs exceed Ford's arbitrary cut-off value of .10 considerably. Since none of the empirical tests of areal data reaggregation methods have demonstrated accuracy as high as that prescribed by Ford, it may be that a higher cut-off value would be more realistic.

<u>The values for states with highly urbanized neighbors</u> (e.g., Delaware, Indiana, Kentucky, New Hampshire, Vermont, and West Virginia) <u>were often severely overestimated</u>. This result is probably the counterpart, at this smaller geographic scale, of the suburban errors that Lam observed. The most glaring errors occurred in Delaware and New Hampshire. These errors can be largely attributed to the relatively larger size of the economic areas that overlap these states.

The small size of Lam's data set (21) limits the validity of parametric statistical tests that could otherwise be used to compare these results with hers. However, the data do suggest some improvement for the pycnophylactic method relative to overlay. In particular, the pycnophylactic method yielded better estimates for only 29% (6) of Lam's planning districts versus 59% (26) states. Although the regression statistics can be used as indicators of the accuracy of the methods, they are only summary statistics. For a particular application, the number of improved estimates that one obtains by using one method or the other is also important.

Given an expectation of equal accuracy for the methods,

a chi-square test for a relationship between performance of the reaggregation method and units used was performed. Although the units used cannot be separated from other differences in the two studies that could cause differences in performance, it is felt that the units are clearly the most important methodological difference.

Table 5.3 shows the contingency table used in this test. The computed chi-square statistic from these data is 5.29. This value allows us to conclude, at the 95% confidence level, that there is in fact a relationship between the performance of the method and the spatial characteristics of the areal units used in the study.

Finally, a binomial difference of proportions test (Hammond and McCullagh, pp. 154-157), similar to a t-test for difference of means, was conducted to determine whether the observed proportion of successes for the pycnophylactic method (.59) is significantly greater than the proportion that would be expected if the methods performed identically (.50). The difference of proportions test statistic has a standard normal probability distribution and had a computed value of 0.882 for this data set. The hypothesis that the proportion of successes is equal to .50 cannot be rejected at the 95% confidence level. There is a fairly high probability that a proportion as large as .59 occurred completely by chance. This result suggests that there is little difference between the two methods for a given set of data.

To summarize, the results reinforce many of the

Table 5.3.	Relationship of	Success of	Reaggregation	Method	to
	Source Units Us	ed			

Method

		Overlay	Pycno.
U n	Census Tract	15 (10.66)	6 (10.34)
t	Econ. Area	18 (22.34)	26 (21.66)

* Expected value in parentheses

findings of previous studies. High errors are more likely to occur in target units estimated from relatively larger source units and in predominantly suburban target units. The overlay and pycnophylactic methods performed nearly identically and yielded fairly accurate and precise estimates. A new finding is that, despite the fact that the pycnophylactic estimating surface was generally a better approximation of the actual one, this improvement was not translated into significantly better target unit estimates. Finally, there is some support for the hypothesis that the size and delineation criteria of the source units can affect the relative accuracy of the two methods. For this individual study, the units selected resulted in superior performance for the pycnophylactic method.

CHAPTER VI. SUMMARY AND CONCLUSIONS

This study has compared the performance of two of the best available methods for reaggregating areal data into different units.

The primary motivation for the work was the finding, in a previous study (Lam, 1980), that there was no difference in the performance of the methods. This was a puzzling result because of the difference in their estimating surfaces, the approximations of the actual surface that are aggregated into the target units. The overlay method aggregates the source units themselves, a stepped estimating surface, and the pycnophylactic method aggregates a smooth estimating surface interpolated from the source units. A characteristic of both of these methods that improves their performance relative to others is that their estimating surfaces both preserve the total value, the volume, within the source units.

The units selected for this study, economic areas and states, have spatial characteristics that 1) are markedly different than those used previously (census tracts and planning districts), and 2) should result in greater accuracy for the pycnophylactic method. In particular, the units were much larger and the delineation criteria for the economic areas suggested an intra-unit distribution that is approximately bivariate normal, rather than the homogeneous distribution hypothesized for census tracts. These factors probably favored the overlay method in Lam's study.

The results, like Lam's, do not strongly favor either method. Although both methods are fairly accurate and precise, there was no significant difference in their estimates. Also, although the pycnophylactic method had a higher proportion of successes than overlay, it was not significantly greater than the proportion that would be expected if the methods performed identically. The estimates were compared only after the pycnophylactic interpolation process had almost completely converged.

Comparing the results with those from Lam's study, a significant difference was found in the relative success of the two methods. There is some support, then, for the hypothesis that the spatial characteristics of the units affect the relative performance of the methods.

The highest errors occurred most frequently when the target units intersected source units that were relatively large or highly urbanized. In order to maintain the volume-preserving constraint for the source units, the negative errors (usually located at the urban centers) in the estimating surface had to be offset by positive errors in the neighboring (suburban) areas. These positive errors were then allocated to the suburban target units.

In general, the better the fit of the estimating surface to the actual one, the more accurate the target estimates should be. However, this study has demonstrated that the improved fit may not be translated into better estimates because of the offsetting of positive and negative errors when they are aggregated into the target

units. The pycnophylactic surface was a better representation of the actual one, but there were no significant differences in the pycnophylactic and overlay estimates for target units.

The maps of the errors in the estimating surfaces show that the two methods suffer from a common deficiency: they do not capture extremes in the actual surface well. The estimating surfaces for both methods hover about the mean density values of the source units, so they under- or over-estimate locations that deviate from the means. While the pycnophylactic method is a theoretical improvement over the overlay method, it misrepresents the actual surface in a manner somewhat analogous to the interpolation of a topographic surface using distance-weighted averaging of control points. That is, the interpolated surface is continuous and smooth, or "undulating", while the actual surface is continuous but sometimes "jagged".

The presence of sudden extremes may be a characteristic of population (and related) density surfaces, and such extremes may even be largely limited to a range of geographic scales. The overlay and pycnophylactic methods should perform better when the actual density surface has few extremes.

CHAPTER VII. RECOMMENDATIONS FOR FURTHER STUDY

The aggregation of estimating surfaces into target units is the step in the areal data reaggregation process that distinguishes it from the problem of geographic surface modelling. The objective of surface models is usually to represent some actual surface as accurately as possible. This study has shown that the aggregation step acts as a sort of filter on the errors in the estimating surface. A better estimating surface does not necessarily produce better target estimates. This problem notwithstanding, accurate surface modelling is still a worthy objective in the study of areal data reaggregation: an accurate estimating surface will probably do more to provide good estimates than any other factor. With this in mind, three avenues for future research on areal data reaggregation are suggested.

First, different variables should be used in future empirical tests. The pycnophylactic method will probably model an undulating surface better than the seemingly jagged population density surface used here.

Second, Clarke's fourier adaptation of the pycnophylactic method should be applied to economic areas or similar units. The Clarke algorithm is designed specifically to capture periodicity, such as that exhibited by geographic central places. It is this method, more than any of the other available ones, that has the potential to capture local extremes in the surface.

Third, and finally, recent research (Clarke, 1984; Wallin, 1984) is demonstrating that the pycnophylactic concept, interpolation constrained by volume-preservation, is very flexible. Many different assumptions, or types of detail, can be applied to the surface while preserving the source unit values. One of the major reasons that the pycnophylactic method did not convincingly out-perform overlay in this study is that it interpolates a "peak" within source units only in special cases, i.e., neighbors of lower density. Almost all of the economic areas have one predominant "peak", an SMSA, near their centers. Thus, a bivariate normal distribution was proposed here as a model of the population density surface within an economic area. To constrain a pycnophylactic surface to bivariate normality within source units is a worthwhile research task that could significantly improve the pycnophylactic estimates for this data set.

BIBLIOGRAPHY

Boneva, L., D. Kendall and I. Stefanov (1971), "Spline Transformations: Three New Diagnostic Aids for the Statistical Data-Analyst", Journal of the Royal Statistical Society, Ser. B, vol. 33, no. 1, pp. 1-70.

Clarke, K. C. (1984), "Two-Dimensional Fourier Interpolation for Uniform Area Data", <u>Technical Papers</u>, <u>50th Annual Meeting of the American Society of</u> Photogrammetry, vol. 2, pp. 835-845.

Coulson, M. R. C. (1978), "Potential for Variation: A Concept for Measuring the Significance of Variations in Size and Shape of Areal Units", <u>Geografiska Annaler</u>, Ser. B, vol. 60, no. 1, pp. 48-64.

Crackel, T. (1975), "The Linkage of Data Describing Overlapping Geographical Units -- A Second Iteration", Historical Methods Newsletter, vol. 8, no. 3, pp. 146-150.

Ford, L. (1976), "Contour Reaggregation: Another Way to Integrate Data", <u>Papers</u>, <u>13th</u> <u>Annual</u> <u>Conference</u> <u>of</u> <u>the</u> <u>Urban</u> <u>and</u> <u>Regional</u> <u>Systems</u> <u>Association</u>, vol. 11, pp. 528-575.

Goodchild, M. F. and N. S-N. Lam (1980), "Areal Interpolation: A Variant of the Traditional Spatial Problem", Geo-Processing, vol. 1, pp. 297-312.

Hammond, R. and P. McCullagh (1974), <u>Quantitative</u> <u>Techniques in Geography: an Introduction</u>, Clarendon Press: Oxford, 318 pp.

Lam, N. S-N. (1980), "Methods and Problems of Areal Interpolation", <u>Ph.D.</u> <u>Dissertation</u>, University of Western Ontario, 177pp.

Lam, N. S-N. (1983), "Spatial Interpolation Methods: A Review", The American Cartographer, vol. 10, no. 2, pp.129-149.

Lam, N. S-N. (1985), "A Method for Choropleth Inversion", <u>Technical Papers</u>, <u>45th Annual Meeting of the</u> <u>American Congress on Surveying and Mapping</u>, pp. 365-373. Markoff, J. and G. Shapiro (1973), "The Linkage of Data Describing Overlapping Geographical Units", <u>Historical</u> Methods Newsletter, vol. 7, no. 1, pp. 34-36.

Tobler, W. (1979), "Smooth, Pycnophylactic Interpolation for Geographical Regions", Journal of the American Statistical Association, vol. 74, no. 367, pp. 519-536.

Tobler, W. and J. Lau (1978), "Isopleth Mapping Using Histosplines", <u>Geographical</u> <u>Analysis</u>, vol. 10, no. 3, pp. 273-279.

United States Bureau of the Census (1983), "County and City Data Book, 1983", U.S. Government Printing Office.

United States Bureau of Economic Analysis (1977), "BEA Economic Areas (revised 1977): Component SMSA's, Counties, and Independent Cities", U.S. Government Printing Office.

Wallin, E. (1984), "Isarithmic Maps and Geographical Disaggregation", Proceedings, International Symposium on Spatial Data Handling, vol. 1, pp. 209-217.

