1-1-1992

Accuracy of a GIS-based small-area population projection method used in spatial decision support systems

Panagiotis Lolonis
University of Iowa

Marc P Armstrong
University of Iowa

Claire E Pavlik
University of Iowa

Please see article for additional authors.

ACCURACY OF A GIS-BASED SMALL-AREA POPULATION PROJECTION METHOD USED IN SPATIAL DECISION SUPPORT SYSTEMS

Panagiotis Lolonis
Marc P. Armstrong
Claire E. Pavlik
Sidan Lin

Department of Geography
The University of Iowa
Iowa City, IA 52242

ABSTRACT

This paper describes a recently developed GIS-based population projection method for small spatial units, known as the Modifiable Spatial Filter (MSF) method, and evaluates the accuracy of its projections with respect to a parameter that controls the sample size of the data that are used to estimate projection parameters. This assessment is made with respect to average absolute error, average relative error, distribution of absolute error, and absence of bias for subgroups. Results show that MSF projections appear to be stable when the parameter is greater than fifteen.

INTRODUCTION

Future populations are projected using a wide variety of techniques that differ in their underlying assumptions, the real-world processes modeled, and the input data required (Isserman 1977; Lee and Goldsmith, 1982). Regression based projection techniques (Ericksen, 1974; O'Hare, 1976), for example, associate past population changes with environmental, socioeconomic and political changes that have occurred in a study area. Cohort component techniques (Isserman, 1977; Isserman, 1984), on the other hand, project populations by progressing the population cohorts of a given base year into the future. These traditional population projection methods are most often used to project populations for spatial units with relatively large populations, such as metropolitan areas or states. Many locational planning problems, however, require population projections for much smaller spatial units, such as city blocks or block groups, that contain comparatively small populations. For these problems, traditional methods are inadequate; either the input data required for such methods are unavailable at this scale or the resulting projections are unacceptably inaccurate (National Research Council, 1980; Tayman, 1992).

Recent advances in GIS software and geo-coded databases now enable the development of methods that project populations for arbitrarily small areas. This paper presents one such method, the Modifiable Spatial Filter (MSF) method (Rushton *et al.*, 1992), and assesses the accuracy of its projections against observed values. Previous assessments indicate that when applied at the census block level, the MSF method predicts populations more accurately than traditional cohort component methods (Rushton *et al.*, 1992). In these evaluations of predictive accuracy, however, the minimum population threshold parameter, which controls the size of population sample that is used in the estimation of the projection parameters, was fixed and consequently its effect on projection accuracy is unknown. In this paper, we extend the assessment of the MSF method by evaluating the effects of variation of the minimum population threshold on the accuracy of one-year predictions.

THE MSF METHOD

The MSF method, an adaptation of the cohort component method, uses disaggregated geo-demographic data to estimate projection parameters and project populations for small areas. The MSF method estimates projection parameters for each spatial unit using data not only from that unit but also from neighboring units. The assumption underlying this method is that neighboring spatial units are characterized by similar demographic processes and, therefore, have similar projection parameters. Thus, the variance of cohort component parameters for spatial units with small population cohorts can be reduced by estimating progression rates using pooled data from neighboring areas.

The MSF method consists of five steps. First, the residence locations of individuals living in a study area are determined by linking population data to a digital geographical database using the address matching capabilities of GIS software. Second, the address matched population data are aggregated to spatial units appropriate for the study (e.g. city blocks). This step is performed using the spatial aggregation capabilities of GIS software. Third, the "neighborhood" of each spatial unit, which consists of all units (including itself) that are within a user-specified maximum distance from the unit, is determined using geometrical data stored in the GIS database and special purpose software (e.g. Lolonis, 1992). Fourth, progression rates for each spatial unit and age group are estimated using the population and neighborhood data from the two previous steps. This estimation is made by sequentially accumulating data from the center of the unit's neighborhood to its periphery, until a minimum population level is reached. If the required minimum population level is not reached within the unit's neighborhood, then progression rates for the study area as a whole are assigned to the unit. Currently, this estimation requires special purpose software that implements the MSF method (Lolonis, 1992). Finally, projections are computed using the populations of the "base year", the estimated progression rates, and projections of each year's initial cohort (e.g. projected births).

The estimation of the progression rates in the MSF method is made using the equation:

$$r_i(g,g+1) = \frac{\sum_{j \in F_i(g,g+1)} \sum_{t=T_0+1}^{T_1} n_{jt}(g+1)}{\sum_{j \in F_i(g,g+1)} \sum_{t=T_0}^{T_1-1} n_{jt}(g)} \qquad (1)$$

where:

$r_i(g,g+1)$: the progression rate of spatial unit i and age groups g and $g+1$;
i : basic spatial unit index;
g : age group index;
T_0: initial year of population data;
T_1: final year of population data;
t: index of year in the sequence T_0 through T_1;
j: index for the spatial units within the MSF of the ith spatial unit;
$F_i(g,g+1)$: set of spatial units in the filter area of spatial unit i, for age group g; and
$n_{jt}(g)$: number of individuals at spatial unit j, year t, and age group g.

The *filter area* of unit i for the gth age group is defined as the k closest spatial units in its neighborhood (including itself) that contain an average of at least M students for age groups g and g+1 over the observed years. If this minimum population threshold is not reached within the neighborhood defined for unit i, then the filter area is assumed to be the entire study area. Mathematically, the filter area is defined as:

$$F_i(g,g+1) = \{s_i(1),\ldots,s_i(k), k \le p_i\} \tag{2}$$

$$\text{if} \quad \frac{\sum_{j=1}^{k_i-1}\sum_{t=T_0}^{T_1}\left[n_{s_i(j)t}(g)+n_{s_i(j)t}(g+1)\right]}{2*(T_1-T_0)} < M \tag{3}$$

$$\text{and} \quad \frac{\sum_{j=1}^{k_i}\sum_{t=T_0}^{T_1}\left[n_{s_i(j)t}(g)+n_{s_i(j)t}(g+1)\right]}{2*(T_1-T_0)} \ge M \tag{4}$$

$$\text{otherwise} \quad F_i(g,g+1) = \{\text{ all units in the area}\} \tag{5}$$

in which the additional variables are defined as:

M : minimum population threshold in the filter area;
p_i: number of spatial units in the neighborhood of *i;*
k: number of spatial units in the filter area of *i;*
j: index indicating the order of a spatial unit in the filter area;
s_i(j): *j*th spatial unit in the neighborhood of unit *i*. The elements of that set are ordered in ascending order with respect to distance from spatial unit *i*;
$n_{s_i(j)t}(g)$: population for spatial unit $s_i(j)$ at year *t*, and age group *g*.

Population projections for spatial unit *i*, age group *g*+1, and time period T+1 are computed as:

$$n_{iT+1}(g+1) = r_i(g,g+1) * n_{iT}(g) \tag{6}$$

where:

n_{iT+1}(g+1): population of unit *i* and age group *g*+1 at time T+1;
r_i(g,g+1): progression rate for spatial unit *i* and age groups (*g*,*g*+1); and
n_{iT}(g): population of unit *i* and age group *g* at time T.

In summary, equation (1) evaluates the proportion of the cohort in age group *g*, residing in the filter area of spatial unit *i*, that advanced to the next age group (*g*+1) in the next time period (T+1) over the observed time interval. Equations (3), (4) and (5) define the filter area for each spatial unit. Finally, equation (6) computes the projected population using the progression rates, the past population estimates, and projections of the youngest cohort.

The MSF method addresses the apparent inability of cohort component methods to model the small populations at the block level through pooling the demographic data of proximal areas. The pooling done for a projection is dependent on the value of M, the minimum population threshold, which determines the size of the filter for each projection unit given the spatial distribution of observed cohort component populations and the maximum distance constraint.

Two factors affect the accuracy of the estimates generated using the MSF method. If M is small, then the size of the spatial filter tends to be small and data are pooled over an area that is likely to be characterized by similar demographic processes. Projections computed using small M values, however, tend to be subject to the same problems observed in traditional cohort component models applied to small areas: when population counts within the filter area are low, even single count changes in the numerator or denominator of equation (1) strongly affect the resulting grade progression rate estimate. This results in cohort population projections that may deviate substantially from actual values because of the sensitivity of parameter estimates to small differences in observed data. On the other hand, if M is large, the problem of parameter sensitivity to small changes is reduced by increasing the number observations included in the estimate. The result, however, is that the size of the spatial filters also tends to increase, so that filter areas probably will contain spatial units with dissimilar demographic characteristics. With large M, therefore, the filtered progression rate estimates for a particular unit may include observations from units that are substantially different demographically, thus increasing the likelihood of inaccurate projections.

Previous results (Rushton *et al.*, 1992) demonstrate that with M set to 30 and a distance constraint of one mile, the MSF method provides more accurate one year projections than the traditional cohort component method. While these results indicate the general utility of the method in small area population prediction, they must be extended to include the influence of the minimum population threshold parameter, M, on predictive accuracy. Through such an extension, the tradeoff between parameter sensitivity and spatial homogeneity can be evaluated, leading to recommended M values for prediction. Here, we examine the impact on predictive accuracy of systematic variation in M, comparing observed and estimated data values for block level cohort components. The evaluation of the accuracy of the predictions is based on criteria proposed by the Panel on Small-Area Estimates of Population and Income (National Research Council, 1980).

THE STUDY AREA

The assessment of predictive accuracy of the MSF method presented here is based on the elementary school population of the Iowa City, Iowa, Community School District. In this assessment, census blocks are the spatial units used, while population cohorts are defined as the students of the 7 elementary school grades (K through 6). The district contains 1,187 census blocks and served 5,587 elementary students in 15 elementary schools during the 1991-92 school year. For the analysis, 896 of the 1,187 census blocks, which extend over areas with a well established mailing address system, were used. The remaining 291 census blocks, which are primarily rural and without an established street address system, were excluded from the analysis because the residential locations of students living in those areas could not be determined through address matching.

Official school enrollment records for the five year period 1987-91 containing each student's name, mailing address, and grade were used as the population data source. Enrollment data for the first four school years were used to estimate MSF projection parameters; enrollment data for the final school year, 1991-92, were used to assess the accuracy of projections based on the estimated parameters. In the comparison of projected and actual enrollments, projections for the youngest cohort, kindergarten students, were obtained with a regression-based, trend extrapolation technique (see Lolonis, 1992). Students were assigned to census blocks by address matching using a commercially available GIS software package (Caliper, 1990). In the address matching stage, approximately 90% of student addresses were successfully matched; unmatched student records were not included in the analysis.

ASSESSMENT OF THE ACCURACY OF MSF PROJECTIONS

The accuracy of the cohort projections for minimum threshold parameter values ranging from 0, which implements the traditional cohort component model, to 30 were evaluated by comparing actual and predicted student values by block and grade for elementary school students in the Iowa City Community School District during the 1991-92 school year. Examining these differences district-wide for total enrollments (Table 1), we see that while the traditional cohort component model over-predicts total enrollments by 510, this is approximately halved with the minimum population threshold parameter, M, set to 2, and decreases to 69 as M increases to 30. Thus, at the district level of aggregation the improvement in predictive accuracy of cohort component models due to spatial filtering is clear. In order to assess the accuracy of the method in more detail, we evaluated the differences between actual and predicted values with respect to average absolute error.

average relative error, shape of the error distribution, and absence of bias for population subgroups.

Table 1. District-wide Actual and Projected Elementary School Enrollments by Grade, 1991

M	0	1	2	3	4	5	6	Total
0	791	837	800	801	817	809	720	5575
	50	*71*	*62*	*79*	*94*	*103*	*51*	*510*
2	762	804	781	774	775	748	688	5332
	21	*38*	*43*	*52*	*52*	*42*	*19*	*267*
5	750	798	762	759	762	742	680	5253
	9	*32*	*24*	*37*	*39*	*36*	*11*	*188*
7	749	789	755	757	758	731	679	5218
	8	*23*	*17*	*35*	*35*	*25*	*10*	*153*
10	748	786	749	750	750	730	675	5188
	7	*20*	*11*	*28*	*27*	*24*	*6*	*123*
15	748	781	748	746	741	724	673	5161
	7	*15*	*10*	*24*	*18*	*18*	*4*	*96*
20	749	781	747	742	743	721	669	5152
	8	*15*	*9*	*20*	*20*	*15*	*0*	*87*
25	748	780	745	739	739	720	667	5138
	7	*14*	*7*	*17*	*16*	*14*	*-2*	*73*
30	747	777	742	741	738	721	668	5134
	6	*11*	*4*	*19*	*15*	*15*	*-1*	*69*
Actual	741	766	738	722	723	706	669	5065

Note:
Numbers in *italics* show the differences between projected and actual enrollments.

Average Absolute Error

Average absolute error indicates how different, on average, a projected value is from its corresponding observation. Here, average absolute value was computed by averaging the absolute value of the difference between observed and projected grade population values for each block over the 896 blocks and 7 elementary grades in the study (6,272 observations). Average absolute error was computed for each of the minimum population threshold parameter values using:

$$e_M = \frac{\sum_{i=1}^{N} \sum_{g=g_i}^{g_f} \left| n_i(g) - \hat{n}_{Mi}(g) \right|}{N * (g_f - g_i + 1)} \tag{7}$$

where:

e_M: average error for minimum population threshold value M;
i : basic spatial unit index;
N : total number of spatial units in the study area;
g : age group index;
g_i: first age group;

g_f: last age group;

$n_i(g)$: actual population of unit i and age group g in 1991-92 academic year;

$\hat{n}_{Mi}(g)$: projected population of unit i and age group g in 1991-92.

Table 2. Mean Absolute Error as a Function of the Minimum Population Threshold (M)

Minimum Population Threshold	Number of Observations	Average Absolute Error	Sample Standard Deviation	Standard Deviation of the Mean
0	6272	.445	.986	.012
2	6272	.403	.824	.010
5	6272	.378	.762	.010
7	6272	.371	.751	.010
10	6272	.368	.746	.009
15	6272	.362	.736	.009
20	6272	.359	.734	.009
25	6272	.359	.732	.009
30	6272	.358	.727	.009

For the Iowa City study area, as M increases to approximately 15 (Table 2) the average absolute error decreases, indicating convergence between observed and expected values. For M values from 15 to 30, however, the average absolute error remained approximately constant.

<u>Average Relative Error</u>

While average absolute error indicates how far predicted values are from observed values in the units of measurement used, the average of prediction errors relative to the observed values themselves is an additional criterion of predictive accuracy (Isserman, 1977). If, for example, we find in evaluating a set of predictions that the average absolute error is small but relative errors are large, we may not be satisfied with the accuracy of the projections. Average relative errors for each of the minimum population threshold parameter values was computed as:

$$\mathrm{Re}_M = \frac{\sum_{i=1}^{N}\sum_{g=g_i}^{g_f}\left(\left|n_i(g) - \hat{n}_{Mi}(g)\right| / n_i(g)\right)}{N * (g_f - g_i + 1)} \qquad (8)$$

where:

Re_M: average relative error for threshold value M.

The remaining variables are defined as in equation (7).

The average relative error values indicate a problem symptomatic of small area projections: because of the low counts involved and the fact that only whole numbers are allowed as actual values, relative errors tend to be high (Table 3). On the whole, we observe that the average relative error decreases as M increases to 15 and stabilizes as M increases from 15 to 30.

Distribution of Absolute Errors
The distribution of absolute errors is used to examine aspects of accuracy, such as the dispersion of errors and the presence of an unacceptably large number of outliers, that are not captured by average absolute error or average relative error. Ideally, the distribution of absolute errors should be highly skewed to the left with few outliers. The presence of a large number of outliers would indicate a weakness in the predictive accuracy of the projection method and would decrease its reliability in applications (Isserman, 1977). Also, lack of clustering of error values close to zero may indicate bias in the projection method.

The distribution of the absolute errors of the MSF projections is heavily skewed to the left with few outliers (Table 4). Indeed, for all values of M, approximately 97-98% of the projected values differ by less than three students per census block and grade from the observed values. For each M, most observations fall within the interval [0,1], and their frequency decreases gradually towards zero. The most prominent changes in the distribution occur when M increases from zero (traditional cohort component) to non-zero values (filter). For M greater than two, the changes in the distribution are negligible.

Table 3. Average Relative Error as a Function of the Minimum Population Threshold (M)

Minimum Enrollment Threshold	Number of Observations	Average Relative Error	Sample Standard Deviation	Standard Deviation of Mean
0	2221	0.550	0.620	0.013
2	2221	0.492	0.521	0.011
5	2221	0.451	0.488	0.010
7	2221	0.441	0.479	0.010
10	2221	0.435	0.481	0.010
15	2221	0.428	0.486	0.010
20	2221	0.425	0.488	0.010
25	2221	0.422	0.488	0.010
30	2221	0.420	0.487	0.010

Note:
To avoid divisions by zero in computing average relative errors, we used only observations with non-zero actual enrollments in creating this table.

Table 4. Frequency and Cumulative Percentage of Absolute Errors as a Function of the Minimum Population Threshold

M	[0]	(0-1)	[1-2)	[2-3)	[3-4)	[4-5)	[5-6)	[6-7)	>=7
0	4050	868	935	248	83	41	16	13	18
	65	*78*	*93*	*97*	*99*	*99*	*100*	*100*	*100*
2	3816	1265	884	173	74	30	12	8	10
	61	*81*	*95*	*98*	*99*	*100*	*100*	*100*	*100*
5	3757	1387	852	174	57	23	11	4	7
	60	*82*	*96*	*98*	*99*	*100*	*100*	*100*	*100*
7	3719	1419	872	160	57	24	11	3	7
	59	*82*	*96*	*98*	*99*	*100*	*100*	*100*	*100*
10	3712	1443	852	162	58	24	12	3	6
	59	*82*	*96*	*98*	*99*	*100*	*100*	*100*	*100*
15	3670	1500	839	168	54	23	9	2	7
	59	*82*	*96*	*99*	*99*	*100*	*100*	*100*	*100*
20	3645	1526	847	156	62	18	9	3	6
	58	*82*	*96*	*98*	*99*	*100*	*100*	*100*	*100*
25	3626	1552	832	169	57	18	11	1	6
	58	*83*	*96*	*98*	*100*	*100*	*100*	*100*	*100*
30	3625	1553	831	171	58	16	10	2	6
	58	*83*	*96*	*99*	*99*	*100*	*100*	*100*	*100*

Notes:
Numbers in *italics* show the cumulative percentages.

<u>Absence of Bias in Population Subgroups</u>

While the first three criteria summarize the accuracy of the MSF projections over the entire set of observations, they do not assess systematic biases in the projections of population subgroups. Such biases may occur because of differences in the demographic processes of particular subgroups. Families, for example, may be concerned about the negative effects of migration on the social life and educational progress of their children and may decide not to migrate when children are in certain grades. Thus, if the accuracy of the projection method depends on migration, it may vary from grade to grade. Similar biases in the accuracy of the projections will be observed if migration patterns vary spatially through time. For example, if new housing is built in certain sub-regions of a study area, thus increasing the in-migration rates for those sub-regions, then projections based on progression rates estimated using population data that describe the demographic processes before the growth occurred will systematically underestimate the populations for those areas.

Analyses for subgroup biases were performed by grouping deviations of observed and projected grade by block values with respect to student grades and elementary school attendance areas. When a grouping is made with respect to grades, a decreasing trend in the values of average absolute error is observed for grades two through six (Table 5). The error for kindergarten projections is larger than that for the rest of the grades. This difference, however, could be caused not only by the underlying demographic processes but also by the method that was used to obtain the kindergarten projection; that method (Lolonis, 1992) is based on trend extrapolation and is not part of the MSF method.

Table 5. Average Absolute Error and Sample Standard Deviation by Grade as a Function of the Minimum Population Threshold (M).

	Grade						
M	0	1	2	3	4	5	6
0	.65	.41	.44	.43	.42	.41	.35
	.95	*.99*	*.97*	*.93*	*.93*	*1.25*	*.79*
2	.62	.37	.42	.39	.37	.34	.31
	.92	*.88*	*.93*	*.79*	*.76*	*.74*	*.67*
5	.61	.35	.39	.37	.34	.31	.28
	.89	*.76*	*.88*	*.73*	*.68*	*.68*	*.63*
7	.61	.34	.37	.36	.34	.30	.27
	.89	*.75*	*.86*	*.73*	*.67*	*.64*	*.63*
10	.61	.34	.36	.36	.33	.30	.27
	.90	*.75*	*.83*	*.72*	*.66*	*.65*	*.62*
15	.60	.34	.37	.35	.33	.29	.26
	.89	*.73*	*.83*	*.70*	*.65*	*.63*	*.61*
20	.61	.33	.36	.34	.33	.29	.26
	.89	*.72*	*.82*	*.70*	*.66*	*.63*	*.62*
25	.61	.33	.36	.34	.32	.29	.26
	.89	*.73*	*.82*	*.69*	*.64*	*.64*	*.62*
30	.60	.33	.36	.34	.32	.29	.26
	.88	*.74*	*.81*	*.70*	*.63*	*.63*	*.62*

Notes:
Numbers in *italics* are sample standard deviations.
The number of observations per grade is 896.

The error trends observed above are also evident here. Indeed, with the exception of kindergarten enrollments, the average absolute error decreases as M increases. Again, the most prominent changes occur when M ranges between zero and fifteen.

When subgroups are defined with respect to existing school attendance areas, the projection accuracy exhibits two distinct characteristics. First, the average errors appear to stabilize at a smaller value of M (Table 6) and second, their magnitude and rate of change vary substantially over space. The most accurate projections are obtained for schools serving primarily urban areas with low growth rates, such as Lincoln (1) and Longfellow, while the least accurate projections are obtained for schools, such as Lemme, Lucas, and Wood, that serve areas which have experienced growth in the recent past. The most evident improvements in the accuracy of projections as a function of M are observed in areas around the periphery of the metropolitan area that have experienced growth; the average absolute error for Lincoln (2), Roosevelt (2) and Shimek decreases by 29%, 33% and 38% respectively, as M increases from zero to thirty. For those three schools, the most rapid decrease in the error values occurs when M increases from zero to seven.

Table 6. Average Errors and Sample Standard Deviations by School Attendance Area as a Function of the Minimum Population Threshold

School	0	2	5	7	10	15	20	25	30
C. Central	0.57	0.50	0.47	0.46	0.46	0.45	0.45	0.44	0.44
(64; 441)	*0.96*	*0.84*	*0.76*	*0.75*	*0.75*	*0.77*	*0.76*	*0.73*	*0.72*
Hills	0.65	0.62	0.63	0.63	0.63	0.61	0.60	0.60	0.60
(17; 179)	*1.31*	*1.30*	*1.31*	*1.33*	*1.28*	*1.19*	*1.19*	*1.18*	*1.19*
Hoover	0.46	0.37	0.35	0.34	0.34	0.34	0.33	0.33	0.33
(72; 356)	*0.85*	*0.58*	*0.53*	*0.52*	*0.51*	*0.51*	*0.51*	*0.51*	*0.51*
Horn	0.55	0.50	0.46	0.46	0.46	0.44	0.43	0.43	0.43
(41; 333)	*0.83*	*0.72*	*0.68*	*0.68*	*0.69*	*0.67*	*0.67*	*0.67*	*0.66*
Kirkwood	0.86	0.83	0.77	0.76	0.76	0.76	0.76	0.76	0.76
(34; 417)	*1.28*	*1.23*	*1.11*	*1.11*	*1.13*	*1.13*	*1.11*	*1.10*	*1.10*
Lemme	0.73	0.72	0.69	0.66	0.67	0.66	0.66	0.66	0.66
(31; 325)	*1.08*	*1.08*	*1.08*	*1.05*	*1.05*	*1.04*	*1.04*	*1.04*	*1.04*
Lincoln (1)	0.19	0.18	0.17	0.17	0.18	0.17	0.17	0.17	0.17
(45; 116)	*0.50*	*0.44*	*0.42*	*0.41*	*0.42*	*0.42*	*0.42*	*0.42*	*0.42*
Lincoln (2)	1.11	0.96	0.88	0.83	0.81	0.79	0.79	0.79	0.79
(12; 101)	*1.64*	*1.31*	*1.09*	*0.99*	*0.96*	*0.98*	*0.98*	*0.98*	*0.98*
Longfellow	0.20	0.18	0.16	0.16	0.16	0.16	0.16	0.16	0.15
(165; 303)	*0.54*	*0.42*	*0.41*	*0.40*	*0.41*	*0.40*	*0.40*	*0.40*	*0.40*
Lucas	0.71	0.67	0.62	0.61	0.61	0.60	0.60	0.60	0.59
(43; 465)	*1.44*	*1.41*	*1.20*	*1.20*	*1.22*	*1.20*	*1.20*	*1.20*	*1.18*
Mann	0.26	0.23	0.23	0.22	0.22	0.22	0.22	0.22	0.22
(101; 274)	*0.73*	*0.57*	*0.55*	*0.55*	*0.56*	*0.56*	*0.57*	*0.57*	*0.57*
Penn	0.51	0.47	0.47	0.46	0.46	0.45	0.44	0.45	0.45
(39; 340)	*0.84*	*0.76*	*0.75*	*0.75*	*0.74*	*0.72*	*0.71*	*0.73*	*0.73*
Roosevelt (1)	0.34	0.30	0.28	0.29	0.29	0.28	0.27	0.27	0.27
(63; 214)	*0.67*	*0.55*	*0.56*	*0.55*	*0.55*	*0.53*	*0.53*	*0.53*	*0.53*
Roosevelt (2)	0.70	0.60	0.51	0.49	0.49	0.47	0.47	0.46	0.47
(34; 253)	*2.09*	*1.12*	*0.87*	*0.80*	*0.81*	*0.77*	*0.76*	*0.76*	*0.76*
Shimek	0.50	0.42	0.35	0.35	0.32	0.32	0.31	0.31	0.31
(33; 177)	*1.09*	*0.78*	*0.65*	*0.65*	*0.60*	*0.60*	*0.60*	*0.60*	*0.59*
Twain	0.34	0.32	0.31	0.30	0.29	0.28	0.28	0.29	0.29
(76; 364)	*0.77*	*0.74*	*0.74*	*0.72*	*0.66*	*0.62*	*0.62*	*0.64*	*0.64*
Wood	0.98	0.93	0.88	0.85	0.85	0.84	0.85	0.85	0.83
(26; 407)	*1.27*	*1.22*	*1.22*	*1.20*	*1.21*	*1.21*	*1.21*	*1.18*	*1.17*

Notes:
Numbers in *italics* are sample standard deviations.
The numbers within parentheses below each school name show the census blocks and the 1991-92 total enrollment for that school.

DISCUSSION AND CONCLUSIONS

These analyses suggest that the accuracy of the MSF projections improves as the number of students used to make estimates of progression parameters increases from zero to fifteen. This improvement is generally monotonic and is observed in the trends of the total projected enrollments (Table 1), average absolute error (Table 2), and average relative error (Table 3). The trends are also evident when observations are grouped with respect to grade (Table 5) and space (Table 6). The rate of improvement, however, decreases as M increases and is negligible for values of M greater than fifteen. In the analyses performed here, there is no indication that there are intervals in the values of M for which accuracy fluctuates systematically.

ACKNOWLEDGMENTS

We would like to thank David Smyth for his assistance in developing the database on which the analyses in this paper are based and Trudy Meyers for assistance in data entry.

REFERENCES

Caliper Corporation, 1990, *GisPlus: Geographic Information System*, Newton, MA.

Ericksen, E.P. 1974, A regression method for estimating population changes of local areas: *Journal of the American Statistical Association*, Vol. 69, pp. 867-875.

Isserman, A. 1977, The accuracy of population projections for subcounty areas: *Journal of the American Institute of Planners*, Vol. 43, pp. 247-259.

Isserman, A. 1984, Projection, forecast, and plan: On the future of population forecasting: *APA Journal*, Vol. 50, pp. 208-221.

Lee, E.S. and Goldsmith, H.F. 1982, ed. *Population Estimates: Methods for Small Area Analyses*, Sage Publications, Beverly Hills.

Lolonis, P. 1992, Procedures and Programs for Implementing the Modifiable Spatial Filter (MFA) Method for Making Small-Area Population Projections, (forthcoming, Discussion Paper, Department of Geography, University of Iowa, Iowa City, IA).

National Research Council. 1980, *Estimating Population and Income of Small Areas*, (Panel on Small-Area Estimates of Population and Income), National Academy Press, Washington D.C.

O'Hare, W. 1976, Report on a multiple regression method for making population estimates: *Demography*, Vol. 13, pp. 369-379.

Rushton, G., Armstrong, M.P., Lolonis, P. 1992, Small Area Student Projections Based on a Modifiable Spatial Filter: *Socio-Economic Planning Sciences* (forthcoming).

Tayman, J. 1992, An Evaluation of Subcounty Population Forecasts: *URISA Proceedings*, Vol. IV, pp. 172-179.