Modeling Sensitivity to Accuracy in Classified Imagery: A Study of Areal Interpolation by Dasymetric Mapping*

Peter F. Fisher and Mitchel Langford

University of Leicester

Areal interpolation is the process by which data collected from one set of zonal units can be estimated for another zonal division of the same space that shares few or no boundaries with the first. In previous research, we outlined the use of dasymetric mapping for areal interpolation and showed it to be the most accurate method tested. There we used control information derived from classified satellite imagery to parameterize the dasymetric method, but because such data are rife with errors, here we extend the work to examine the sensitivity of the population estimates to error in the classified imagery. Results show the population estimates by dasymetric mapping to be largely insensitive to the errors of classification in the Landsat image when compared with the other methods tested. The dasymetric method deteriorates to the accuracy of the next worst estimate only when 40% error occurs in the classified image, a level of error that may easily be bettered within most remote sensing projects. **Key Words: areal interpolation, dasymetric mapping, sensitivity analysis, error, accuracy.**

Introduction

A major focus of GIS research has been a concern with the accuracy of spatial data (Goodchild and Gopal 1989). Much of that research has concentrated on the nature and description of error in spatial data (Goodchild 1991; Congalton 1992). Less attention has been paid to its consequences in subsequent analyses. In this paper we examine the propagation of error in a classified Landsat image into the areal interpolation of population counts.

In previous research, we tested the accuracy of a number of methods of areal interpolation (Fisher and Langford 1995). We found that a method based on dasymetric mapping (Wright 1936; Monmonier and Schnell 1984) is a much better method than others tested. Dasymetric mapping involves the use of areas with known or estimated population density to enhance the detail of a choropleth map of population below the detail of the finest enumeration area. As we implement it, dasymetric mapping is used to estimate population distributions on the basis of the location of residential and nonresidential land uses derived from classification of satellite imagery. The rate of misclassification of land covers derived from satellite imagery is, however, notoriously high (Campbell 1987; Congalton 1992), and inaccurate land cover information could clearly have profound effects on the accuracy of the areal interpolation. The aim of the research reported here is to show the effect of the error in the classified image (henceforward referred to as classification error) on the accuracy of the population estimates (estimation error), and thus to explore the sensitivity of the dasymetric method to that classification error.

This paper first reviews the methods of areal interpolation examined here, and then presents the error modeling procedures used. Discussion of the nature of errors in classified Landsat imagery follows, before the methods for simulating errors in a classified image are presented. Results of the analysis and discussion follow.

Methods and Background

Areal Interpolation

Population statistics are invariably collected and published according to some spatial aggregation. In urban areas of the United States the smallest such unit is the census block, while in rural areas it is the enumeration district. These are hierarchically arranged into spatial units of

*We would like to thank Doris Brencke, Philip Jutson, Alan Strachan, David Troughear, David Unwin, and David Walker for their various inputs to the research reported here. The care and assistance of the editorial team at *The Professional Geographer* is much appreciated. This paper is based on one presented at GIS/LIS '93.

Professional Geographer, 48(3) 1996, pages 299–309 © Copyright 1996 by Association of American Geographers.
Initial submission, April 1995; revised submissions, November 1995, December 1995, January 1996; final acceptance, January 1996.
Published by Blackwell Publishers, 238 Main Street, Cambridge, MA 02142, and 108 Cowley Road, Oxford, OX4 1JF, UK.

progressively coarser spatial resolution, block group, census tract, etc. In other countries other terms are used. In England and Wales, the location of research reported here, enumeration districts (EDs) are the finest spatial reporting areas for the census; they are hierarchically aggregated into wards, and then districts. All census data are published only for these spatial aggregations. Analysts, however, commonly want to integrate the census data with data collected according to some other spatial division of the study area, but the boundaries of the census areas are only rarely coincident with the boundaries of the other units. The transformation of attribute data from one zonation to another is known as areal interpolation or cross-area estimation (Goodchild and Lam 1980; Langford et al. 1991; Goodchild et al. 1993), with the areal units of the initial zonation referred to as *source* zones, and those of the subsequent zonation as *target* zones (Flowerdew and Green 1989). Many different methods of areal interpolation have been suggested (Goodchild and Lam 1980; Goodchild et al. 1993). These methods have been divided into cartographic, regression, and surface methods by Fisher and Langford (1995). They tested five methods and showed that cartographic methods provided both the best (dasymetric methods) and worst (areal weighted) estimates.

The areal weighted method is the simplest and most easily implemented approach to areal interpolation. It simply allocates population to target zones according to the relative area of source zones (Equation 1).

$$P_t = \sum_{r=1}^{q} \frac{A_{ts_r} P_{s_r}}{A_{s_r}} \tag{1}$$

where P_t is the estimated population of a target zone,

q is the number of source zones which overlap with that target zone,

P_s is the population of the rth overlapping source zone,

A_s is the area of that rth source zone, and

A_{ts} is the area of geometric overlap between that rth source zone and the target zone.

Obviously, this method is based only on geometric properties of the source and target zones; the underlying assumption is that people are evenly distributed within all source polygons. This assumption is acceptable only if nothing else is known about the population distribution, but even then it is generally inaccurate. People live in collections of houses that occasionally may be evenly distributed across the landscape (and possibly within a source zone), but more commonly they live in clusters of houses interspersed with nonresidential land.

Various methods have been suggested for overcoming the fundamental flaw in the areal weighted method. Of four methods taking into account land use information, Fisher and Langford (1995) show that the dasymetric method provides the best results, probably because it is the only method, other than the areal weighted method, in which a local subset of the source zones is used in interpolation to each target zone. This method was originally developed by Wright (1936) from a concern that choropleth maps do not give even a remotely valid representation of the distribution of population within enumeration areas (the same concern as that expressed above for the areal weighted method). Working in the area of Cape Cod, Wright showed that if control zones of known (or estimated) population density can be located, then the choropleth map may be refined and population density mapped at resolutions finer than the enumeration areas, giving more (and more valid) detail over the standard choropleth map. Wright's method is adapted here to employ land cover mapping from satellite imagery (Campbell 1987) and census enumeration data.

Within a GIS the dasymetric method is most easily implemented if it is simplified compared with that proposed by Wright (1936), where areas of several different population densities were identified. Here the necessary land use information is simply the binary divide between residential and nonresidential use, which is relatively easy to distinguish from classified satellite imagery. The method can be developed in almost any raster GIS environment; the one chosen here is Idrisi (Eastman 1992). A flow diagram that achieves dasymetric mapping is shown in Figure 1, and was implemented as an Idrisi batch file. Three

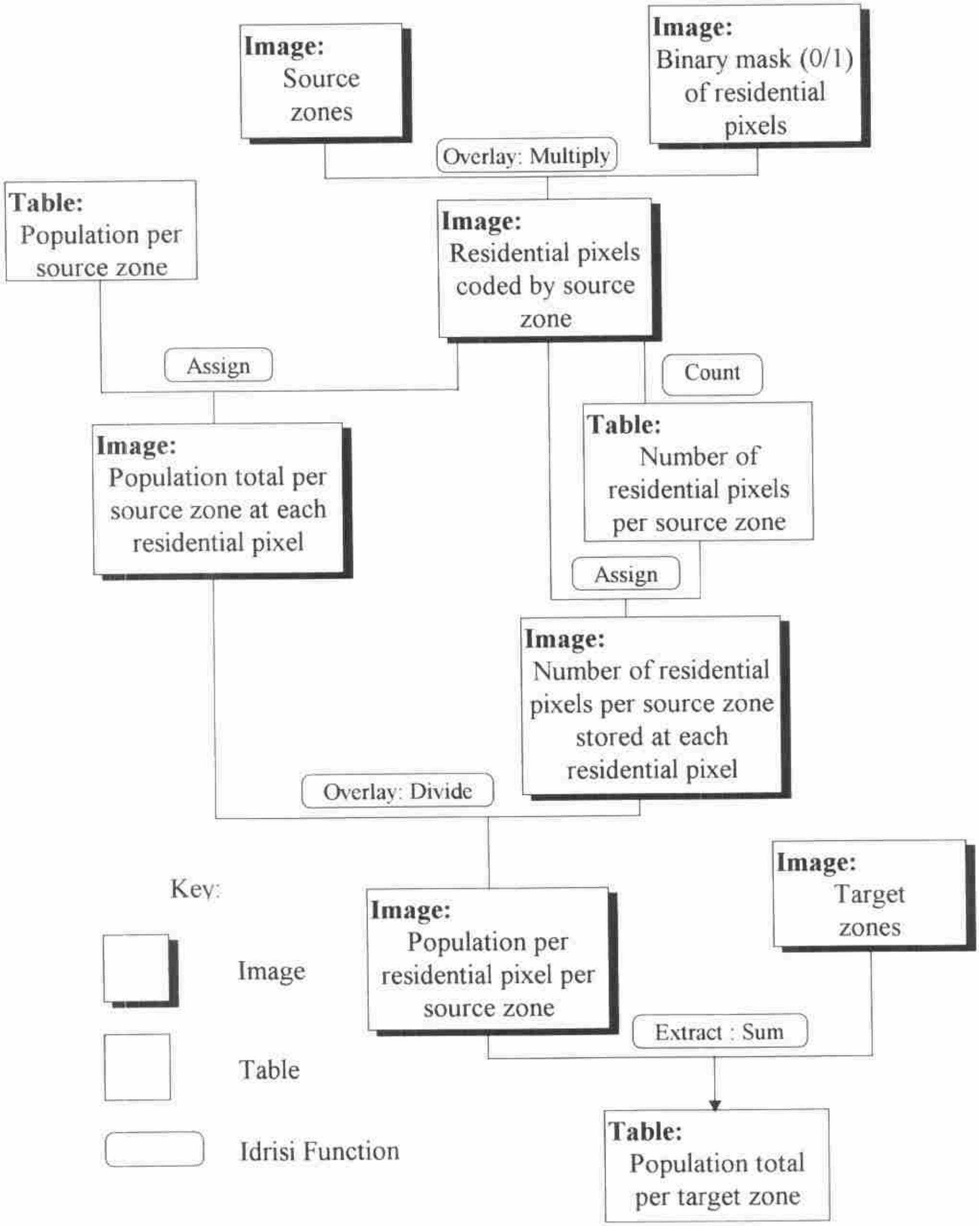

Figure 1: *The process of dasymetric mapping as implemented for the research reported here.*

main inputs of spatial data are used: the extent of source and target zones and the binary mask of residential and nonresidential pixels. A table of source zone populations is also required. The first step in Figure 1 is to find the number of residential pixels within each source zone, R_s, and then map that number back as an attribute of the residential pixels within that source zone, r. In parallel the population per source zone, P_s, is mapped to each residential pixel within a zone. These two images (residential pixels per source zone and population per source zone both with values only in residential pixels) are then divided to give an estimate of the population in each residential pixel where the value of all residential pixels within a source zone will be the same. With the image of target zones it is then easy to count the population of any target zone as the sum of populations of all residential pixels within it, P_t. The number of residential pixels within the target zone, R_t, do not have to be found, but effectively this is an implementation of Equation 2.

$$P_t = \sum_{r=1}^{q} \frac{R_{t_r} P_{s_r}}{R_{s_r}} \tag{2}$$

We have extended the method to employ three classes of land use, with considerable increase in computational complexity but only minor improvements in accuracy; results are not reported here.

Simulating Errors in Classified Imagery

The assignment of pixels in the image to land cover types is remarkably unreliable (Campbell 1987). Determination of pixels of residential housing, for example, may be confused by the fact that houses are rarely the size of the pixel, but are mixed with all manner of other land covers. The remote sensing literature is replete with accuracy measures that are derived from a comparison and cross tabulation of the cover type classified at a location and the cover type actually present, the so-called confusion matrix (Campbell 1987). The most common measure of accuracy in a classified scene is the percent correctly classified, or the overall accuracy. It is essentially the proportion of sites where the classified and the actual land covers are coincident, expressed as a percentage of all the sites examined. Overall accuracy in a classified remotely sensed image rarely exceeds 90%, but is commonly better than 80%. Indeed, the widely quoted standard for land cover mapping from remote sensing suggests that an acceptable accuracy is 85% (Anderson et al. 1976). In the work reported here, percent accuracy is taken to be (100 - percent error).

Fisher (1991, 1994) presented computer algorithms that use a number of different measures of classification accuracy, including the overall accuracy, to assign any pixel in a classified satellite image to alternative random cover types. This method can be used to generate multiple versions of possible land cover images, each version being a realization of the integration of the classified image with the accuracy information. In the current project, various levels of the overall accuracy in the classified image are modeled, because overall accuracy gives the most general and, over all cover types, the worst case for the effect of errors in the classified image.

Two variants of this basic simulation approach are used here. In the first, a pixel is simply reassigned to any cover type where all cover types have an equal chance of being chosen. This method has no constraints. A variant method preserves areas; the frequencies of occurrence of the land cover types in the initial image are determined, and then used as prior probabilities in the simulation algorithm, so that all resulting realizations of the image will have the same proportions of the different cover types as the original. Fisher and Langford (1995) showed that all of these regression methods yielded worse population estimation errors than the dasymetric method.

Study Area

The three western census Districts of the county of Leicestershire in England are the test area for the research reported here (Fig. 2). Data from 985 enumeration districts of the 1981 census of population were used for analysis, and the 49 wards into which they are arranged were taken as the source zones for analysis. Multiple sets of target zones were found by the same method as that used by Fisher and Langford (1995), which is based on Openshaw's (1977) algorithm. Essentially, the enumeration districts form elemental zones, and are randomly aggregated into a smaller number of larger areal units (Fig. 3). In this way the actual populations of the target zones are known and can be compared with the populations interpolated from the source zones.

A 1984 TM image of the area was registered to the census zones. Seven primary land cover types were identified using the maximum likelihood classifier, including high and low density residential, industrial, water, woodland, agricultural land, and quarries. These were re-

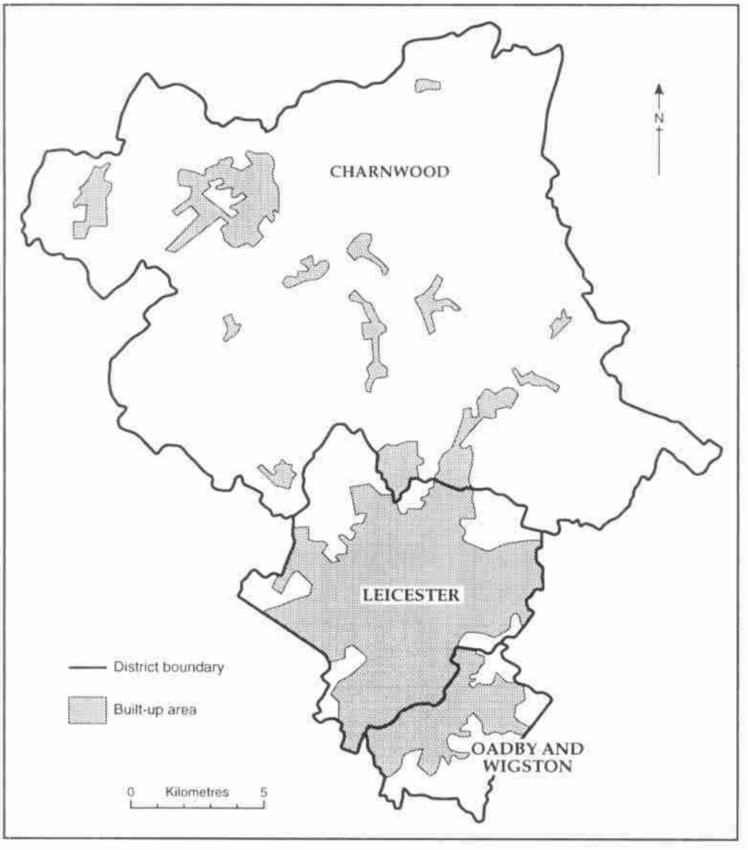

Figure 2: *The study area of western Leicestershire.*

duced to just residential and nonresidential cover types for dasymetric mapping. Field checking of the area showed that 20% of pixels were classified erroneously (80% overall accuracy).

Experimental Design

Two sets of tests were run. In the first set, four different approaches to interpolation were performed: (1) areal weighted interpolation and (2) dasymetric interpolation without simulating classification error in the land cover, and dasymetric interpolation with (3) unconstrained simulation of classification error and (4) simulated classification error with area preservation. The level of classification error randomized was 20%, being that determined by ground truth checking (Artzen 1992), and the interpolation was performed onto 10, 50, and 100 target units. For each of the four experimental situations and for each of the three numbers of target units, 250 different versions (each a realization of the random generation process) of the target units and (where appropriate) the land cover map with simulated classification error by both the methods outlined above were found.

In the second set of tests, unconstrained classification error was simulated in the land cover data between 0 and 100% at 10% increments, and population was interpolated to alternative versions of 100 target zones. Again, each error level was the subject of 250 realizations of the simulation process.

Reports of the estimation error for both experiments are given as the root mean

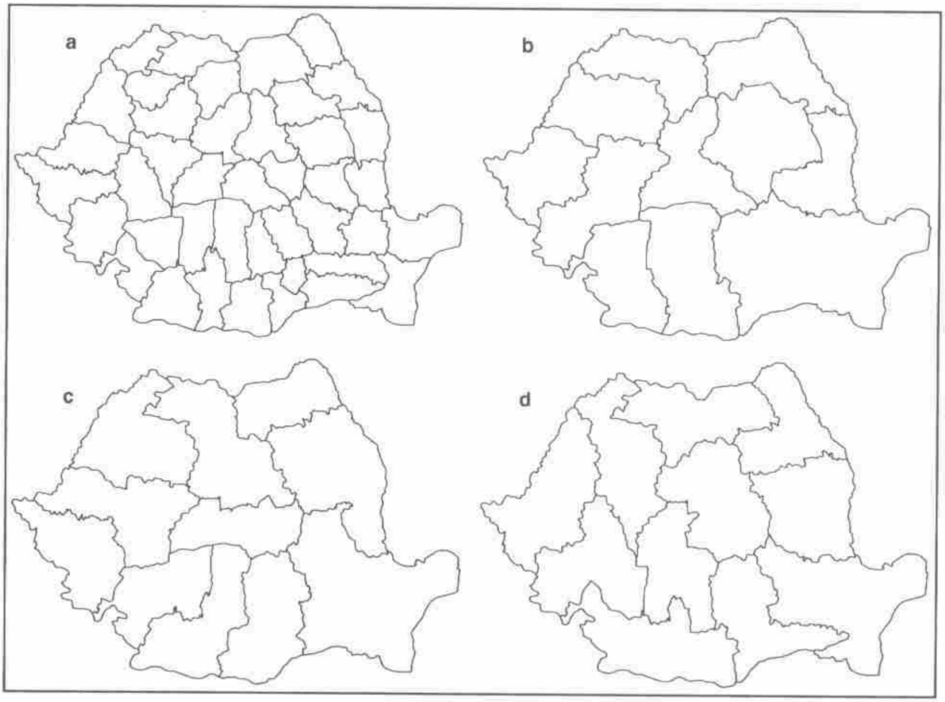

Figure 3: *Enumeration districts form elemental zones (A), and can be aggregated into many possible higher level spatial units (B, C, D).*

squared error of the population estimates (RMSE), as defined in Equation 3.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{j=1}^{m} (P_{aj} - P_{tj})^2}{m}} \qquad (3)$$

where P_{tj} is the estimated population of the jth target zone,

P_{aj} is the actual population of the jth target zone, and

m is the number of target zones in the experiment.

Results

Table 1 presents 12 different sets of results. The mean and coefficient of variation of the root mean squared error for each different realization of the randomization process are reported for the areal weighted method and the dasymetric method with the original land cover map treated as true, and for the dasymetric method with 20% classification error (80% accuracy) simulated in the overall accuracy without constraint and with the area preserving approach. The first column shows the results of the areal weighted method, and in each of the other three columns the results of different dasymetric models are presented. The second column shows results when no error was simulated in the classified Landsat image. In these tests, the original classified Landsat image is taken as correct, and therefore the first two columns can be taken as end points for the error model, because the assumptions of the areal weighted method means that it should return the same results as a dasymetric model with 100% classification error, or white noise, in the classified image. Indeed, when 100% classification error was modeled (see below), the null hypothesis that it was the same as the areal weighted method was accepted at

Table 1 *Root Mean Squared Errors and Coefficients of Variation in the Estimation of Populations 10, 50 and 100 Target Zones*

Interpolator	Areal weighted	Dasymetric model: Original classified land cover	Dasymetric model: 80% overall accuracy simulated	Dasymetric model: 80% accuracy simulated with area preserving
m = 100	mean population in targets = 4597			
Mean RMS error	2669	1225	1311	1291
(Coeff of var)	(0.58)	(0.27)	(0.28)	(0.28)
Minimum RMS error	2200	958	1067	1022
	(0.48)	(0.21)	(0.23)	(0.22)
Maximum RMS error	3353	1515	1546	1648
	(0.72)	(0.33)	(0.33)	(0.35)
m = 50	mean population in targets = 9194			
Mean RMS error	3575	1629	1744	1716
(Coeff of var)	(0.39)	(0.18)	(0.19)	(0.19)
Minimum RMS error	2621	1196	1184	1138
	(0.28)	(0.13)	(0.13)	(0.12)
Maximum RMS error	4529	2190	2372	2263
	(0.49)	(0.24)	(0.26)	(0.25)
m = 10	mean population in targets = 45977			
Mean RMS error	5758	2639	2840	2772
(Coeff of var)	(0.13)	(0.06)	(0.06)	(0.06)
Minimum RMS error	2176	1084	967	990
	(0.05)	(0.02)	(0.02)	(0.02)
Maximum RMS error	11851	5270	5402	5395
	(0.26)	(0.11)	(0.12)	(0.12)

the 95% significance level; the two methods (areal weighted and dasymetric with 100% classification error) yield effectively the same result.

In the final two columns of Table 1, the results of modeling 20% classification error (reflecting comparison with ground truth) by both the method for overall accuracy and the area preserving variant are given for three different numbers of target zones. The results show that modeling error in the Landsat classification yields values of the interpolated populations for which the estimation error lies between the areal weighted and the dasymetric methods. At the 95% significance level, the t test shows that the results reported for each pair of values in each row of the table are significantly different (the null hypothesis is rejected). Thus, although the two results of error modeling shown in any row are very similar in both the mean and the spread of RMSE values, they cannot be considered to be samples drawn from the same population. Estimates with the area preserving method are always very slightly more accurate than when only the overall accuracy is modeled. Although the absolute values of the RMS error increase with decreasing numbers of target zones, the pattern with respect to the degree of estimation error in the different experiments is exactly the same.

Due to the relative speed of processing, and the similarity of results from the two methods for simulating classification error, only the overall accuracy was simulated over the whole range of levels of classification errors yielding a full sensitivity analysis. The mean, maximum, and minimum RMS error values for each are plotted against the amount of classification error in Figure 4, and the percentage increase in estimation error from both the preceding value and the original is shown in Table 2. The results show a gentle increase in the maximum estimation error up to about 60% classification error (only a 36% change in estimation error over the original value at 60% classification error), and then a steep climb to the final 100% classification error (125% change). The change in mean and minimum values, on the other hand, shows a climb on a more steadily increasing curve. Histograms of the distribution of selected levels of error show that the range of RMSE values increases with increasing error; Figures 5c and 5d are almost identical, having a broad spread within the same limits and bimodal distributions, while the

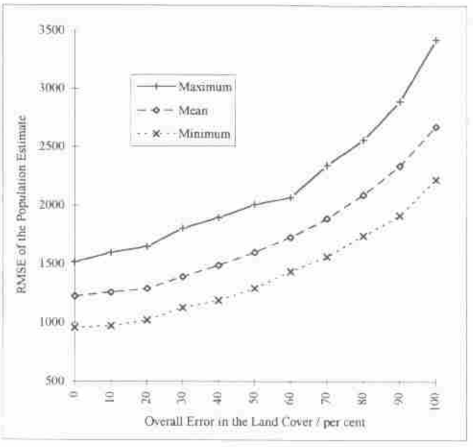

Figure 4: *The variation in minimum, mean, and maximum RMS errors in the population estimates with changing simulated overall error in the classified Landsat scene.*

other two are much closer to normal distributions and have single modal values with progressively smaller means and ranges as the error is reduced. In short, the increase in classification error is accompanied by an increase in estimation error, in skewness of the distribution of the estimation error, and in deviation from a near-normal distribution. With 100% classification error simulated, the distribution of RMSE in the estimates is equivalent to the area weighted method.

Statistical significance testing shows that no pairwise tests of distributions of neighboring estimation error values yield any pairs where the null hypothesis must be accepted; all are different. Specifically, the error of the estimates in the 0% model is significantly different from that where 10% classification error is modeled, the 10% is different from that with 20%, and so on. In other words, there is no distinct break point when the classification error can be shown to have a statistically different effect on the estimates as it increases.

Comparison with Other Interpolation Methods

In this section the error distribution of the dasymetric method with variable levels of land cover accuracy is compared with other methods of cross-area estimation.

Other methods of spatial interpolation include those based on surface interpolation and on regression. Three regression models were suggested by Langford et al. (1991); they use the areas of land cover and land use types at different levels of generalization and derived from satellite imagery as the independent variables. The *simple* model uses just the number of pixels in residential use as the independent variable, the *focused* model uses two classes of residential land as independent variables, and the *shotgun* method employs seven different land cover types. All methods apply global coefficients of linear regression equations estimated from the properties of source zones to predict the populations of target zones. The error distributions of the population estimates from these methods are reported by Fisher and Langford (1995).

The error statistics (RMS error and coefficient of variation) of estimations for these methods are reported in Table 3. These esti-

Table 2 *Change in Average Root Mean Squared Error of the Population Estimates with Increasing Simulated Error in the Classified Landsat Image*

Rate of error simulated in the Landsat image	Percent increase in RMS error over RMS with no classification error			Percent increase in RMS error over previous level of classification error		
	Maximum	Mean	Minimum	Maximum	Mean	Minimum
10	5.5	2.8	1.7	5.5	2.8	1.7
20	8.8	5.4	6.7	3.1	2.5	4.9
30	19.1	13.5	17.8	9.5	7.7	10.5
40	25.0	21.5	24.1	4.9	7.0	5.3
50	32.4	30.7	35.1	5.9	7.5	8.8
60	36.4	41.1	50.1	3.0	7.9	11.1
70	54.6	54.2	63.3	13.3	9.3	8.8
80	69.2	70.8	81.7	9.4	10.7	11.3
90	91.0	91.1	100.0	12.9	11.9	10.0
100	125.9	118.6	132.0	18.3	14.4	16.0

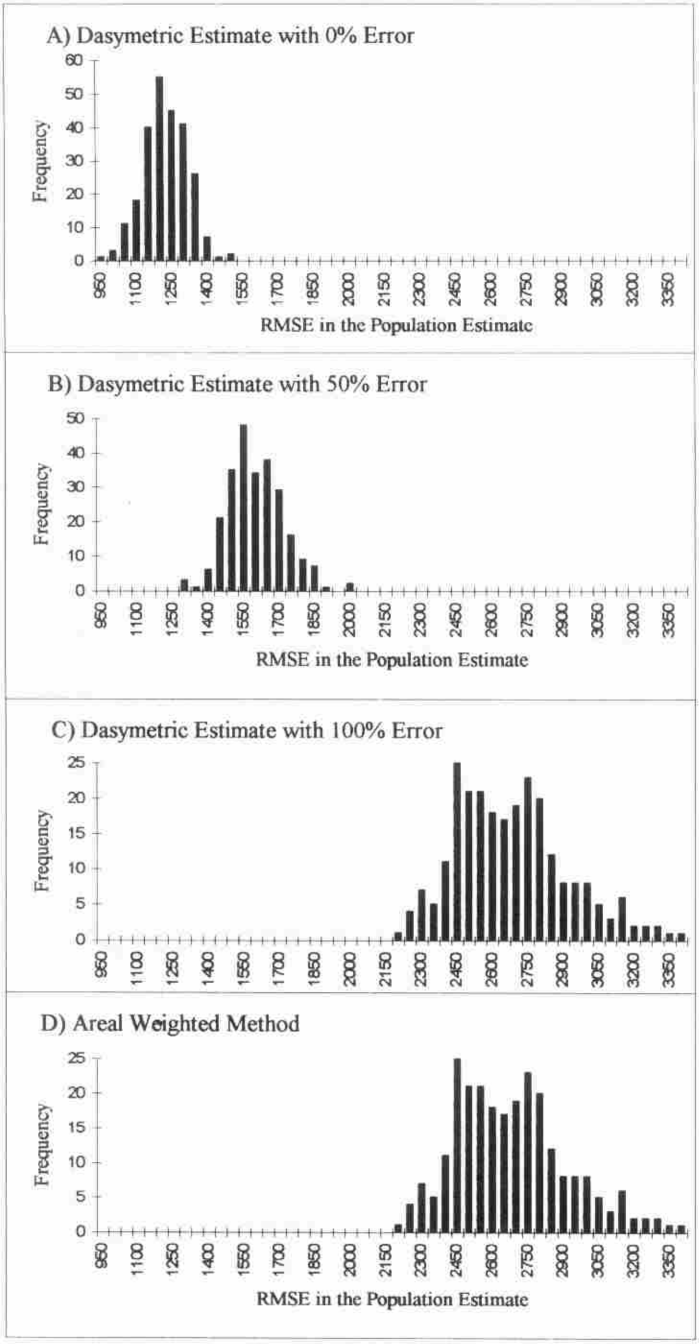

Figure 5: *The distribution of the individual RMS errors in the population estimates for each of 250 simulations for the (a) 0, (b) 50 and (c) 100% errors and (d) for the areal weighted method.*

Table 3 *Root Mean Squared Errors and Coefficients of Variation of Population Estimates for Each of Three Regression Models for Areal Interpolation to 100 Target Zones*

	Simple	Focused	Shotgun
Mean RMS error	1895	1725	1553
(Coefficient of variation)	(0.41)	(0.37)	(0.31)
Minimum RMS error	1688	1507	1313
	(0.36)	(0.32)	(0.28)
Maximum RMS error	2249	2002	1847
	(0.49)	(0.43)	(0.40)

mates are based on 250 interpolations to 100 target zones and are comparable with the results shown in Figure 5. Table 3 can be compared with results presented in Fisher and Langford (1995, Table 3d); small variations between the two tables are due to their being based on different sets of realizations. The three methods produce slightly different levels of estimation error from each other (Table 3; for a full discussion see Fisher and Langford 1995). When these are compared with the range of estimation errors in the dasymetric method and different levels of error in the land cover are simulated, it can be seen that the minimum RMSE in any of the regression models (1,313 in the shotgun method) coincides with the minimum curve of the dasymetric method when the error in the Landsat classification is between 40 and 50% (Fig. 5). The maximum RMSE of the regression models (2,248 for the simple method) coincides with the maximum curve between 60 and 70% error in the land cover. This comparison shows that of the methods tested, the dasymetric method is not only the most accurate, but it is relatively stable; errors of up to 40% in the classified Landsat image still yield better estimates of the interpolated populations than any other method tested.

Conclusion

The sensitivity analysis reported here shows that:

- just 10% classification error in the Landsat scene gives a significantly different result from the population estimation errors when the original Landsat scene is used; but
- up to 40% error can be incorporated in the Landsat scene, before estimations using the dasymetric method deteriorate to the level of accuracy of any other method tested.

In other words, the dasymetric method is generally robust against error in classified satellite imagery, since classification errors as high as 40% are rarely reported in the literature, especially when using a TM scene that is classified into only seven original cover types.

The reason for the relative robustness of the dasymetric method under classification error must be due to aggregating pixels into source and target zones. That is to say the frequencies of pixels in different land classes may not vary significantly within a zone even when up to 40% classification error is being simulated. This is supported by research by Strahler (1981), Gershmel and Napton (1982), and Franklin et al. (1986), who all show that classification errors at the pixel level in a raster dataset can be large (classification accuracy) without impacting the accuracy of estimates of regional amounts (inventory accuracy).

In summary, this research shows that the dasymetric method is not only relatively accurate for areal interpolation, but is also robust to error in the land cover classification. Furthermore, in applications based on the frequency of cover types where pixels are aggregated into larger zones, even quite large errors in classified imagery may be relatively unimportant. ■

Literature Cited

Anderson, J. R., E. E. Hardy, J. T. Roach, and R. E. Witmer. 1976. *A Land Use and Land Cover Classification System for Use with Remote Sensor Data.* USGS Professional Paper No. 964. Washington, DC: U.S. Geological Survey.

Artzen, Pym. 1992. Personal communication.

Campbell, James B. 1987. *Introduction to Remote Sensing.* New York: Guilford.

Congalton, Russell G. 1992. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment* 37:35–46.

Eastman, Ronald. 1992. *Idrisi User's Guide, Version 4.* Worcester, MA: Idrisi Project, Clark University.

Fisher, Peter F. 1991. Modelling soil map-unit inclusions by Monte Carlo simulation. *International Journal of Geographical Information Systems* 5:193–208.

Fisher, Peter F. 1994. Visualizing the accuracy in classified remotely sensed images. *Photogrammetric Engineering and Remote Sensing* 60:905–10.

Fisher, Peter F., and Mitchel Langford. 1995. Modelling the errors in areal interpolation between zonal systems by Monte Carlo Simulation. *Environment and Planning A* 27:211–24.

Flowerdew, Robin, and Michael Green. 1989. Statistical methods for inference between incompatible zonal systems. In *The Accuracy of Spatial Databases,* ed. Michael F. Goodchild and Sucharita Gopal, 239–47. London: Taylor & Francis.

Franklin, J., T. L. Logan, C. E. Woodcock, and A. H. Strahler. 1986. Coniferous forest classification and inventory using Landsat and digital terrain data. *IEEE Transactions on Geoscience and Remote Sensing* GE-24:139–49.

Gershmel, Philip J., and D. E. Napton. 1982. Interpretation of resource data: Problems of scale and spatial transferability. In *Proceedings of the Annual Meeting of the Urban and Regional Information Systems Association,* 471–82. Washington, DC: URISA.

Goodchild, Michael F. 1991. Issues of quality and uncertainty. In *Advances in Cartography*, ed. Jean-Claude Muller, 113–39. London: Elsevier.

Goodchild, Michael F., and Sucharita Gopal. 1989. *Accuracy of Spatial Databases.* London: Taylor & Francis.

Goodchild, Michael F., and Nina S.-N. Lam. 1980. Areal interpolation: A variant of the traditional spatial problem. *Geoprocessing* 1:297–331.

Goodchild, Michael F., Luc Anselin, and U. Deichmann. 1993. A framework for the areal interpolation of socioeconomic data. *Environment and Planning A* 25:383–97.

Langford, Mitchel, David J. Maguire, and David J. Unwin. 1991. The areal interpolation problem: Estimating population using remote sensing in a GIS framework. In *Handling Geographical Information: Methodology and Potential Applications,* ed. Ian Masser and Michael Blakemore, 55–77. London: Longman.

Monmonier, Mark, and G. Schnell. 1984. Land use and land cover data and the mapping of population density. *International Yearbook of Cartography* 115–21.

Openshaw, Stan. 1977. Algorithm 3: A procedure to generate pseudo-random aggregations of N zones into M zones, where M is less than N. *Environment and Planning A* 9:169–84.

Strahler, Alan H. 1981. Stratification of natural vegetation for forest and rangeland inventory using Landsat digital imagery and collateral data. *International Journal of Remote Sensing* 2:15–41.

Wright, J. K. 1936. A method of mapping densities of population with Cape Cod as an example. *Geographical Review* 26:103–10.

PETER FISHER is Senior Lecturer in the Department of Geography at the University of Leicester, Leicester LE1 7RH, United Kingdom. His primary research interest is the study of uncertainty in spatial data.

MITCHEL LANGFORD is Lecturer in Remote Sensing in the Department of Geography at the University of Leicester, Leicester LE1 7RH, United Kingdom. His primary area of research is integration of spatial data.