'Intelligent' Dasymetric Mapping and Its Comparison to Other Areal Interpolation Techniques

Jeremy Mennis

Department of Geography and Urban Studies, 1115 W. Berks St., 309 Gladfelter Hall, Temple University, Philadelphia, PA 19122
Phone: (215) 204-4748, Fax: (215) 204-7388, Email: jmennis@temple.edu

Torrin Hultgren

Department of Geography, UCB 360, University of Colorado, Boulder, CO 80309
Email: torrin47@yahoo.com

Abstract

In previous research we have described a dasymetric mapping technique that combines an analyst's subjective knowledge with a sampling approach to parameterize the reapportionment of data from a choropleth map to a dasymetric map. In the present research, we describe a comparison of the proposed dasymetric mapping technique with the conventional areal weighting and 'binary' dasymetric mapping techniques. The comparison is made using a case study dasymetric mapping of U.S. Census tract-level population data in the Front Range of Colorado using land cover data as ancillary data. Error is quantified using U.S. Census block-level population data. The proposed dasymetric mapping technique outperforms areal weighting and certain parameterizations of the proposed technique outperform 'binary' dasymetric mapping.

1 Introduction

A dasymetric map depicts a statistical surface as a set of enumerated zones, where the zone boundaries reflect the underlying variation in the surface and represent the steepest surface escarpments (Dent, 1999). Dasymetric mapping has its roots in the work of Russian cartographer Semenov Tian-Shansky (Bielecka, 2005) and American J.K. Wright (1936). Later significant contributions to dasymetric mapping were made by O'Cleary (1969) and, concurrent with more recent advances in GIS and environmental remote sensing, researchers in spatial analysis (Goodchild et al., 1993; Wu et al., 2005). The process of dasymetric mapping can be considered the transformation of data encoded in a choropleth map to a dasymetric map (Eicher and Brewer, 2001). Dasymetric mapping uses an ancillary data set to reapportion data from the choropleth map zones to the dasymetric map zones. The vast majority of dasymetric mapping research has focused on population data, though dasymetric mapping can be applied to any punctiform data which can be modeled as a statistical surface.

In previous research we described a new íntelligent' dasymetric mapping technique, so-called because it combines an analyst's subjective knowledge with an empirical sampling approach to parameterize the relationship between the ancillary data and underlying statistical surface for purposes of reapportionment (Mennis, 2003; Mennis and Hultgren, 2005). Here, an analyst may manually 'preset' an ancillary class to a particular population density, or allow the technique to estimate the population density of the class by sampling choropleth map zones associated with that ancillary class. These choropleth map zones may be sampled using the 'centroid,' 'contained,' or 'percent cover,' methods. With the centroid method, choropleth map zones are associated with an ancillary class if their centroid falls within that class. With the contained method, choropleth map zones are associated with an ancillary class if they are wholly contained with that class. With the percent cover method, choropleth map zones are associated with an ancillary class if their overlap with a particular class exceeds a user defined threshold (e.g. 80%).

In the present research, we present a comparison of the intelligent dasymetric mapping technique with two conventional approaches, areal weighting and 'binary' dasymetric mapping. Areal weighting is the most basic form of areal interpolation whereby a homogeneous data distribution is assumed to occur within each choropleth map zone. Thus, a dasymetric map zone's population can be calculated by summing the population of the choropleth map zones with which the dasymetric map zone overlaps, where each choropleth map zone contributes the product of its total population multiplied by the percentage of its total area that overlaps the dasymetric map zone (Goodchild and Lam, 1980). The 'binary' dasymetric mapping technique is simply the exclusion of population from uninhabited areas, however they may be defined, and

the consequent apportionment of the entire population to the inhabited areas of the map (Eicher and Brewer, 2001).

2 Data and Methods

The comparison of methods is based on the dasymetric mapping of U.S. Census tract-level total population data for the Front Range of Colorado to sub-tract units (Figure 1). The ancillary data are land cover data generated from manual interpretation of 1996-1997 aerial photography as part of the U.S. Geological Survey's Front Range Infrastructure Resources Project (Stier, 1999). For the case study, each polygon was classified as one of the following land covers (Figure 2): high density residential, low density residential, non-residential developed, vegetated, or water.

Using these population and land cover data, a series of maps were created using the intelligent dasymetric mapping technique, as well as using areal weighting and the traditional binary dasymetric mapping technique. Each of the different sampling methods – containment, centroid, and percent cover – was used. For the percent cover sampling method, percent cover thresholds of 70%, 80%, and 90% were employed. Each sampling method was also applied using no manually preset ancillary class data density values, and manually preset values of zero data density for the non-residential developed and water land covers. In addition, a regions layer
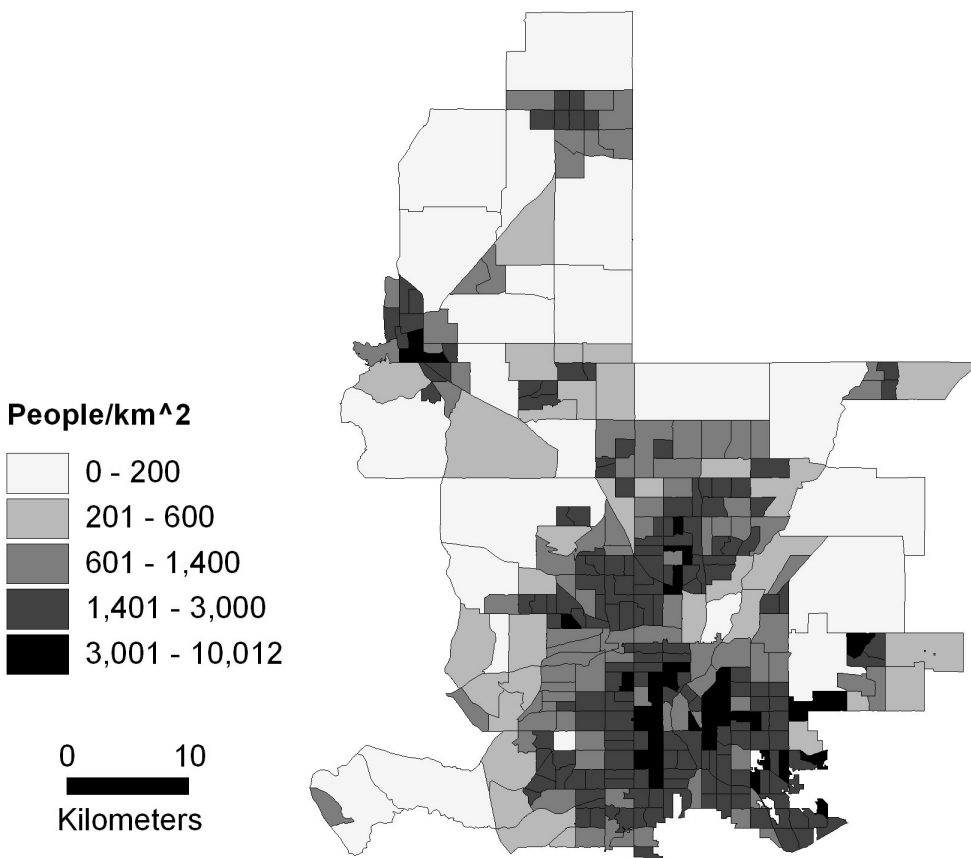

Figure 1. Tract-level map of total population for the study area.

of the counties was also employed to support a independent calibration of the intelligent dasymetric mapping technique for each individual region (Figure 2). Each sampling method was run with regions and without the use of regions. In all, the following 19 maps were created:

Conventional Approaches
1. Areal Weighting
2. Binary (zero data distributed to non-residential developed and water land covers; areal weighting used to distribute the data to remaining land covers)

Intelligent Dasymetric Mapping Technique
3. Centroid Sampling without Regions and with Presets
4. Contained Sampling without Regions and with Presets
5. Percent Cover (70%) Sampling without Regions and with Presets
6. Percent Cover (80%) Sampling without Regions and with Presets
7. Percent Cover (90%) Sampling without Regions and with Presets
8. Centroid Sampling with Regions and Presets
9. Percent Cover (70%) Sampling with Regions and Presets
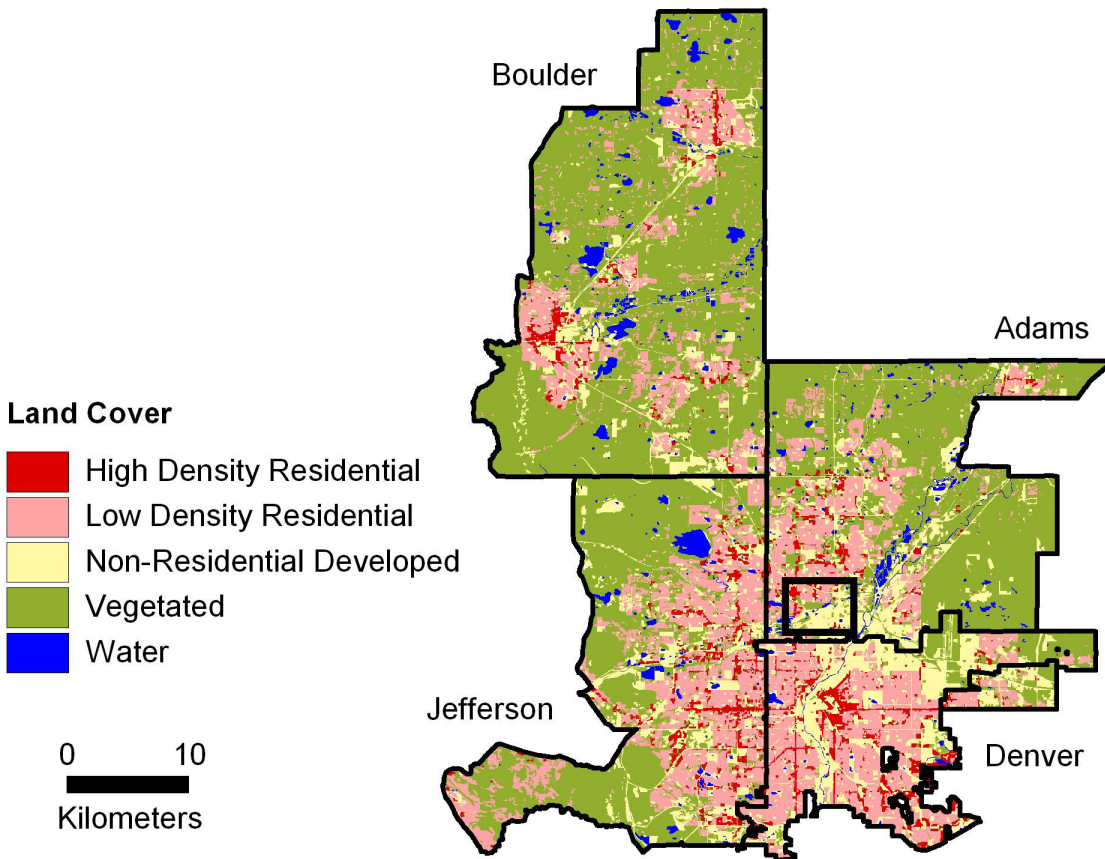10. Percent Cover (80%) Sampling with Regions and Presets

Figure 2. Land cover map of the study area (detail of boxed area shown at bottom). County boundaries, clipped to the study area, and county names are also shown for

11. Percent Cover (90%) Sampling with Regions and Presets
12. Centroid Sampling without Regions and without Presets
13. Percent Cover (70%) Sampling without Regions and without Presets
14. Percent Cover (80%) Sampling without Regions and without Presets
15. Percent Cover (90%) Sampling without Regions and without Presets
16. Centroid with Regions and without Presets
17. Percent Cover (70%) with Regions and without Presets
18. Percent Cover (80%) with Regions and without Presets
19. Percent Cover (90%) with Regions and without Presets

The difference between the estimated and actual population data for each U.S. Census block was then calculated, and the root mean square (*RMS*) error calculated for each tract, where the *RMS* is the average count error of all the blocks within each tract. The *RMS* was then normalized by the tract's actual population to yield the coefficient of variation (*CV*; Eicher and Brewer, 2001). These *CV* scores are entered into an analysis of variance (ANOVA) test to determine whether there is a significant difference in means among the *CV* values of the 19 different maps. Because the Levene statistic indicates that the assumption of homogeneity of variances among the different groups is rejected, the Tamhanes T2 post-hoc test is used to indicate whether there is a significant difference in means between each pairwise combination of the 19 maps.

3 Results

As an example of the resulting dasymetric maps, Figure 3 shows the map of total population produced using the intelligent dasymetric mapping technique using centroid sampling with regions and presets (#8). Clearly, the map presented in Figure 3 offers a far more detailed depiction of population density than the analogous choropleth map shown in Figure 1. This is particularly true in suburban and exurban areas, where the tracts tend to be larger and the land cover tends to be particularly heterogeneous at the transition from urban to rural land uses. Figure 4 shows a block-level count error map for the map shown in Figure 3, where count error is calculated as the actual population subtracted from the estimated population of the block. The mean count error is zero and the standard deviation is 84. Clearly, a far greater area of blocks is subject to overestimation, as compared to underestimation, at greater than one standard deviation. This reflects the fact that relatively large rural blocks tend to be overestimated while relatively small urban blocks tend to be underestimated. In the study region, these overestimated rural areas occur primarily on the western border, where the plains meet the foothills of the Rocky Mountains. These areas are typically large swaths of sparsely populated shrubland, encoded as part of the vegetated class on the land cover map (Figure 2).
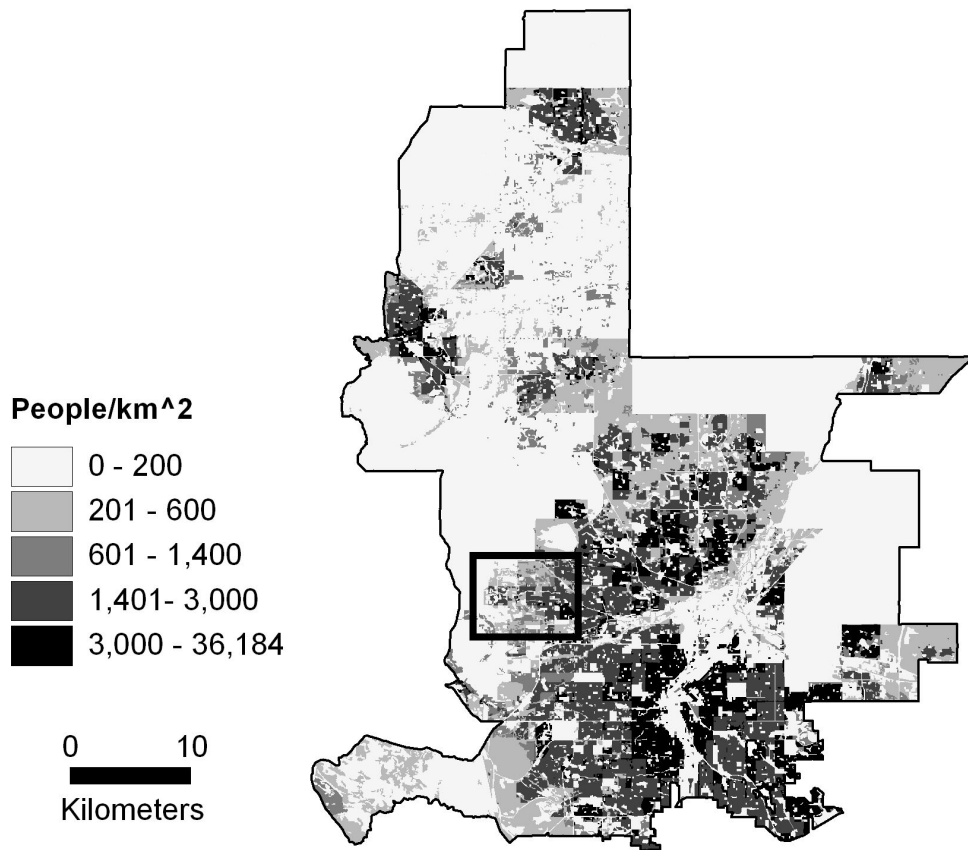
Figure 3. The dasymetric map of total population density produced using the centroid sampling method with regions and presets of zero density for non-residential developed and water land covers. The class interval and color schemes are the same as for Figure 1.

Figure 5 provides a chart of the mean *CV* scores for each of the 19 methods. Areal weighting (#1) has the highest *CV* value, indicating relatively poor performance compared to the other methods. Those intelligent dasymetric mapping methods using presets (#3-#11) generally outperform those methods without presets, though this pattern is tempered by the variability introduced by the use of regions and the different threshold settings for the percent cover sampling method. Those methods using presets also outperform the binary method; and the binary method's performance is approximately equal to that of those methods which did not use presets.

The ANOVA reveals that there are significant differences in means among the methods (Table 1). Table 2 reports the Tamhane's post-hoc test of significant difference in mean for each pairwise combination of methods. There is no significant difference in mean *CV* between areal weighting and the binary method. All the intelligent dasymetric mapping parameterizations using presets were significantly different from areal weighting, and more than half were significantly different from the binary dasymetric mapping technique.
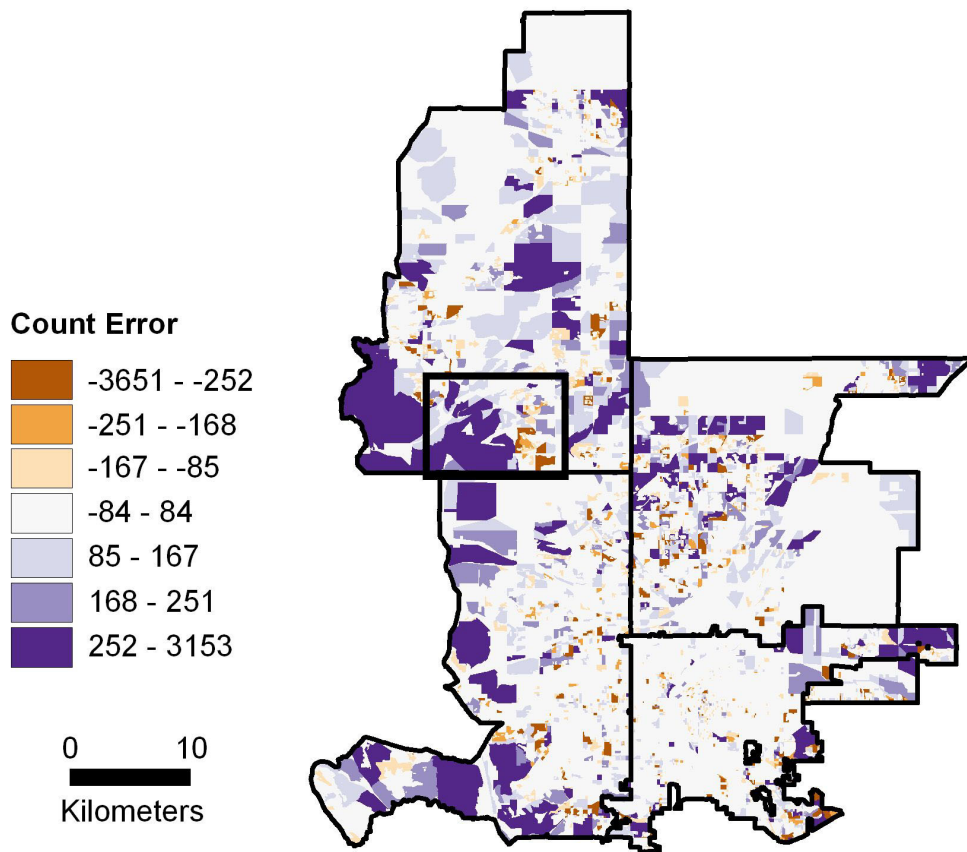
Figure 4. A map of the count error by block for the dasymetric map presented in Figure 3. Class intervals are by standard deviation from the mean error, which is zero. Red areas indicate underestimation of population while blue areas indicate overestimation.

Table 1. Results of the ANOVA of the *CV* Scores of the 19 Methods.				
		Mean Square	**F**	**Significance**
Total Population	Between Groups	0.002857	5.985452	1.11368E-14
	Within Groups	0.000477		

4 Conclusion

This research suggests that the intelligent dasymetric mapping technique outperforms areal weighting and can outperform binary dasymetric mapping with appropriate parameter settings. The use of preset data density values, even for relatively straightforward relationships, such as prohibiting population from areas covered by water, is particularly important for deriving better quality maps of population distribution. The use of regions did not improve the performance of the technique, though this is likely due to the weak utility of county boundaries as indicators of regional parameterization of the functional relationships between land cover and population density. In future research we will extend this study by testing other variables besides

total population. In addition, we have programmed the intelligent dasymetric mapping technique to output a summary statistics file with each dasymetric map, in order to inform the analyst about the character of the sampling and the overall quality of the dasymetric map.
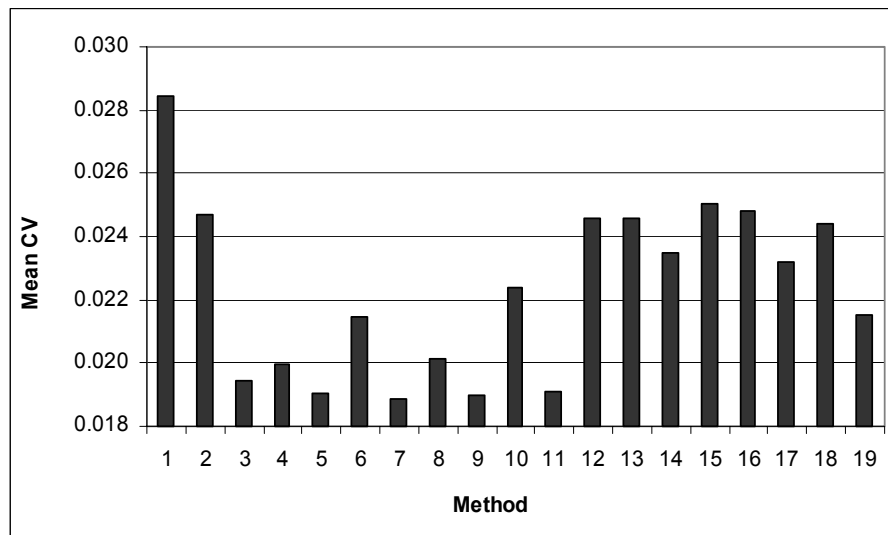

Figure 5. Mean *CV* score of each method.

References

Bielecka, E, "A dasymetric population density map of Poland", in *Proceedings of the 22nd International Cartographic Conference*, July 9-15, A Coruna, Spain (CD).

Dent, B D, 1999, *Cartography: Thematic Map Design*, Fifth Edition (WCB/McGraw-Hill).

Eicher, C L and Brewer, C A, 2001, "Dasymetric mapping and areal interpolation: implementation and evaluation" *Cartography and Geographic Information Science* **28** 125-138.

Goodchild, M F, Anselin, L, and Deichmann, U, 1993, "A framework for the areal interpolation of socioeconomic data" *Environment and Planning A* **25** 383-397.

Goodchild, M F and Lam, N S, 1980, "Areal interpolation: a variant of the traditional spatial problem" *Geo-processing* **1** 297-312.

McCleary, G F, Jr., 1969, *The Dasymetric Method in Thematic Cartography*. Unpublished Ph.D. dissertation, University of Wisconsin.

Mennis, J, 2003, "Generating surface models of population using dasymetric mapping" *The Professional Geographer* 55 31-42.

Mennis, J. and Hultgren, T., 2005. "Dasymetric mapping for disaggregating coarse resolution population data" *Proceedings of the 22nd Annual International Cartographic Conference*, July 9-16, A Coruna, Spain (CD).

Stier, M, 1999, *Temporal land use and land cover mapping.* Paper presented at the American Society for Photogrammetry and Remote Sensing (ASPRS) Conference, Portland, Oregon (WWW document, http://rockyweb.cr.usgs.gov/frontrange/publications.htm, last accessed January 3, 2006).

Wright, J K, 1936, "A method of mapping densities of population" *The Geographical Review* 103-110.

Wu, S-S., Qiu, X, and Wang, L, 2005, "Population estimation methods in GIS and remote sensing: a review" *GIScience and Remote Sensing* **42** 80-96.

Table 2. Results Of ANOVA Tamhanes T2 Posthoc Test of Significant Difference in Mean *CV* Score Between Each Pair-Wise Combination of Dasymetric Mapping Methods.

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1																			
2																			
3	x	x																	
4	x																		
5	x	x																	
6	x																		
7	x	x																	
8	x																		
9	x	x																	
10	x																		
11	x	x																	
12					x		x		x		x								
13					x		x		x		x								
14																			
15																			
16			x		x		x		x		x								
17																			
18							x		x										
19	x																		

Note: Methods are listed 1-19 on the X and Y axes (see text for the method associated with each number). A significant difference between methods at the 0.10 level is indicated by a 'x' at the intersection of two methods.