Assessing Population Surface Models using the Northern Ireland Census Grid Square Resource

C. D. Lloyd[1], D. J. Martin[2], I. G. Shuttleworth[1]

[1]School of Geography, Archaeology and Palaeoecology, Queen's University, Belfast, BT7 1NN
Telephone: +44 (0)28 9097 3478, Fax: +44 (0)28 9097 3212, Email: c.lloyd@qub.ac.uk

[2]School of Geography, University of Southampton, Highfield, Southampton, SO17 1BJ

1. Introduction

Direct comparison of populations using census data for different time periods is usually problematic because of (i) change in the zonal units used to report counts and (ii) change in the questions asked or the categories used for producing census outputs. This paper is concerned with the first of these two problems. In the absence of consistent zonal units, it is often necessary to reallocate counts from one set of zones to another or to generate a population surface which can be used in its own right or the counts aggregated in another set of zones. Assessment of the performance of areal interpolation procedures often relies on using synthetic data. In the case of population surface models direct assessment of estimates on a cell-by-cell basis is usually impossible due to the absence of actual population counts. The present paper makes use of gridded counts provided as an output of the 2001 Northern Ireland Census of Population to assess the performance of methods for generation of population surfaces.

A variety of routines for generation of population surfaces have been developed (for example, Tobler, 1979; Martin, 1989). In this paper the approach of Martin (1989, 1996) is applied and assessed. Through use of Northern Ireland grid square census data, different kinds of errors are identified and the importance of selection of an appropriate data neighbourhood and distance decay function is assessed. The paper closes with some preliminary recommendations for those who wish to generate population surface models but do not have access to grid square data.

2. Data and methods

2.1 Data

Total population surfaces were generated from Output Area population weighted centroids. These surfaces were compared to counts provided within 100 m^2 grid cells. In 1971 grid square counts were provided in Great Britain and Northern Ireland, but only in Northern Ireland was release of grid-square data continued beyond that census. In 1971, 1981 and 1991, 1 km^2 grid squares were available for the whole of Northern Ireland while 100 m^2 grid squares were provided only for urban areas. In 2001, both 1 km^2 grid squares and 100 m^2 grid squares were made available for the whole of Northern Ireland. The grid-square product is described by Shuttleworth and Lloyd (submitted).

2.2 Population surface modelling

Martin (1989) outlines a method for mapping population from zone centroids. With this approach, each zone centroid is visited in turn and the mean intercentroid distance is calculated within a predefined search radius. This measure indicates the unknown areal extent of zones in the region and it is used to calibrate a distance-decay function that assigns weights to cells in the output grid. The cells that are closest to the zone centroid receive the largest weights, while those cells estimated to be located in the maximum areal extent of the zone receive the smallest weights. The population is then redistributed in the surrounding region using these weights. A given cell in the output grid may receive population values from one or more centroids, or may remain unpopulated as it is beyond the area of influence of any of the centroid locations. Bracken and Martin (1995) apply the method for linking 1981 and 1991 censuses of Britain, producing a dataset available at http://census.ac.uk/cdu/software/surpop/.

Using this approach, population is preserved globally (the sum of populations in the zones is the same as the sum of populations in the population surface), but the sum of the number of people in a given zone does not necessarily correspond to the overlapping area in the population surface. That is, no account is taken of zone boundary location and population may be gained from or lost into neighbouring zones through applying the method. For this reason, Martin (1996) presented an adapted version of the method. In this modified version, a rasterised zone map with the same cell size as the required output surface is acquired first. As population is redistributed, the weights for cells located outside the current zone are automatically set to zero — it is locally mass preserving. Both the original (global mass preservation) and updated (local mass preservation) versions of the method are applied and assessed in this paper.

3. Results

The analysis was based on 2001 OA population weighted centroids using total population, reallocating counts to a grid with 50m spacing with the same minimum and maximum x and y co-ordinates as the 100m grid data. The counts on 50m cells were then aggregated to form 100m grid cells that could be compared directly with the grid square count data. Table 1 gives, for the whole of Northern Ireland, summary statistics when the estimated surfaces were subtracted from the 100m grid data counts (so, negative values indicate under-estimation) for search radii of 250 m and 500 m and for both global and local versions of the routine. The mean error suggests that the as the search radius is increased the tendency to under-estimate tends to increase.

Type mass preserve.	Search radius	Max. neg. error	Max. pos error	Mean error	Std. dev.
Global	250	-2561.021	495.987	-7.351	24.501
Global	500	-2529.398	152.559	-8.055	22.748
Local	250	-2560.856	206.859	-7.455	24.205
Local	500	-2515.798	137.366	-8.139	22.890

Table 1. Northern Ireland: population surface estimation errors (estimates – 100m grid cell counts), distance decay parameter = 1. Local = mass preservation within OAs.

The results in Table 1 are for a distance decay parameter of 1, which approximates to linear distance decay. Alternative distance decay parameters are likely to yield more accurate estimates in different situations. In densely populated urban areas a large distance decay parameter value is likely to be more appropriate than in sparsely populated rural areas. In Table 2, error summary statistics are given for a search radius of 500 m and with distance decay parameter values of 0.25 and 2.

Type mass preserve.	Distance decay par.	Max. neg. error	Max. pos error	Mean error	Std. dev.
Global	0.25	-2543.217	105.572	-8.445	23.025
Global	2	-2526.193	187.487	-7.729	22.863
Local	0.25	-2521.212	143.382	-8.417	22.984
Local	2	-2519.405	146.720	-7.894	23.012

Table 2. Northern Ireland: population surface estimation errors (estimates – 100m grid cell counts), search radius = 500. Local = mass preservation within OAs.

Tables 3 and 4 are the equivalent of Tables 1 and 2 respectively but for the Belfast Local Government District rather than for Northern Ireland as a whole.

Type mass preserve.	Search radius	Max. neg. error	Max. pos error	Mean error	Std. dev.
Global	250	-1268.321	362.673	-7.991	44.350
Global	500	-1234.875	101.938	-10.460	41.441
Local	250	-1262.994	206.859	-8.422	44.394
Local	500	-1230.042	137.366	-9.846	42.518

Table 3. Belfast: population surface estimation errors (estimates – 100m grid cell counts), distance decay parameter = 1. Local = mass preservation within OAs.

Type mass preserve.	Distance decay par.	Max. neg. error	Max. pos error	Mean error	Std. dev.
Global	0.25	-1239.084	97.099	-11.805	42.568
Global	2	-1241.165	108.311	-9.375	41.013
Local	0.25	-1225.985	143.382	-10.375	42.614
Local	2	-1237.622	146.720	-9.340	42.727

Table 4. Belfast: population surface estimation errors (estimates – 100m grid cell counts), search radius = 500. Local = mass preservation within OAs.

The figures suggest that, as expected, a large distance decay parameter is more appropriate in densely occupied urban areas as the mean errors tend to be closer to zero for large values (i.e., 1, 2) than for a value of 0.25. The local mass preserving approach gives less accurate estimates, judging by the mean and the standard deviation of the errors, than does the original global mass preserving approach. This may be partly due to the size of the grid used — 50 m cells sometimes straddle two OAs and forcing all of the population of that cell to belong to one OA is erroneous. Future work will assess the impact of using smaller grid sizes.

Figure 1 shows, for the Belfast area, estimation errors for a search radius of 500 m with local mass preservation. The estimation errors appear visually random. There is some suggestion that errors are larger (whether negative or positive) in areas where the density of OA centroids is greater. It also appears that there is a tendency to under-estimate where the population densities are larger and this is intuitively sensible.

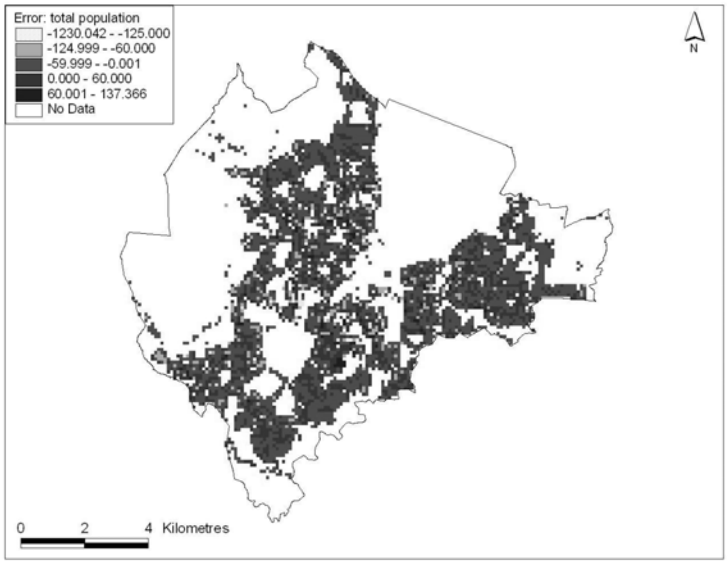

Figure 1. Population surface estimation errors (estimates – 100m grid cell counts): 500 m search radius, local mass preservation.

4. Conclusions and future work

This brief paper provides results from preliminary analyses. The intention is to provide results that other users with no access to grid square data can use to inform the approach that they employ. By assessing differences due to method, parameters (e.g., search bandwidth) and underlying population characteristics gaining an insight into appropriate procedures in particular situations is feasible.

Future work will focus on assessing spatial variation in estimation errors. Detailed examination of the impact locally of applying different search radii and different distance decay parameters will be conducted. Following this work, the possibility of using, for example, different distance decay parameters can be considered.

5. Acknowledgements

The Northern Ireland Statistics and Research Agency (NISRA) are thanked for the provision of 2001 Northern Ireland Census of Population grid square data.

6. References

BRACKEN, I. and MARTIN, D., 1995, Linkage of the 1981 and 1991 Censuses using surface modelling concepts. Environment and Planning A, 27, pp. 379–390.

MARTIN, D., 1989, Mapping population data from zone centroid locations. Transactions of the Institute of British Geographers, New Series, 14, pp. 90–97.

MARTIN, D., 1996, An assessment of surface and zonal models of population. International Journal of Geographical Information Systems, 10, pp. 973–989.

SHUTTLEWORTH, I. G. and LLOYD, C. D., submitted, Are Northern Ireland's communities dividing? Evidence from the Census of Population 1971–2001. Environment and Planning A.

TOBLER, W. R., 1979, Smooth pycnophylactic interpolation for geographical regions Journal of the American Statistical Association 74, pp. 519–530.