Creating Population Surfaces using Information on the Spatial Structure of Variables

Christopher D. Lloyd[1], Paul Williamson[1], Gemma Catney[1], Chris Brunsdon[1], and Alex Singleton[1]

[1]Department of Geography and Planning, School of Environmental Sciences,
University of Liverpool, Roxby Building, Chatham Street, Liverpool, L69 7ZT UK
Tel. (+44) 0151 794 2857
C.D.Lloyd@Liverpool.ac.uk

KEYWORDS: Areal interpolation, Kernel estimation, Spatial dependence, Census data, Scale

1. Introduction

The reallocation of counts from one set of zones to another (areal interpolation) is a common objective in Census research. Examples include to account for boundary changes between Censuses, or from higher to lower level geographies. Possible approaches include kernel smoothing, to create a population grid, and regression based on land use data. Variables may exhibit very different spatial patterns – for example, there may be more variation in employment or educational status across an area than there is with respect to car ownership. Thus, the approach used to create a population grid in the first case should be adapted to account for the heterogeneity of a variable. Such approaches are needed since any analysis based on area data, such as counts for wards provided as outputs from the UK Census of Population, is partly a function of the size and shape of those zones. In addition, where zones change between Censuses, direct comparisons for different Census dates are not possible. If the 2011 Census is to be the last in the UK, then there will be an even greater need for flexible approaches to mapping populations using diverse data sources.

Most current attempts to create population surfaces are based on (i) kernel smoothing used in isolation (see, for example, Martin 1989, 1996) or (ii) areal reallocation informed by external datasets such as land use data. In the present research, possible benefits of combing these two approaches are explored. Specifically, a kernel smoothing approach, with the weighting adjusted according to the variable being mapped, will be used to reallocate counts subject to land use classes. The paper makes three specific contributions. Firstly, it will provide population surfaces for a range of variables for the 1991, 2001 and 2011 Censuses, thus providing an original perspective on population change (initially in England, but expanding to include the rest of the UK). Secondly, it will systematically assess the performance of a range of methods for generating population surfaces. Thirdly, using the most recent release of Census outputs, the resulting series of outputs will be the most up to date for the UK.

2. Data and methods

2.1 Data

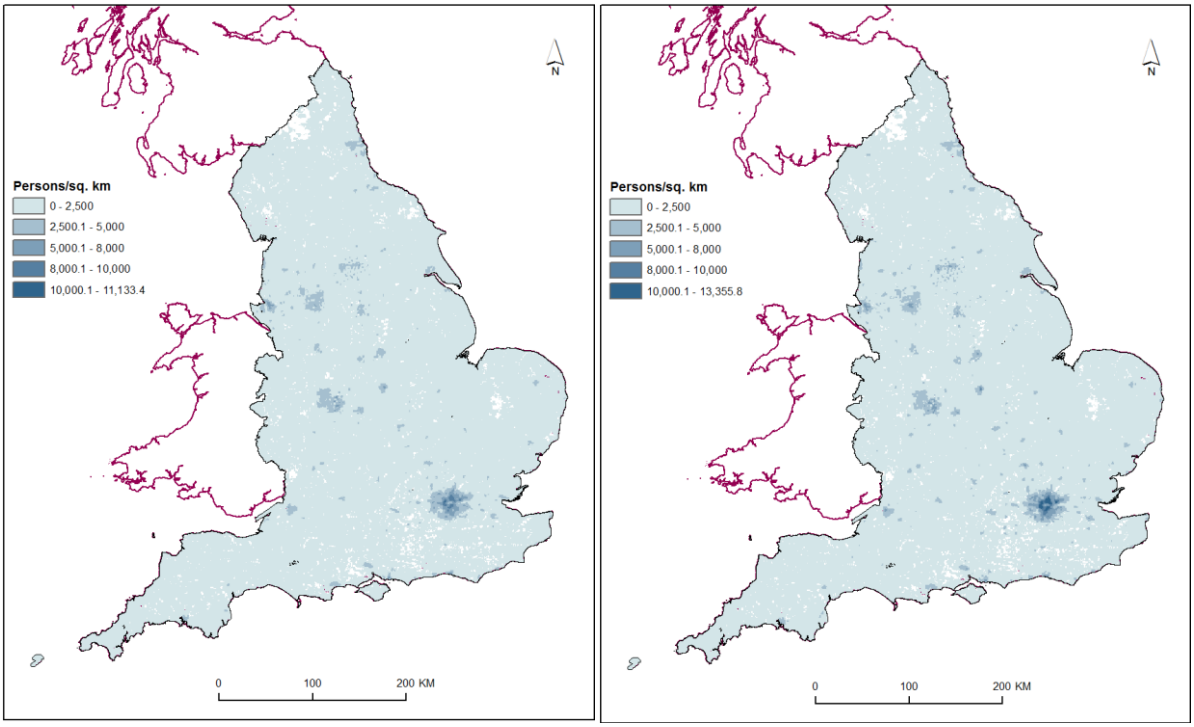
The analysis focuses on total population for 1991 (by Ward; *n* = 8619) and 2001 (by Census Area Statistics (CAS) Wards; *n* = 7969), with 2011 Census data to be included once they are available. In addition, some results are discussed relating to the number of persons over the age of 65, and the numbers of persons with a limiting long term illness (LLTI) in 1991. Land use data are derived from Meridian ™2 vector data, available through the Ordnance Survey OpenData initiative.

2.2 Methods

Surfaces are generated from the centroids of CAS Wards using kernel estimation, using the quartic kernel (Bailey and Gatrell, 1995), but future work will make use of the SurfaceBuilder program developed by David Martin[1] (or code based on this). The initial approach consists of conducting kernel estimation using a mask defined by particular land uses where there are likely to be no people – woodland in the example below.

3. Analysis

Figure 1 shows a population surface representing the total population of England in (a) 1991 and (b) 2001. These grids were generated with kernel smoothing with a bandwidth of 2.5km, and wooded areas were excluded from the population surfaces. Clearly, this is one very specific example of a landuse feature which is taken into account.



(a) (b)

Figure 1. Persons per square km in (a) 1991, (b) 2001: 1km square grid.

[1] http://www.public.geog.soton.ac.uk/users/martindj/davehome/software.htm

Figure 2 shows the 2001 population totals in Figure 1(b) subtracted from those for 1991 (Figure 1(a)). The most notable gains are in the London area and, less obviously, other urban areas including Birmingham and major cities of the north west.

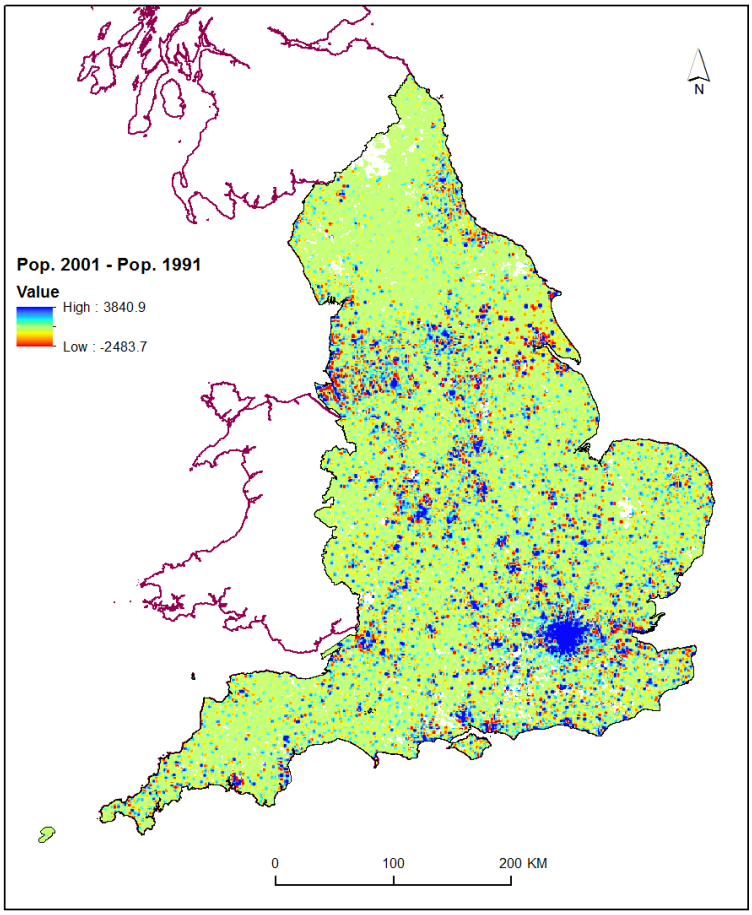
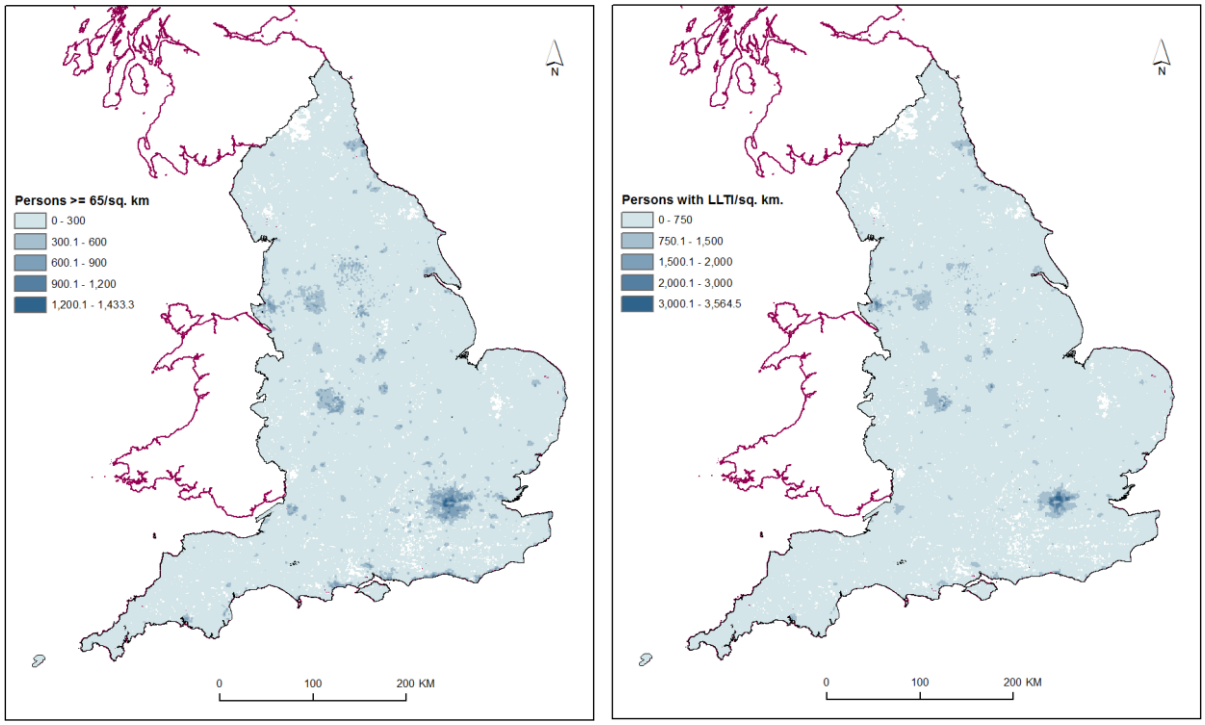


Figure 2. Persons per square km in 2001 - Persons per square km in 1991: 1km square grid.

This approach can readily be extended to individual variables such as persons by age, or by health status. However, variables may have very different spatial distributions and, therefore, the shape and size of the kernel used could be adapted to these variables. Figure 4 shows (a) Persons >= 65 per square km in 2001 (b) Persons with a LLTI per square km in 2001. In this case, both surfaces were generated with the same form of kernel, but the spatial structure of these two variables is quite different. As one indicator, the Moran's *I* spatial autocorrelation coefficient for Persons >= 65 (%) by CAS Ward using inverse distance weighting (with row-standardised weights) is 0.36, for LLTI (%) is 0.57. Thus, LLTI (%) is more clustered than Persons >=65, and, for the age variable, a relatively 'flatter' distance decay function might be appropriate.



(a) (b)

Figure 3. (a) Persons >= 65 per square km in 2001 (b) Persons with a LLTI per square km in 2001: 1km square grid.

4. Discussion and conclusions

The paper outlines some provisional results of population surface models with the dual objectives of (i) exploring population change in England between 1991 and 2001 (with extension to 2011 in due course) and (ii) assessing how kernel smoothing and areal reallocation approaches can best be used in isolation and in combination.

An obvious addition to this analysis is the use of geostatistical approaches to population surface modelling. The present analysis provides some context to future work which will fully assess when it is useful to make use of information on the spatial structure of individual variables through variograms estimation and modelling, and deconvolution to estimate the point support variogram. In cases where rich population data are available in conjunction with spatially-detailed secondary data sources, simple areal reallocation approaches may be sufficient. However, in more sparsely populated rural areas in particular, it may be that spatial correlation structure of variables has a more important role in terms of the accuracy of population counts re-allocated to grid cells.

5. Acknowledgements

The Office for National Statistics are thanked for provision of the data. Office for National Statistics, 2001 Census: Digitised Boundary Data (England and Wales) [computer file]. ESRC/JISC Census Programme, Census Geography Data Unit (UKBORDERS), EDINA (University of Edinburgh)/Census Dissemination Unit. Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

References

Bailey, TC and Gatrell AC (1995) *Interactive Spatial Data Analysis*. Longman Scientific and Technical, Harlow

Martin D (1989) Mapping population data from zone centroid locations *Transactions of the Institute of British Geographers, New Series* 14 90 – 97

Martin D (1996) An assessment of surface and zonal models of population *International Journal of Geographical Information Systems* **10** 973 – 989

Biography

Chris Lloyd is a Senior Lecturer in Human Geography. His research focus is on spatial data analysis, and in particular on local spatial statistics and the exploration of spatial scale. He has a particular interest in the conceptualisation and measurement of residential segregation, and is also engaged in analyses of distortion in historic maps.

Paul Williamson is a Senior Lecturer in Human Geography. He is a Population Geographer with interests in microsimulation modelling and Census data analysis.

Gemma Catney is a Leverhulme Trust Early Career Research Fellow engaged in her project 'Geographies of ethnic and social segregation in England and Wales, 1991-2011'. She is a Population Geographer by background, and in addition to segregation by ethnic group in Britain and religion in Northern Ireland, she has research interests in internal migration with an ethnic group dimension.

Chris Brunsdon is Professor of Human Geography at the University of Liverpool. He has a particular interest in the modelling and analysis of social data, the visualisation of geographical data and the use of local statistics.

Alex Singleton is a Lecturer in Urban Planning. He has research interests in geodemographics and the geography of access to higher education.