Exploring the utility of grids for analysing long term population change

C.D. Lloyd [a,*], G. Catney [a], P. Williamson [a], N. Bearman [b]

[a] Centre for Spatial Demographics Research and Department of Geography and Planning, School of Environmental Sciences, University of Liverpool, UK
[b] Clear Mapping Co, Waterside House, Falmouth Road, Penryn, Cornwall, UK

ARTICLE INFO

ABSTRACT

This paper details an innovative approach which enables the analysis of small area population change across four decades. Population surfaces are generated using small area data (enumeration districts or output areas) for each Census from 1971 to 2011 inclusive. The paper details the methods used in the creation of these surfaces, and discusses the rationale behind this approach, arguing that grids represent the most appropriate model for assessing population distributions. Methods for grid creation are tested using pre-existing population grids for Northern Ireland as a benchmark. The method developed is then applied to create population grids for the rest of the UK for 1971, 1981, 1991, 2001 and 2011. The changing population structures of small areas across these five time points are explored here to illustrate the value of this approach. The publically-available data resource – the final product of the 'PopChange' project – will facilitate exploration of long-term changes in populations over small areas. The paper argues that maximum advantage could be taken of the 'big data revolution' if such data were gridded in a similar way, allowing them to be placed in a longer-term historical context, using tools made available through the PopChange project.

1. Introduction

Analyses of change over time in small geographical areas are restricted by the availability of common variables and geographies (Martin, Dorling, & Mitchell, 2002). Many such analyses use variables which have similar, but not identical, definitions, and results are then interpreted with the caveat that definitional and 'true' change are conflated to some degree. If there are differences in the size and/or shape of geographical zones used in such comparative analyses, then it is necessary to compare areas on a 'best fit' basis, or to transfer counts from the original (source) zones to a set of zones which are common for the time periods being compared. There are many reasons why analyses of long-term population change and, therefore, approaches for making datasets comparable over multiple years, are important. As an example, areas with a former dependency on heavy industry and subsequent industrial decline are often associated with poorer general health than areas where the labour market has remained buoyant (Stillwell, Norman, Thomas, & Surridge, 2010). In this context, developing approaches which enable us to explore long-term health patterns over small areas would be invaluable. By measuring how much deprivation has increased in areas over several decades, or by identifying areas with stubbornly high levels of deprivation, we can begin to unpick how persistence or change in deprivation and in other characteristics of populations relates to key outcomes such as health status or educational attainment. We can also begin to evaluate policy and practice in local and national government. Intervention strategies have focused on targeting areas with particular social needs, but how far have these interventions altered the deprivation trajectories of neighbourhoods? The rationale for the development of locally-based strategies (e.g., see Broughton, 2016) can be more fully assessed using measures of deprivation for long time periods.

This paper presents PopChange, a new resource for Britain for the period 1971 to 2011, which enables such questions to be answered. The resource developed is for Britain and the methodology is assessed using data for Northern Ireland, with the end result that comparable data are available for the whole of the UK (Britain and Northern Ireland). The paper also includes some case studies which illustrate its potential benefits; and some experimental results which support the approach to population surface creation finally adopted.

Previous research has sought to explore change in population characteristics over small areas of the UK and elsewhere. The 'Linking Censuses Through Time' project (Dorling, Martin, & Mitchell, 2001; Martin et al., 2002) sought to enable the comparison of 2001 Census data for Britain with data for 1971, 1981 and 1991. Norman, Rees, and Boyle (2003) reallocate population counts for 1991 to 2001 to a common set of zones (wards) using postcode centroid densities. Walford and Hayles (2012) detail a resource which provides an array of Census variables for common geographies from 1971 to 2001 for Britain. These studies are all based on consistent geographies for irregular zones, while the focus here is on surfaces, which have the advantage of removing reliance on any zonal system developed based on the population

* Corresponding author.
E-mail address: c.d.lloyd@liverpool.ac.uk (C.D. Lloyd).

distribution at one particular time point. Martin (1989, 1996) outlines a kernel estimation method for the creation of population surfaces (gridded population values) from standard output geographies comprising spatially and temporally irregular zones, thereby allowing direct comparison of data for different time periods. Martin's method has been tested using gridded population data for Northern Ireland (Martin, Lloyd, & Shuttleworth, 2011). Another approach based on smoothing using distance decay functions is detailed by Deng, Frantz, and Araoz (2017). Other methods for generating grids from irregular source data include pycnophylactic interpolation (Tobler, 1979), dasymetric approaches (e.g., Kim & Yao, 2010; Mennis, 2003) and geostatistical (kriging-based) methods (Goovaerts, 2008; Kyriakidis, 2004). Grid modelling is a specific variant of areal interpolation; reviews of areal interpolation methods (but not focusing on generation of grids) include Flowerdew and Green (1994), Gregory and Ell (2005) and Lloyd (2014).

Gridded counts offer several significant advantages over irregular zones and, in some countries, gridded population data are provided. In Northern Ireland, 1 km and 100 m cells containing an array of variables were published as outputs from all Censuses from 1971 to 2011 inclusive. In Britain, gridded counts were produced in 1971 but not for later Censuses. The 1971 counts were used to create the *Census Atlas of Britain* (CRU/OPCS/GROS, 1980). Gridded counts for Britain were not produced for the 1981 Census largely on the grounds of cost and "users were asked to pay for the grid referencing of the household records and this led to a proposed charge by OPCS of seven and a half times that for standard areas, such as EDs" (Denham & Rhind, 1983, p. 56). A general argument for not providing gridded counts is that it avoids the possibility of accidental data disclosure through the differencing of counts for gridded and statutory census geographies (Duke-Williams & Rees, 2002). Outside of the UK, gridded population data are available in several countries including Estonia, Finland, the Netherlands and Sweden. Eurostat, the agency responsible for production of EU-wide statistical information, requires that member states provide gridded total population grids[1] and for those countries which do not provide gridded outputs as standard (e.g., the UK, at least for the duration of its membership of the EU) these must be estimated. Population grids have been developed in many other national contexts, including in the USA (Mennis, 2003), and across Europe (Batista e Silva, Gallego, & Lavalle, 2013; Gallego, 2010). Estimated grids are developed and assessed for three countries (Vietnam, Cambodia, and Kenya) by Stevens, Gaughan, Linard, and Tatem (2015). Recent examples for Britain are 2011 population grids developed by Murdock et al. (2015; covering England and Wales) using Census, postcode and building data, and the Britain-wide grid produced using Census and land cover data by Reis et al. (2016).

All areal data are subject to the modifiable areal unit problem (MAUP) whereby the results of analyses are a function of the size and shape of areal units (Openshaw, 1984; Openshaw & Taylor, 1979; Wong, 2009). However, with grids the analyses are simplified as all units are of the same size and shape, and scale effects can be explored through simple aggregation of cells. In addition, a population grid 'smooths' out spatial population discontinuities which are an artefact of the underling arbitrary statutory geographies. More generally, grids represent populations which are arguably more true to the real world, and where there are no people there may be no cells (as with the approach applied here) – unlike standard areal data which tend to cover all land areas in the study region. This is of particular value in any studies which seek to assess interactions between areas, or clustering – there is likely no social meaning in a cluster which includes neighbouring zones whose shared boundaries are actually entirely unpopulated. Data on irregular zones are often constructed based on administrative criteria or statistical characteristics (e.g., population homogeneity) and grids avoid, to some degree, the potential subjectivities involved in the design of standard output geographies. Grids allow

assessment of change without using a set of irregular zones which were designed for one particular Census year – thus suburban areas at the edge of large cities may have been represented by large zones in 1971 but small zones in 2011 as their population densities grew. They allow straightforward assessment of changes in results with changes in scale – their size and shape are constant and they can be aggregated to explore how (for example) the clustering of population groups changes as the spatial resolution is coarsened. In addition, the availability of gridded population counts opens up the wide array of possibilities offered by image processing methods (e.g., see Sonka, Hlavac, & Boyle, 2015). Other researchers have made cases for gridded population data (e.g., Martin, 1996). An obvious additional benefit of population grids is that they are flexible and they can easily be compared to other grid models including environmental data (Gallego, 2010).

This paper details an approach to the generation of population grids using Census data for 1971, 1981, 1991, 2001 and 2011 for Britain. The resulting resource, *PopChange*, has been developed as a part of an Economic and Social Research Council (ESRC) funded project 'Population change and geographic inequalities in the UK, 1971–2011'. The *PopChange* project entailed:

- Identification of comparable variables from the UK (England and Wales, Scotland, Northern Ireland) Censuses of 1971, 1981, 1991, 2001 and 2011
- Creation of population surfaces for Britain (England, Scotland and Wales) for all comparable variables (1 km cells nationally and, in due course, 100 m cells for urban areas)
- Provision of population surfaces, code in R programming language (see R Core Team, 2016) to grid user-supplied data, and an interactive online atlas of population change (see https://popchange.liverpool.ac.uk/)
- Provision of project meta-data, including a project introduction and detailed background material

The gridded data made available have been generated from population data for each Census for the smallest areal units available (enumeration districts or output areas, depending on the year), with postcode centroid intensities used to help reallocate counts from input zones to output grids. The population surface modelling procedure was developed using Northern Ireland as a test case since counts for small area irregular zones *and* grids are available.

The data used are set out next: the data for the Northern Ireland benchmark study are detailed first, followed by the data for Britain. Next, the basic method for population surface modelling adopted is outlined, followed by some potential variants. Following this, the results of testing of the variants using data for Northern Ireland are summarised. The analyses of grids derived for Britain are detailed next; gridded total population counts are provided as examples of how population grids for multiple time points can be beneficial for enhancing our knowledge of changing population geographies.

2. Data

2.1. Northern Ireland

In Northern Ireland (NI), gridded population counts have been provided since 1971 (see Shuttleworth & Lloyd, 2009 for a summary) and thus these represent an opportunity to assess the accuracy of estimated population surfaces based on irregular source zones for the same national context (UK) as the resource. As context for the main focus in this paper, 1 km grid cell population counts for 2011 in NI[2] are used as a baseline and small areas (SAs; $n = 4537$, mean population $= 399$)

Table 1

Input zone counts and population totals taken from the source tables (GB-level Census counts for specified population base). The mean and maximum populations are for 1 km grid cells for all cells (populated and unpopulated with Britain; 239,855 cells in total). *From 1981 onwards the Census did not count all present people plus visitors. Including people who are not unusually resident, the 1981 count would have been 54,285,422 (1.4% higher).

Year	No. of input zones (ED or OA)	Population base	Census table	Total Census population	Mean population per 1 km cell	Maximum population
1971	125,476	Persons present	SAS01	53,845,575	224	22,098
1981	129,963	Usually resident	SAS01	53,556,911*	223	17,011
1991	147,613	Usually resident	SAS01	54,854,596	228	18,677
2001	218,038	Usually resident	KS001	57,103,927	238	21,623
2011	227,759	Usually resident	KS101	61,371,315	256	23,676

as source zones, with postcode centroids used to reallocate counts from SAs to target grid cells. The errors of estimation are then explored by subtracting the estimated values from the observed values. This process was undertaken for counts of the population by religion (specifically number of persons identifying as Catholic by religion or 'religion brought up in') and by limiting long term illness (LLTI), as well as the total number of households. The counts of Catholics and person with an LLTI were selected on the basis of the very different degrees of spatial dependence they exhibited, with religion varying much more smoothly than LLTI (see Lloyd, 2010 for an analysis of the spatial structure of population variables in NI). This element of the analysis builds on work by Lloyd (2017), who assessed the construction of population surfaces using land use data rather than postcodes. The latter provide more accurate estimates and, unlike land use data, are available for the early 1980s onwards; thus they were chosen for use in the present study.

2.2. Britain

The *PopChange* resource includes a wide range of variables which are comparable across all or some of the Censuses for 1971–2011 (see the project website for the full list). In the present paper, the focus is on total persons. This is used to highlight some key characteristics of the grid square resource and to demonstrate some of the ways in which the data can be analysed. Table 1 details the number of input (source) zones and the total population contained in the zones used as input to the grid creation process. The population base used in each case is also indicated. The 1971 base is *all present plus visitors*; the bases for other years are *usually resident*. Usual residence is generally defined as the address in the UK at which a person spends the majority of the time; in most cases this corresponds to a permanent or family home. The specific definition of usual residence varies between Censuses: for 1981, it is all present plus absent residents; for 1991 it is all present plus absent (with the addition of imputed wholly absent households). Note that the *PopChange* resource does not account for undercount in the 1991 Census (Dorling and Simpson, 1993), but future versions are intended to provide updated estimates. In 2001, the One Number Census scheme was used to estimate the total resident population[3] (and see ONS, 2004). The definition of usually resident for 2011 is provided by ONS (2009). The alternative bases for the Censuses of 1971 to 2001 inclusive are discussed by Walford and Hayles (2012). In a discussion about UK-level analyses, Walford (2002) notes a 3% difference across the study area in the total population for 1981 using the 1971 persons present base and the 1981 usual residents base; for 1991 the usually resident base (accounting for absent households) gave a total population count some 3% higher than the 1981 base. While the differences in population bases between Census years included in this study will affect results, the magnitude of the differences is likely to be small. Note also that totals in different Census area tables for the same year can vary because of adjustments made under the One Number Census scheme (Rees, Parsons, & Norman, 2005).

In this paper, population grids are generated for all Censuses from 1971 to 2011 inclusive for total persons. Fig. 1 summarises the main

inputs to each grid, noting the need for pre-processing of Census and postcode data where postcodes were used to inform reallocation of counts, as described above. The source zones are enumeration districts (EDs; 1971, 1981 and 1991 for England and Wales; 1971 for Scotland) or output areas (2001 and 2011 for England and Wales; 1981, 1991, 2001 and 2011 for Scotland). The data for 1971 and 1981, in particular, could not be used 'as is' and some modifications were required in these cases.

ED boundaries for 1971 are not digitally available, only ED centroids (as well as Thiessen polygons created from them). Therefore, the approach taken was to join 1971 ED centroids to 1981 EDs given that these small areas are the closest in date. More information on the procedure applied for processing data for each Census year is provided by Lloyd, Bearman, Catney, and Williamson (2017).

2.3. Postcode centroids as proxies for population density

Previous research has shown that postcodes are a proxy for population density and that they therefore provide a suitable basis for reallocating population counts. Norman et al. (2003) use household counts for each postcode whereby the number of postcode centroids in an area of overlap between source and target zones is used to determine the proportion of the source zone population to be allocated to the zone of intersection. However, household counts/delivery points by postcode are not available for the 1971, 1981 and 1991 Censuses. Instead postcodes were weighted equally, allowing a consistent approach to be employed for all Census years, avoiding the complexity of alternative assumptions and variable estimation accuracy for each Census year. Other possible alternatives include splitting residential and workplace postcode centroids, or using postcode areas rather than centroids where these are available. Also, landuse data or urban/rural classifications could be used in combination with postcodes to inform reallocation of counts. Future work will test refinements to the present approach where appropriate data are available for some or all Census years.

3. Population surface generation

The transfer of population counts between incompatible geographies can be conducted using areal weighting. Areal weighting was used as the basis for exploring three approaches to the creation of grids: (i) basic areal weighting, (ii) areal weighting using postcode centroid intensities and (iii) smoothing of outputs obtained using (ii). Each of these approaches is outlined in turn.

3.1. Areal weighting

Areal weighting entails the overlay of the source zones, *s* (EDs, OAs or SAs), and target zones, *t* (1 km grid cells), and the proportional allocation of the source zone population to the target zones which it overlays:

$$\hat{z}_t = \sum_{s=1}^{N} \frac{A_{st}}{A_s} z_s \tag{1}$$

---

[3] http://www.ons.gov.uk/census/2001censusandearlier/designandconduct/theonenumbercensus.

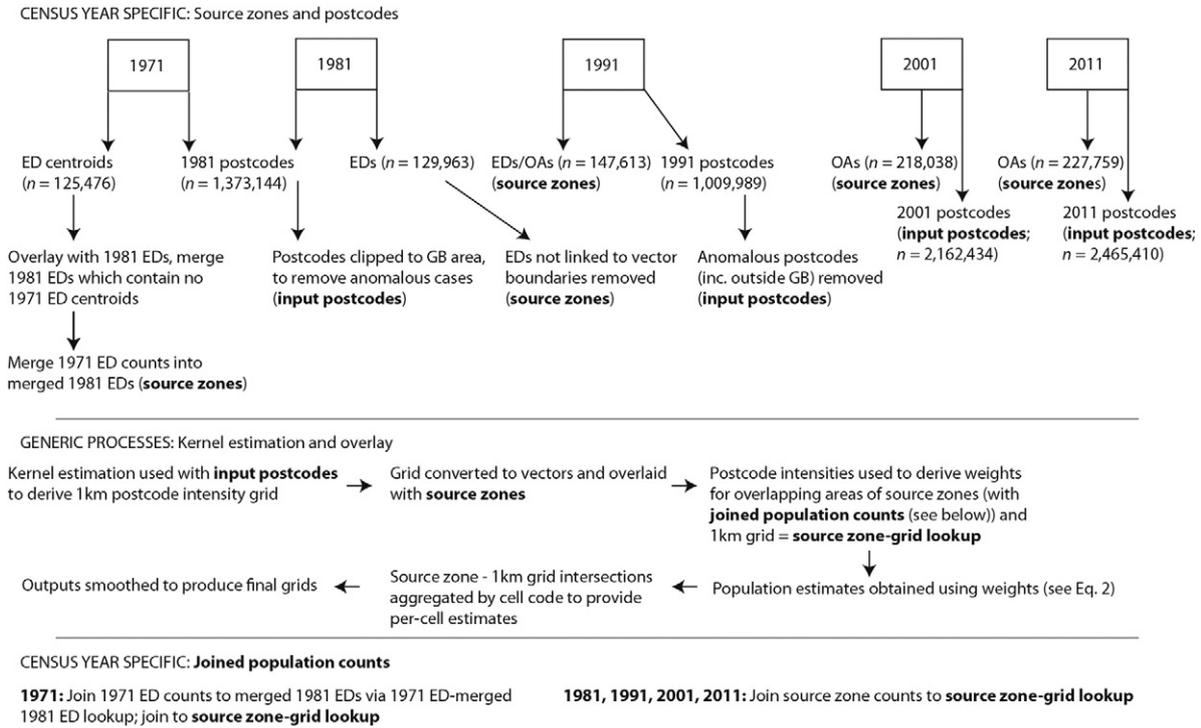CENSUS YEAR SPECIFIC: Source zones and postcodes

Fig. 1. Inputs and key steps in grid generation.

where $\hat{z}_t$ is the estimated population for the target zone t, A_{st} is the area of the zone of intersection between s and t and A_s is the area of source zone s. Overlay provides the basis of the first stage of the procedure used in this paper.

3.2. Areal weighting using postcode centroid intensities

While Eq. (1) is based solely on areas, ancillary data can be used to enhance the reallocation process by providing information on variations in population densities within source zones. In this study, postcode centroid intensity was used to determine weights to assign to overlapping areas of source (e.g., OAs) and target (1 km grid cells) areas. The estimation procedure applies these weights, λ, as follows (Gregory & Ell, 2005):

$$\hat{z}_{st} = \frac{\lambda_{j(c)}A_{st}}{\sum_{k=1}^{N}\lambda_{j(k)}A_{sk}} z_s \qquad (2)$$

where $\hat{z}_{st}$ is the estimated population for the zone of intersection between s and t and A_{st} is its area; $\lambda_{j(c)}$ is the weight for the specified control zone; $\lambda_{j(k)}$ is the weight for zone of intersection k and A_{sk} is the area of the zone of intersection k within source zone s and there are N zones of intersection.

The weights are determined using kernel intensity estimation (KE), which provides an estimate of the intensity of a set of points (here postcode centroids) for a regular grid. The intensity values can be used to determine what proportion of the population of a source zone (e.g., OA) should be reallocated to an overlapping grid cell. This was implemented in ArcGIS with a kernel function based on the quartic kernel function (Silverman, 1986, p. 76). The procedure was as follows:

1. Computed 1 km intensity grid using KE (for a 750 m, 1 km, 1.5 km, 2.5 km, or 3.5 km search radius)
2. Converted the raster grid derived from (1) into a vector grid (by overlay of raster cell centroids and joining these to the 1 km vector grid)
3. Overlaid vector grid with source zones (SAs/EDs/OAs)

4. Determine population of areas of intersection (SA/ED/OA and 1 km grid cell) using areal weighting following Eq. (2) (weights are postcode centroid intensities which are constant across 1 km cells)
5. Aggregate population estimates for areas of intersection by grid cell, giving population estimates for 1 km cells.

3.3. Smoothing gridded values

An additional step considered following stage 5 above was to smooth cell values using a 3 by 3 cell smoothing filter. Smoothing is used to allow values to vary between adjacent grid cells contained completely within the same source zone, while ensuring that the *sum* of the grouped cells remained the same so that the *total* population is unchanged. The amount of smoothing varies: spatially more variable population attributes, such as LLTI, will change more (proportionately) after smoothing than those which are more continuous (e.g., ethnicity) (see Lloyd, 2017). This is theoretically desirable as it prevents the counts of different variables drawn from the same source zone being distributed identically across the grid cells within the source zone, better reflecting the reality 'on the ground' within each source zone (see Lloyd, 2015 for a relevant discussion). Grid cells split across multiple source zones do not require smoothing as the counts for each variable already derive a distinct grid cell geography from their variation over the source zones contributing to the grid cell.

A point-based approach to smoothing was adopted to allow rescaling of the grid cell populations so that their sum matches the population of the source zone in which they are located. The alternative, also smoothing cells which overlap more than one source polygon, makes it problematic to ensure that people are not allocated to a cell outside of the relevant source zone, since there is no direct means of rescaling the population locally after smoothing.

One downside of the smoothing process adopted is that the non-split cells may end up with denominators which sum to a different amount than the sum of the numerators - e.g., for housing tenure: HH (households) owner occupied + HH private rented + HH social rented does not necessary equal HH total. Even so, we feel that this approach is conceptually superior to an approach which does not smooth grids on a

variable-by-variable basis, particularly bearing in mind that the total population or total households derived from different sets of counts will generally vary by only a very small amount. Taking one test as an example, for ethnic groups in 1991 the maximum difference between the summed numerators, using eight ethnic groups, and the total population for any 1 km cell was five people.

4. Testing models using existing grid data

Pre-existing gridded data for Northern Ireland (NI) were used as the basis for testing gridding methods. In this case, population grids were generated from small areas (SAs) using several variants of the surface modelling approach, and accuracy of the estimates was assessed by subtracting the estimates from the observed grids and exploring the resulting error grids. The approaches used were:

- Basic areal weighting: overlay of SAs and 1 km grid and estimate of grid population using proportion of SA falling in each 1 km cell (see Eq. (1))
- Postcode intensity weighting using kernel bandwidths of 750 m, 1.0, 1.5, 2.5 and 3.5 km (Eq. (2))
- As previous, for 1.0 km bandwidth (corresponding to smallest estimation errors for the bandwidths assessed), with the end result smoothed.

Testing comprised two stages, one using total household (HH) counts and one using counts of Catholics and persons with a LLTI (for the reasons detailed in Section 2.1). The first stage was used as HH counts were the only variable available at all locations for the 2011 NI grid square product. For other variables, counts were provided only for cells that contain at least 30 usual residents in 10 households. The co-efficient of determination for total persons versus HH was 0.99, indicating that household counts are a suitable basis for assessing generation of grids which includes counts of persons, as well as households. The testing focuses on the generation of grids for counts of Catholics and persons with a LLTI, but with reference to the results for total HH where appropriate.

Table 2 details the error summary statistics for each model and for both counts of Catholics and person with a LLTI. Note that the larger errors for Catholics than for persons with an LLTI are a function of the fact that numbers of Catholics are much greater than numbers of persons with a LLTI. The correlation coefficients suggest strong positive relationships between observed values and estimates in all cases. The root mean

Table 2
Estimation errors for Catholics by community background and LLTI for Northern Ireland. Errors are computed as estimated 1 km grid cell value minus observed 1 km cell value. 'PC' is postcodes; the number following PC is the kernel bandwidth (in km). The correlation coefficient, r, relates the observed values to the estimates.

	Method	Min	Mean	Max	RMSE	r
Catholics	Basic	−1162.24	2.84	580.06	60.37	0.975
Catholics	PC 0.75	−660.56	2.91	1293.70	45.87	0.985
Catholics	PC 1.0	−819.16	2.91	1131.32	44.63	0.986
Catholics	PC 1.5	−1016.73	2.85	865.98	49.65	0.982
Catholics	PC 2.5	−1170.09	2.85	679.73	55.27	0.978
Catholics	PC 3.5	−1226.40	2.85	710.78	58.46	0.976
Catholics	PC 1.0 + smooth	−819.16	2.91	1131.32	44.84	0.986
Catholics	PC 1.5 + smooth	−1016.73	2.90	865.98	50.28	0.982
Catholics	PC 2.5 + smooth	−1170.09	2.85	679.73	55.28	0.978
LLTI	Basic	−311.29	0.17	158.91	21.90	0.984
LLTI	PC 0.75	−293.78	0.18	369.90	15.73	0.990
LLTI	PC 1.0	−311.99	0.18	323.04	15.56	0.991
LLTI	PC 1.5	−325.47	0.18	277.32	17.49	0.988
LLTI	PC 2.5	−288.36	0.18	216.86	19.91	0.985
LLTI	PC 3.5	−298.45	0.18	173.75	21.20	0.984
LLTI	PC 1.0 + smooth	−311.99	0.18	323.04	15.59	0.991
LLTI	PC 1.5 + smooth	−325.47	0.18	277.32	17.69	0.988
LLTI	PC 2.5 + smooth	−288.36	0.18	216.86	19.92	0.985

square error (RMSE) is taken as a measure of magnitude; this suggests that the use of postcode intensities is beneficial given the much smaller RMSE values for the PC methods than for the basic (areal weighting) method. Further, the specific kernel bandwidth selected clearly has an impact on the accuracy of the results and the 1 km bandwidth corresponds to the smallest RMSE for both counts of Catholics and persons by LLTI.

The errors are smallest for a 1 km bandwidth for total HH, as well as for counts of Catholics and persons with a LLTI, and so this bandwidth was selected. With a smaller bandwidth, populations are concentrated close to postcodes. This is desirable where population densities are high and also where the postcode centroids represent well local population centres. Using 2011 data for Northern Ireland this condition is met in most locations. A bandwidth of 1 km correctly identified a much larger proportion of unpopulated (<0.5 households) cells than did other bandwidths (81.3% of a total 2561 unpopulated cells compared to 62.9% for a 1.5 km bandwidth). The ability to identify unpopulated cells is crucial in population grid generation and this further justifies the selection of the 1 km bandwidth. An alternative approach which adapts the bandwidth according to postcode centroid density could be assessed. One possible approach would be to use an urban/rural classification to determine kernel bandwidths; a problem with such an approach is that a comparable scheme would be needed for each Census year, but this is not available. Given the complexities of using data for multiple time periods and with different levels of quality this simpler approach was preferred. A refinement to the selected approach (1 km bandwidth estimation of postcode centroid intensities; PC 1 and results for 1.5 and 2.5 km are also shown) was also evaluated: smoothing neighbouring cells which fall entirely with input zones (as described above). Judging by Table 1, this has minimal impact on the errors, but the smoothing step was retained for generation of population grids for Britain on the grounds that smoothing is conceptually desirable given the different spatial structure of the counts, as analyses of their spatial autocorrelation indicates (see Lloyd, 2017). These observations hold both for counts of the population by religion and by limiting long term illness (LLTI). Smoothing is likely to have a bigger impact for 100 m grids as more variation and more cells (proportionately) will be smoothed given that they are more likely to be un-split by source zones than are 1 km cells.

The observed grids for Catholics and LLTI generated using PC 1 are shown in Figs. 2a and 3a respectively, while the corresponding errors (estimated values minus observed values) are shown in Figs. 2b and 3b. The observed grids show higher density populations in larger urban areas including Belfast (mid-east) and Derry/Londonderry in the north-west. Empty spaces are found in the more sparsely populated west of NI and in a large patch of the north east (containing the Antrim Plateau) and in the south east (Mourne Mountains). While both maps would be more informative if the counts were expressed as percentages, they are shown as counts for illustrative purposes, since these are the output from the grid generation process. (Of course, percentages can be generated from the derived grids of counts.)

There are no systematic trends in larger errors, although as expected the errors are largest in urban areas since the population densities are higher (see Figs. 2b and 3b). There is little difference in the spatial distribution of errors between unsmoothed estimates (not shown) and smoothed estimates. For both Catholics and LLTI, there are multiple cases of cells with large positive errors adjacent to cells with large negative errors. These reflect cases of locally large population densities; an obvious example is a tower block. While postcodes help to define internal variations in source zones they do not precisely locate very high density population areas within source zones. As a result, the cell containing the tower block would likely have an under-estimated count, while adjacent cells would contain an over-estimated count. Such a situation could only be avoided by having spatially-detailed information on the precise location of such areas. The results suggest that the modelling procedure provides fairly accurate estimates in most locations
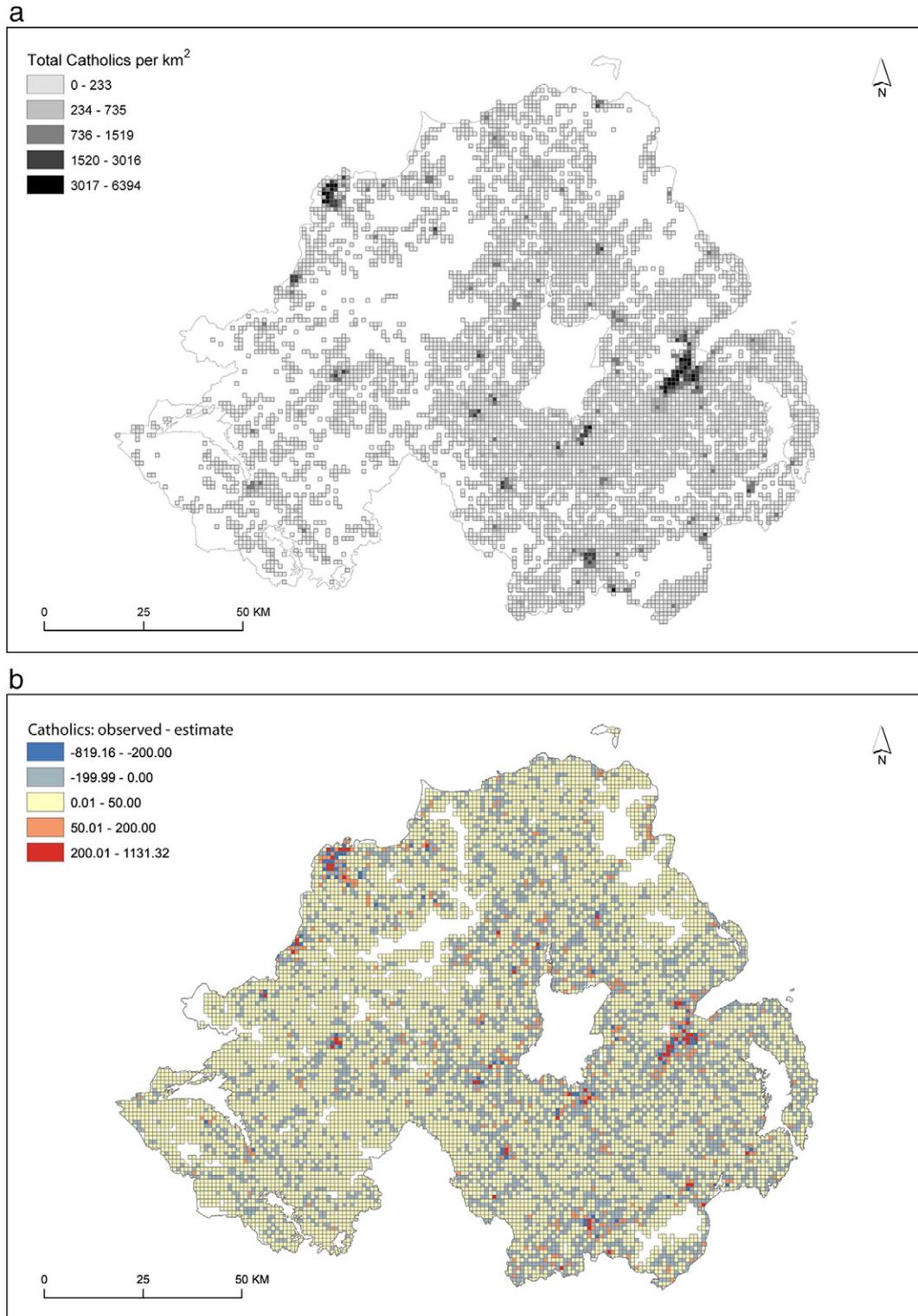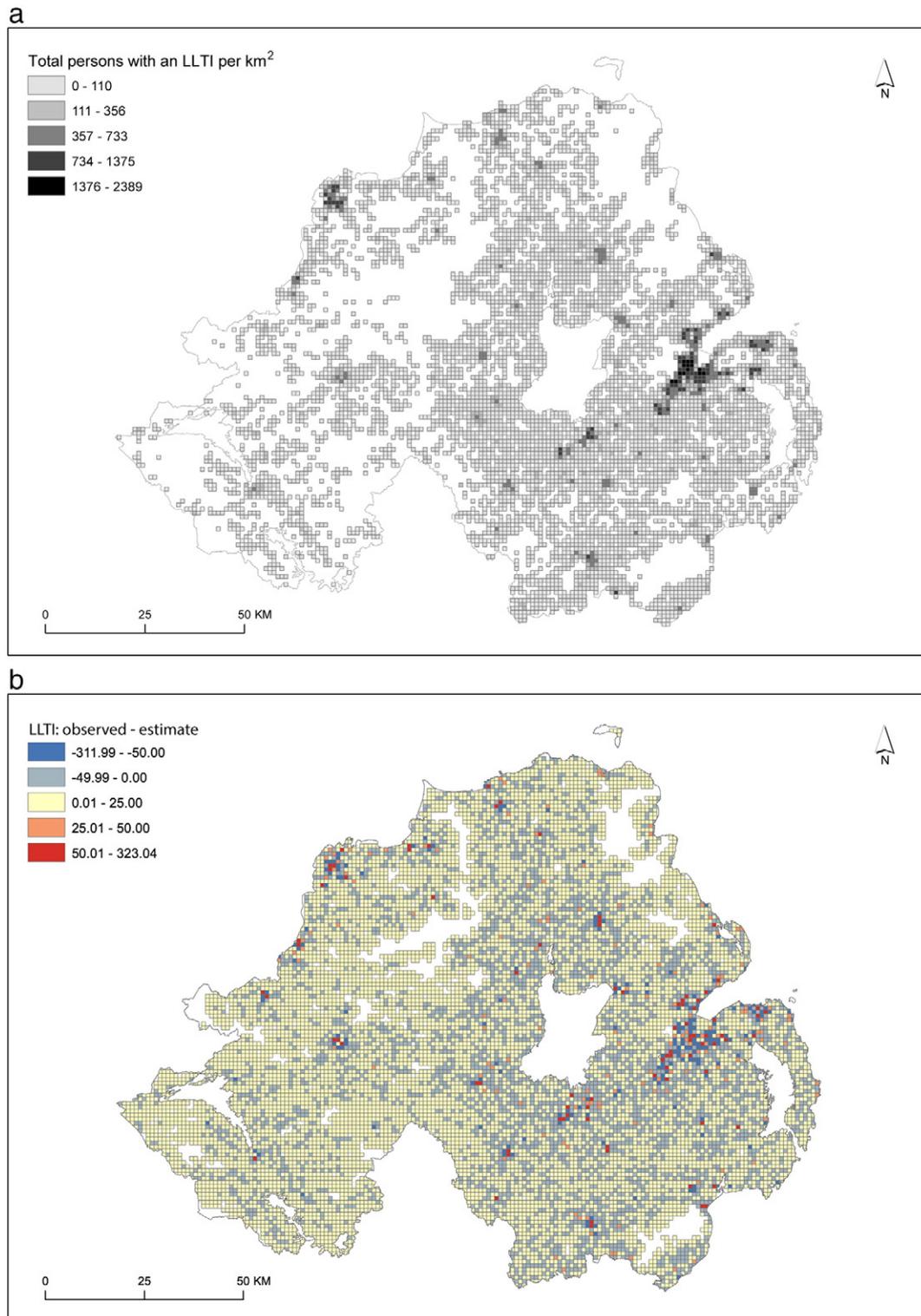
a

b

Fig. 2. (a) Catholics grid for NI for 2011, (b) Catholics estimate-observed grid population.
Source: 2011 Census: Northern Ireland Grid Square Data.

(small negative or positive errors), with larger errors in some urban areas. In these later cases, larger errors tend to be localised and thus, in effect, larger errors are only at the 2 km scale. In other words, a grid with a coarser resolution would contain relatively smaller errors. In summary, the approach was considered suitable as a means for transferring population counts from irregular source zones to 1 km grids. However, the nature of the grids for Britain (detailed below) as *estimates* should be recalled in any study which makes use of them.

5. Population change in small areas of Britain

The approach developed using NI grid data as a test case was applied to small area (ED or OA-level) data for 1971, 1981, 1991, 2001 and 2011 for counts of total persons for Britain. Estimated total persons for a 1 km grid for 2011 for Britain (with an insert for London) are shown in Fig. 4a, while the differences between these figures and their equivalents for 1971 (2011–1971) are in Fig. 4b. The distribution of the population is

a

Total persons with an LLTI per km^2

- 0 – 110
- 111 – 356
- 357 – 733
- 734 – 1375
- 1376 – 2389

0   25   50 KM

b

LLTI: observed - estimate

- -311.99 - -50.00
- -49.99 - 0.00
- 0.01 - 25.00
- 25.01 - 50.00
- 50.01 - 323.04

0   25   50 KM

Fig. 3. (a) LLTI grid for NI for 2011, (b) LLTI estimate-observed grid population.
Source: 2011 Census: Northern Ireland Grid Square Data.

strongly positively skewed and the choice of colour mapping has a major impact on the contrasts between values in the maps. After experimentation, a linear stretch between the minimum and maximum was used for Fig. 4a while Fig. 4b was produced using a linear stretch based on 2.5 standard deviations.[4] Estimates are shown

only for cells which are estimated to contain people (in practice, 0.5 persons or above, noting that fractions of people are possible using this approach). Empty areas accord with expectation; for example, with large unpopulated areas in the Highlands of Scotland. The map of differences (Fig. 4b) contains a considerable amount of information; it shows population decreases in many urban areas including Glasgow, Newcastle, Manchester, Liverpool, Birmingham and central London, with increases in the outskirts of
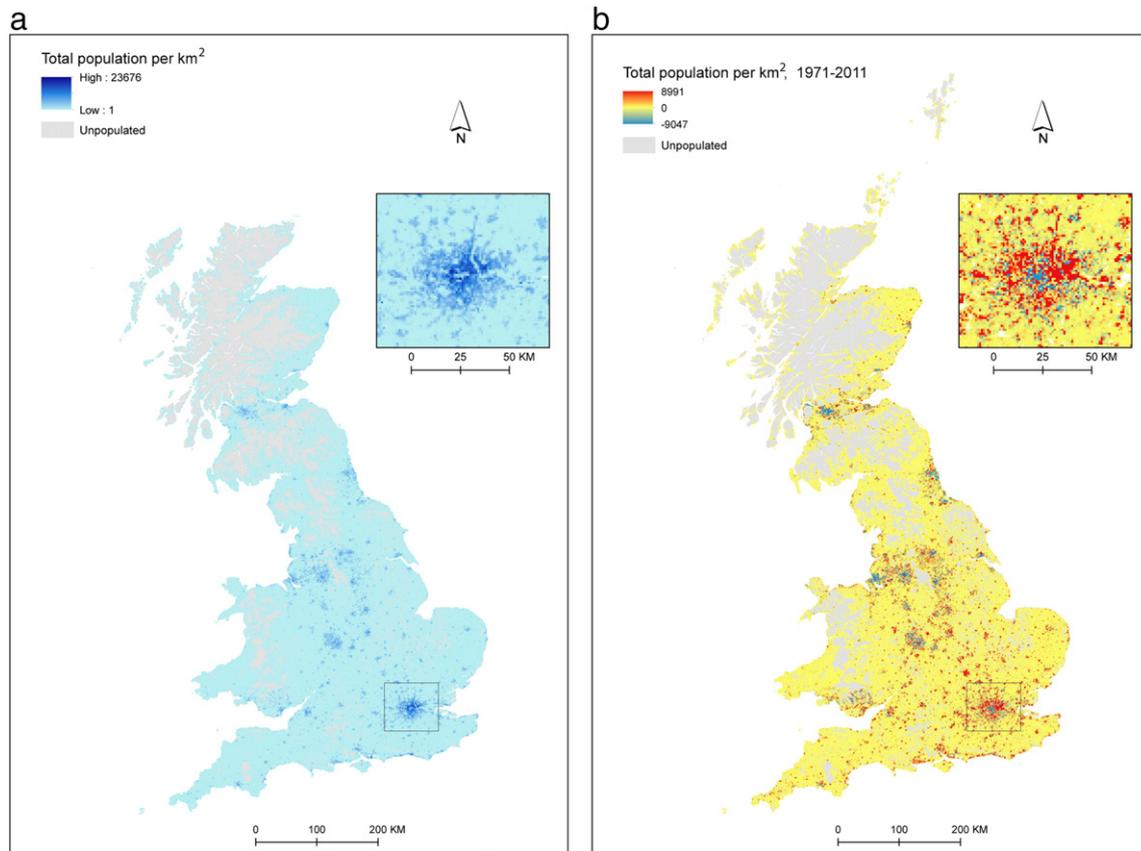
Fig. 4. (a) Total persons in 2011, (b) Total persons in 2011–1971, with London area insert. Threshold is >0.5 persons.

London and other areas across, most notably, the south east of England.

A key benefit of grids is that zones are not derived using some prior condition, such as population density in one Census year, which may not be appropriate for another Census year. For example, if an area is sparsely populated and later becomes a site of urban expansion then zones constructed using data for the earlier period will not be suitable for later ones. For this reason the growth in population – e.g. in the outskirts of London – is better represented using fine-grain, density independent grids than it is through the use of irregular density-dependent (and potentially temporally-variable) zones.

For the same reason, grids are also well-suited to measuring changes in population density across the country as a whole. Table 3 shows the percentage of total persons who lived in 1 km by 1 km cells with above the specified threshold population. These figures provide evidence for counter-urbanisation from areas with higher population densities between 1971 and 1981, followed by a gradual increase in urbanisation from 1981 to 2001, and a large increase from 2001 to 2011. The share of the population in moderate density areas (i.e., >1000 to ≤2500 persons per square km) increased decennially from 1971 to 2011. Over time, the correlation between the population and

population change reversed. For cells with ≥25 persons in all Census years the correlation between population count in a given Census year (e.g., 1971) and the population change from that Census year to the next (e.g. 1971 to 1981) is as follows: 1971: −0.292; 1981: −0.171; 1991: 0. 064; 2001: 0.429. A map of population change between 2001 and 2011 (not shown) indicates that there were large increases in population in most urban areas, in contrast to the large declines evidenced in many urban areas between 1971 and 2011 (Fig. 4b; showing 2011 counts minus 1971 counts). This correlation disappears when the population change is expressed as a percentage of the base year (e.g. 1971–81 change as a % of 1971 population): 1971: −0.043; 1981: −0.104; 1991: −0.078: 2001: −0.021. Collectively, these results reflect the observations of Champion (1989) for the period 1971 to 1981, and the longer-term trends observed by Champion (2008) with respect to England. However, grids allow a local scale, more nuanced analysis, of population change than is provided by, for example, district-level analyses.

As well as permitting the analysis of population change using spatially- and temporally-consistent geography, grid cells also support the use of spatial filters and other raster analysis tools. Spatial filters (see Sonka et al., 2015) are commonly used to smooth (reduce contrast in) or sharpen (increase contrast in) images. Filters can be used to, for example, highlight areas of pronounced change, or as a basis for segmentation of areas with similar characteristics. Space prohibits extensive exploration of the application of spatial filters here. Instead, for those unfamiliar with spatial filters we provide one simple example to illustrate the general point. Fig. 5 is a map of local standard deviations computed using 3 by 3 cells to illustrate the application of a standard image processing tool; this is simply the standard deviation computed for a cell and its immediate neighbours. The local standard deviation picks out the edges of urban areas; comparison with Fig. 4a shows that it is effective in highlighting small urban areas which are not clearly apparent in a map of total persons. An extension to such an approach would be to

Table 3
Percentage of total persons who lived in 1 km by 1 km cells with a population above the threshold.

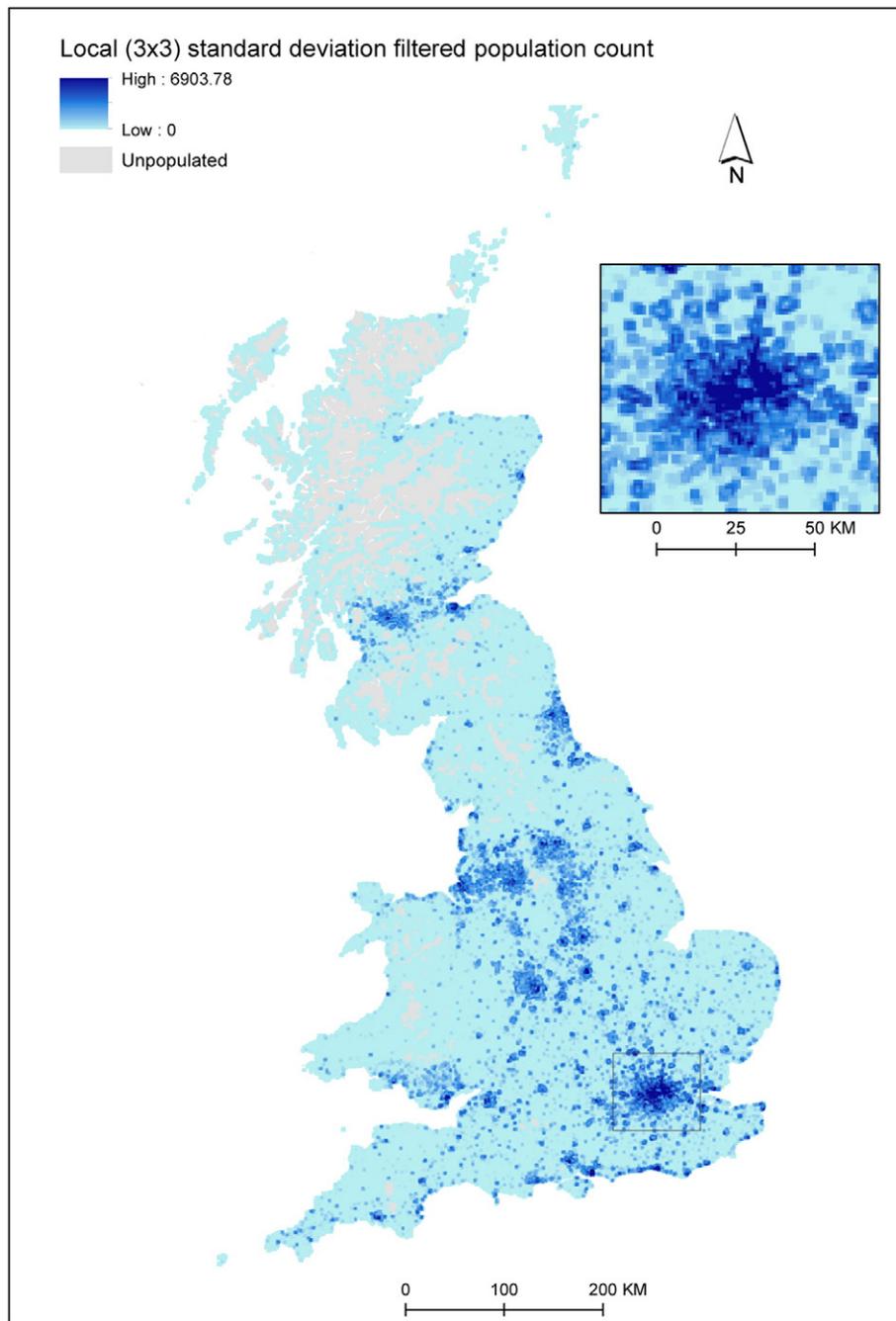| Threshold	1971	1981	1991	2001	2011
20,000	0.12	0.00	0.00	0.03	0.10
15,000	1.13	0.12	0.18	0.43	1.44
10,000	5.93	2.71	2.68	3.63	5.56
5000	28.32	25.86	23.46	22.79	26.99
2500	56.97	60.69	60.37	60.53	62.70
1000	76.11	80.49	80.83	81.59	82.72
100	94.21	94.73	95.73	95.36	95.65

Fig. 5. Standard deviation (3 by 3 cell window): Total persons in 2011.

apply it in creating a typology of population change. In addition, alternative local statistics could be computed to capture information on the nature of areas; these could include maxima (maximum cell value in a neighbourhood) or an alternative scheme for picking out edges such as the Laplace operator.

6. Discussion and conclusions

Testing using data for NI suggested that, not surprisingly, the use of postcodes resulted in much more accurate estimates than were obtained using simple spatial overlay. The kernel bandwidth used proved to be sensitive to population density and the quality of the postcode data available. Overall a 1 km bandwidth was preferred. Smoothing made no difference to the accuracy of estimates, but, we have argued, should

be utilised as variables are likely to have differing spatial structures within the same source zone (see, for example, Lloyd, 2016).

The provisional analyses of grids for Britain for 1971 and 2011 demonstrate the value of being able to explore estimated population change over small areas. The difference maps suggest that there were large decreases in total population (2011–1971) in some urban areas, with increases in parts of London and other areas away from the larger cities and towns. The total population grids for each Census year are available via the *PopChange* web resource (https://popchange.liverpool.ac.uk/). Fig. 6 is a screen dump of the *PopChange* raster calculation tool showing the difference between the total population in 2011 and in 2001. The *PopChange* resource enables, for the first time, analysis of small area change in Britain (and the UK as a whole with integration of grid square counts for Northern Ireland) over a forty year period. It will be interesting to see how changes identified using the grid data compare with

Fig. 6. PopChange web resource (https://popchange.liverpool.ac.uk/).
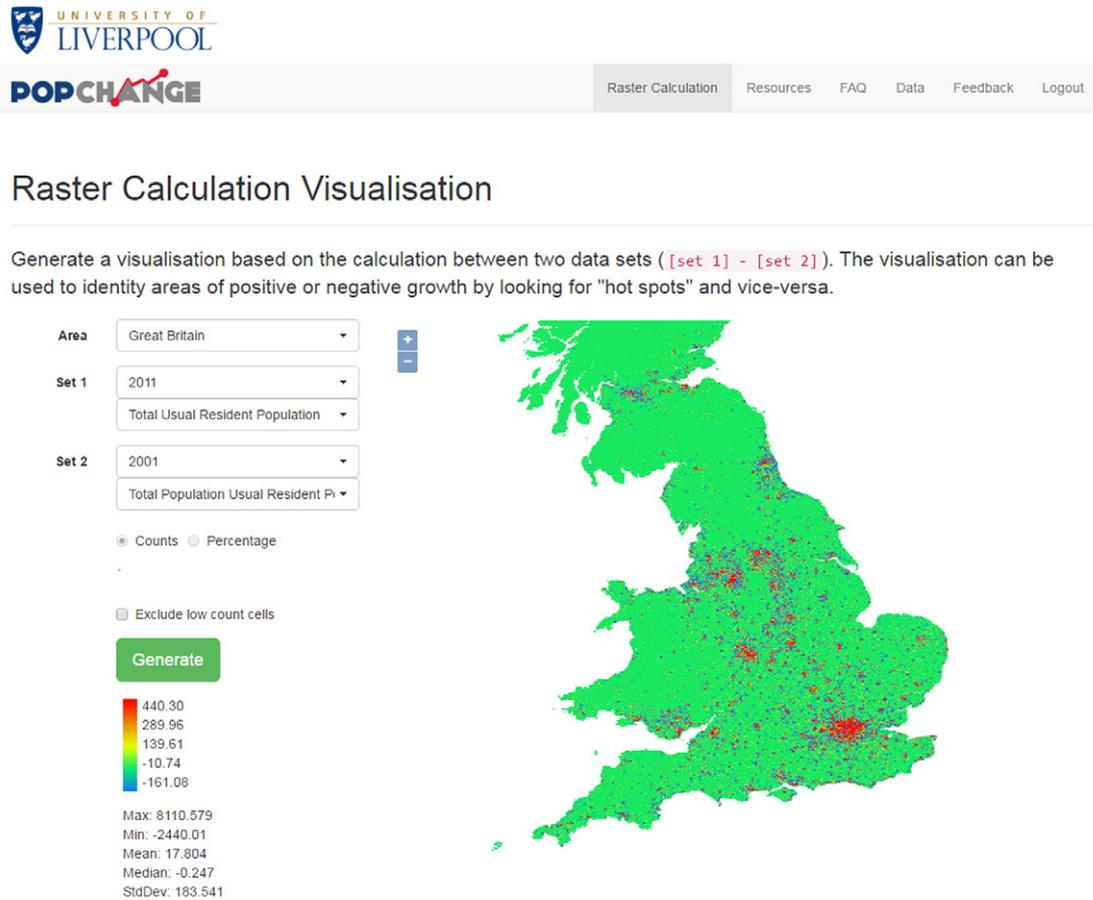
changes for Britain during the same period found using small area vector geographies (Norman, 2016). All data generated are freely available, as are tools for undertaking basic analyses and guidance on use of the resource. The resource will enable geographically- and attribute-rich analyses of population change in the UK and, specifically, the ways in which the population has become more or less geographically unequal.

Developments to the population grid resource will be detailed on the project website. This paper has focused on the use of data for 1 km grid squares. Ongoing work is developing 100 m grid cells (using a slightly different procedure) and these grids will also be made freely available in due course; these will enable even more localised patterns to be explored. There are several other obvious ways in which the *PopChange* resource could be enhanced. More information will be provided on definitions of questions and output categories across a wider range of variables. Additional sources of ancillary information could be used to generate more accurate gridded population estimates. While a consistent approach which was applicable across all Census years was preferred here, 'optimal' grids could be provided where additional ancillary information is available. For recent Census years (especially 2001 and 2011), detailed land use data, or building outlines as well as postcode area boundaries and address or population-weighted postcodes, could be used to increase the quality of estimates. The OpenPopGrid initiative (Murdock et al., 2015) has produced 10 m population grids for 2011, whereby Ordnance Survey Vector Map District building polygons are used to redistribute counts via dasymetric mapping. PopChange could be enhanced using these data, although equivalent data are not available for earlier Censuses. The NI-based study provided information on uncertainty in estimates, and such information could be used to develop formal models of uncertainty for the estimates so as to better inform users of the limitations of the estimates. Also,

more work could be undertaken on the determinants of estimation errors (building on the work of Lloyd & Firoozi Nejad, 2014). In addition, it would be worthwhile to undertake comparative work on the relative benefits of grid-based estimates as compared to estimates made for sets of consistent irregular zones. Incorporating other data sources such as data on births and mortality as well as on land use change would increase further the utility of the resource. The methods developed in the present study could be applied in other national contexts. Generating population grids across multiple countries would enable meaningful cross-country comparisons for core themes such as urban-rural migration or the geographies of inequalities.

While the resource includes data up to 2011, the most recent UK Census year, software tools are provided to allow users to 'grid' their own data, thereby allowing linkage to other sources of information about the population post-2011, as well as estimates made between Census years. Knowing about population change in local areas, and not just who lives in areas now, represents new opportunities to explore significant unanswered research questions. How has the global recession of 2008 impacted on inequalities across the country? With links to data post-2011, the possible effects of austerity could also be assessed. Political affiliations too are obviously partly a function of area population histories. The *PopChange* resource offers a means of beginning to deconstruct the geography of Brexit by allowing us to consider the characteristics of 'Leave' and 'Remain' areas, and how such areas differ in terms of their population histories. Are there differences between areas with persistently high levels of deprivation, compared to areas which have undergone considerable recent changes? How far do changes in exposure to diversity in residential areas (as measured by country of birth or ethnicity) relate to political affiliation, or levels of social trust? Such context is vital to understanding how areas and their

populations evolve, and there is much that can be learned by combining diverse data sources for multiple time points, as the *PopChange* resource is beginning to show.

Much is written about 'Big data' and the possibilities that exciting new datasets offer for addressing major societal questions. Yet much more can be done to chart the social, demographic and economic trajectories of small areas using data which were collected a generation ago and more. Beyond its academic potential, the resource offers considerable opportunities for influencing public discourse and informing policy action. *PopChange* engages public users; members of the public can explore their local neighbourhoods: where has the population has grown or shrunk in the last 40 years? What has happened to neighbourhood unemployment levels? It will help researchers to explore the relationships between health status and long-term deprivation, and policy-makers to assess the success of local intervention strategies, as well as to suggest new ways of targeting resources. The resource brings a new and important perspective to debates about divisions, inequalities and the ways in which people in the UK live together or apart.

Acknowledgements

References

Batista e Silva, F., Gallego, J., & Lavalle, C. (2013). A high-resolution population grid map for Europe. *Journal of Maps, 9*, 16–28.

Broughton, N. (2016). Devolution and poverty. SMF Briefing: Can we Tackle Poverty at a Local Level? London: Social Market Foundation. http://www.smf.co.uk/wp-content/uploads/2016/06/Social-Market-Foundation-Publication-Devolution-and-Poverty-Barrow-Cadbury-FINAL.pdf.

Census Research Unit/Office of Population Censuses and Surveys/General Register Office (Scotland) (1980). *People in Britain: A census atlas.* London: HMSO.

Champion, A. G. (1989). Counterurbanization in Britain. *The Geographical Journal, 155*, 52–59.

Champion, T. (2008). Population change in England since 1981: Is an 'urban renaissance' really underway? *Géocarrefour, 83*, 79–86.

Deng, Y., Frantz, A., & Araoz, A. (2017). Population surface generation: Separating urban and rural. *Cartographica: The International Journal for Geographic Information and Geovisualization, 52*, 20–31.

Denham, C., & Rhind, D. (1983). The 1981 census and its results. In D. Rhind (Ed.), *A census user's handbook* (pp. 17–88). London: Methuen.

Dorling, D., & Simpson, S. (1993). Those missing millions: implications for social statistics of undercount in the 1991 census. *Radical Statistics, 55*, 14–35.

Dorling, D., Martin, D., & Mitchell, R. (2001). Linking Censuses Through Time – Final Report. Easy Access to 2001 comparable 1991, 1981 and 1971 data. Economic and Social Research Council (ESRC) Award H507255151. http://www.researchcatalogue.esrc.ac.uk/grants/H507255151/read, Accessed date: 26 July 2016.

Duke-Williams, O., & Rees, P. (2002). Testing user-requested geographies. In P. Rees, D. Martin, & P. Williamson (Eds.), *The census data system* (pp. 339–353). Chichester: John Wiley & Sons.

Flowerdew, R., & Green, M. (1994). Areal interpolation and types of data. In S. Fotheringham, & P. Rogerson (Eds.), *Spatial analysis and GIS* (pp. 121–145). London: Taylor and Francis.

Gallego, J. (2010). A population density grid of the European Union. *Population and Environment, 31*, 460–473.

Goovaerts, P. (2008). Kriging and semivariogram deconvolution in the presence of irregular geographical units. *Mathematical Geosciences, 40*, 101–128.

Gregory, I., & Ell, P. (2005). Breaking the boundaries: Geographical approaches to integrating 200 years of the census. *Journal of the Royal Statistical Society, Series A, 168*, 419–437.

Kim, H., & Yao, X. (2010). Pycnophylactic interpolation revisited: Integration with the dasymetric-mapping method. *International Journal of Remote Sensing, 31*, 5657–5671.

Kyriakidis, P. C. (2004). A geostatistical framework for area-to-point spatial interpolation. *Geographical Analysis, 36*, 259–289.

Lloyd, C. D. (2010). Exploring population spatial concentrations in Northern Ireland by community background and other characteristics: An application of geographically weighted spatial statistics. *International Journal of Geographical Information Science, 24*, 1193–1221.

Lloyd, C. D. (2014). *Exploring spatial scale in geography.* Chichester: John Wiley and Sons.

Lloyd, C. D. (2015). Assessing the spatial structure of population variables in England and Wales. *Transactions of the Institute of British Geographers, 40*, 28–43.

Lloyd, C. D. (2016). Spatial scale and small area population statistics for England and Wales. *International Journal of Geographical Information Science, 30*, 1187–1206.

Lloyd, C. D. (2017). Creating population surfaces for the analysis of small area change. In D. Swanson (Ed.), *The Frontiers of applied demography* (pp. 431–448). Dordrecht: Springer.

Lloyd, C. D., Bearman, N., Catney, G., & Williamson, P. (2017). *Mapping population change in Britain using the PopChange resource. CSDR working paper 1*Centre for Spatial Demographics Research, University of Liverpool.

Lloyd, C. D., & Firoozi Nejad, B. (2014). Surface models and the spatial structure of population variables: Exploring smoothing effects using Northern Ireland grid square data. *Computers, Environment and Urban Systems, 48*, 64–72.

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

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

Martin, D., Dorling, D., & Mitchell, R. (2002). Linking censuses through time: Problems and solutions. *Area, 34*, 82–91.

Martin, D., Lloyd, C., & Shuttleworth, I. (2011). Evaluation of gridded population models using 2001 Northern Ireland census data. *Environment and Planning A, 43*(8), 1965–1980.

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

Murdock, A. P., Harfoot, A. J. P., Martin, D., Cockings, S. and Hill, C. (2015). OpenPopGrid: An open gridded population dataset for England and Wales. Southampton: GeoData,

University of Southampton. http://openpopgrid.geodata.soton.ac.uk/OpenPopGrid_ProductDocumentation.pdf Accessed 26 July 2016.

Norman, P. (2016). The changing geography of deprivation in Britain: 1971 to 2011 and beyond. In T. Champion, & J. Falkingham (Eds.), *Population change in the United Kingdom* (pp. 193–214). London: Rowman and Littlefield.

Norman, P. D., Rees, P. H., & Boyle, P. J. (2003). Achieving data compatibility over space and time: Creating consistent geographical zones. *International Journal of Population Geography, 9*, 365–386.

ONS (2004). *Census 2001 definitions.* London: The Stationery Office.

ONS (2009). Final population definitions for the 2011 census. ONS: Titchfield. http://www.ons.gov.uk/ons/guide-method/census/2011/the-2011-census/2011-census-questionnaire-content/final-population-definitions-for-the-2011-census.pdf, Accessed date: 26 July 2016.

Openshaw, S. (1984). The modifiable areal unit problem. *Concepts and techniques in modern geography 38.* Norwich: GeoBooks.

Openshaw, S., & Taylor, P. J. (1979). A million or so correlation coefficients: Three experiments on the modifiable areal unit problem. In N. Wrigley (Ed.), *Statistical applications in the spatial sciences* (pp. 127–144). London: Pion.

R Core Team, R. (2016). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. https://www.R-project.org, Accessed date: 8 May 2017.

Rees, P., Parsons, J., & Norman, P. (2005). Making an estimate of the number of people and households for output areas in the 2001 census. *Population Trends, 122*, 27–34.

Reis, S., Steinle, S., Carnell, E., Leaver, D., Vieno, M., Beck, R. and Dragosits, U. (2016). UK gridded population based on census 2011 and land cover map 2007. NERC

Environmental Information Data Centre. http://dx.doi.org/10.5285/61f10c74-8c2c-4637-a274-5fa9b2e5ce44 Accessed 26 July 2016.

Shuttleworth, I. G., & Lloyd, C. D. (2009). Are Northern Ireland's communities dividing? Evidence from geographically consistent census of population data, 1971–2001. *Environment and Planning A, 41*, 213–229.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis.* London: Chapman and Hall.

Sonka, M., Hlavac, V., & Boyle, P. (2015). *Image processing, analysis and machine vision* (Fourth edition). Stanford: Cengage Learning.

Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PloS One, 10*(2), e0107042.

Stillwell, J., Norman, P., Thomas, C., & Surridge, P. (2010). Spatial and social disparities. In J. Stillwell, P. Norman, C. Thomas, & P. Surridge (Eds.), *Spatial and social disparities: Understanding population trends and processes. 2.* (pp. 1–15). London: Springer.

Tobler, W. R. (1979). Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association, 74*, 519–530.

Walford, N. S. (2002). *Geographical data: Characteristics and sources.* Chichester: Wiley.

Walford, N. S., & Hayles, K. N. (2012). Thirty years of geographical (in)consistency in the British population census: Steps towards harmonisation of small-area census geography. *Population, Space and Place, 18*, 295–313.

Wong, D. (2009). The modifiable areal unit problem (MAUP). In A. S. Fotheringham, & P. A. Rogerson (Eds.), *The SAGE handbook of spatial analysis* (pp. 105–123). London: SAGE Publications.