Transactions in GIS

Research Article Transactions in GIS, 2013, 17(4): 563–588

Temporal Uncertainty in a Small Area Open Geodemographic Classification

Christopher G. Gale and Paul A. Longley

Department of Geography, University College London

Abstract

The 2001 Output Area Classification (2001 OAC) is an open source geodemographic classification of the UK built exclusively from 2001 UK Census data. There has been considerable user interest in its applicability to subsequent time periods, particularly given the potential propensity of characteristics and attributes in some areas to change during inter-censual periods. Users often purchase commercial geodemographic classification products in the belief that purely census-based classifications such as the 2001 OAC are uniformly unreliable because there is no temporal updating of input data. Yet there is evidence to suggest that whilst some UK neighborhoods are prone to sudden changes, many others change very little over protracted time periods. Using measures that are available at the small area level, temporal uncertainty indicators can be constructed to identify those areas that are less stable. Using mid-year population estimates and dwelling stock data, this article develops three temporal uncertainty indicators. These provide a reliable means of gauging the stability or otherwise of neighborhood conditions. The conclusion from this is that while a large number of small areas in the UK do experience change over time, this change is not uniform in either degree or distribution, or by geodemographic type.

1 Introduction

Geodemographic classifications are small area measures that provide summary indicators of the social, economic and demographic characteristics of neighborhoods (Adnan et al. 2010). The aggregations of individual and household characteristics and neighborhood attributes have been widely used for resource planning and allocation in both the commercial and public sectors (Shelton et al. 2006). The intellectual heritage of geodemographics can be traced back to the work on urban studies by the Chicago School. These conceptual beginnings of urban ecology and social area analysis provide a framework for social measurement to be empirically undertaken to better understand neighborhood characteristics (see, for example, Longley 2005). The desire to generalize urban social patterning in the 1970s led Richard Webber to develop a branch of applied urban studies that would later be termed “geodemographics”. While initially devised for use in the public sector, successful geodemographic applications developed most rapidly in the private sector, using proprietary solutions such as CACI’s (London, UK) Acorn and Experian’s (Nottingham, UK) Mosaic. Despite their cost, commercial products such as these dominate the UK market and are widely utilized across multiple industries, in significant part because ancillary sources are used to enrich and update the classifications – many users equate “frequently updated” with “best”, despite the fact that the provenance of some of the ancillary sources is unknown (see CACI 2009, Experian 2010). The provision of incremental updates over time is much vaunted in the marketing of commercial classifications.

Address for correspondence: Christopher G. Gale, Department of Geography, University College London, Gower Street, London, WC1E 6BT, UK. E-mail: c.gale.10@ucl.ac.uk

 doi: 10.1111/tgis.12035

Table 1 The lowest two hierarchical levels of 2001 UK Census Geography (Data sources: ONS, NRS and NISRA)

	Hierarchy Level 1	Hierarchy Level 2
England and Wales	Output Areas: Total number – 175,434 Average resident population – 297	Lower Layer Super Output Areas: Total number – 34,378 Average resident population – 1,500
Scotland	Output Areas: Total number – 42,604 Average resident population – 336	Data Zones: Total number – 6,505 Average resident population – 800
Northern Ireland	Output Areas: Total number – 5,022 Average resident population – 119	Super Output Areas: Total number – 890 Average resident population – 1,900

An alternative to commercial classifications is the 2001 Output Area Classification (2001 OAC), created for the Office for National Statistics (ONS) by Vickers and Rees (2007). This classification is freely available in the public domain and is the outcome of a well-documented and easily replicable methodology. It however lacks any inter-censual updates and relies solely on data from the 2001 UK Census. The difference between the public and commercial geodemographic products is clear; the 2001 OAC is freely available, but of limited relevance in areas of the country that have changed rapidly over the last decade; while commercial products have been "freshened up" using a range of sources that are either of unknown provenance or made available at coarser levels of granularity than the small neighborhood areas for which the classifications were devised.

This article seeks to identify an alternative to the temporal updating methods that are used in commercial classifications by utilizing small area measures of change to indicate how reliable or otherwise the 2001 OAC geodemographic assignments have become. The 2001 OAC is grounded at the scale of the Output Area (OA) – on average housing 264 individuals and 110 households (Vickers et al. 2005). Table 1 provides an explanation of 2001 UK Census Geography used in the classification. The 2001 OAC assigns each OA to one of seven Supergroups, 21 Groups and 52 Subgroups in the three-tier hierarchical classification shown in Table 2 (Vickers et al. 2005). While the 2001 OAC provides a benchmark of neighborhood conditions as of the date of the 2001 UK Census, 29 April 2001, there has been considerable (actual and potential) user interest in its applicability to subsequent time periods, although this requires understanding of the distinctive nature of small area change across the UK. A recent survey of user requirements for such classifications carried out by the ONS (ONS 2012a) has suggested that users gravitate towards commercial geodemographic classification products because of the perception that the 2001 OAC is unreliable in the absence of updating or temporal change measures.

At the time of writing (August 2012), the imminent availability in the UK of small area statistics from the 2011 UK Census provides the opportunity to build a new OAC. But this alone will not address the inherent issue of the spatially variable degradation in reliability over time, or counter commercial systems' edge in the coming years. In fact, much of urban theory (e.g. Hoyt 1939) posits that the characteristics of most neighborhoods do not change rapidly, and empirical studies substantiate this view (see Longley et al. 2011). Users nonetheless state that they would like annual updates to a new OAC (ONS 2012a).

Table 2 The 2001 Output Area Classification Hierarchy (Vickers et al. 2005)

Supergroups	Groups	Subgroups
1 Blue Collar Communities	1a Terraced Blue Collar	1a1
		1a2
		1a3
	1b Young Blue Collar	1b1
		1b2
	1c Older Blue Collar	1c1
		1c2
		1c3
2 City Living	2a Transient Communities	2a1
		2a2
	2b Settled in the City	2b1
		2b2
3 Countryside	3a Village Life	3a1
		3a2
	3b Agricultural	3b1
		3b2
	3c Accessible Countryside	3c1
		3c2
4 Prospering Suburbs	4a Prospering Younger Families	4a1
		4a2
	4b Prospering Older Families	4b1
		4b2
		4b3
		4b4
	4c Prospering Semis	4c1
		4c2
		4c3
	4d Thriving Suburbs	4d1
		4d2
5 Constrained by Circumstances	5a Senior Communities	5a1
		5a2
	5b Older Workers	5b1
		5b2
		5b3
		5b4
	5c Public Housing	5c1
		5c2
		5c3
6 Typical Traits	6a Settled Households	6a1
		6a2
	6b Least Divergent	6b1
		6b2
		6b3
	6c Young Families in Terraced Homes	6c1
		6c2
	6d Aspiring Households	6d1
		6d2
7 Multicultural	7a Asian Communities	7a1
		7a2
		7a3
	7b African-Caribbean Communities	7b1
		7b2

The increasing proliferation of government open data sources appears to offer some solutions – as of August 2012 the government data.gov.uk website offers over 8,500 datasets (see http://www.data.gov.uk). Yet despite progress in improving the availability and dissemination of open data, very few government datasets are currently available at the OA level. Additional complications arise out of the different data dissemination conventions in England and Wales, in Scotland and in Northern Ireland – something that hinders the creation of a UK-wide Index of Multiple Deprivation, for example. Recourse to coarse grained open data potentially allows regular updating of open classifications, although at the expense of representing local detail. This is far from ideal in many applications of geodemographic classifications. The alternative, explored here, is to use only the small number of measures obtainable at the OA level to construct temporal uncertainty indicators. These can then be compared at national and regional scales (see Figure 1). Our motivation is to identify areas in which significant change in demo-

Figure 1 UK study area locations
Contains Ordnance Survey Data © Crown Copyright and Database Right 2011.

Table 3 Population change (Data sources: ONS, NRS and NISRA)

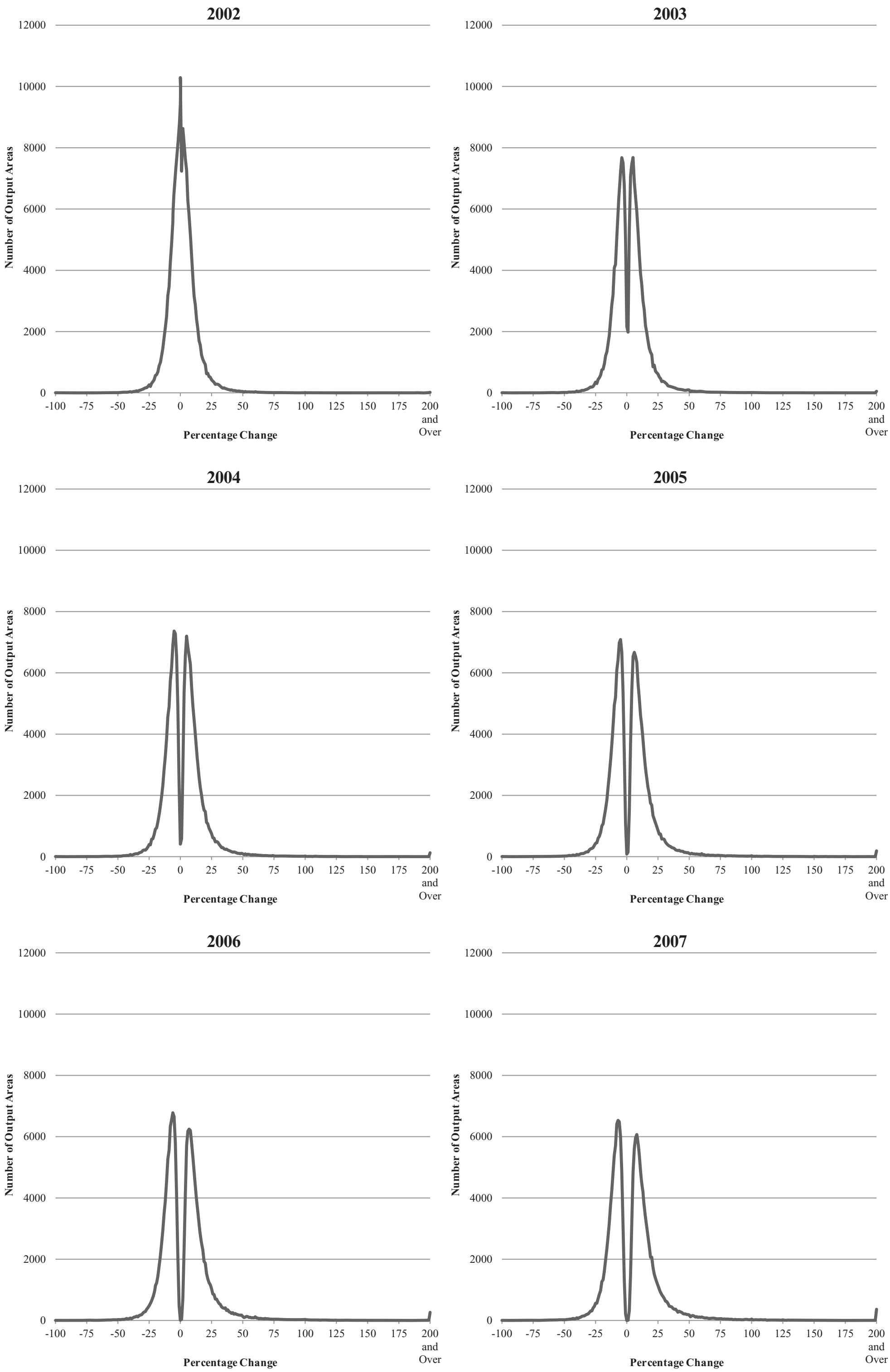
Country	2001 Total	Proportion of total UK population	2010 Total	Proportion of total UK population	2001 to 2010 Change
England and Wales	52,359,975	88.58%	55,240,475	88.72%	+5.50%
Scotland	5,062,011	8.56%	5,222,014	8.39%	+3.16%
Northern Ireland	1,685,267	2.85%	1,799,398	2.89%	+6.77%
United Kingdom	**59,113,485**	**100%**	**62,261,887**	**100%**	**+5.33%**

graphic compositions has occurred at the small area level. These are not designed to update the classification, but rather provide a first step to indicate where updating is likely to be necessary. Any area deemed to have experienced significant change over time is flagged as being potentially uncertain and its geodemographic class assignment unreliable. This notion of temporal uncertainty has not hitherto been explored in any detail within the wider discourse of geodemographics: rather, the main focus of uncertainty within geodemographics has been at the initial stage of creation of a classification, specifically with respect to the cluster assignment procedure. For example, there are uncertainties inherent in the assignment of any area to a supposedly watertight category, especially where clusters are not tightly defined in multivariate space (Openshaw 1995). The "fuzzy geodemographics" proposed by Openshaw (1989) is a potential resolution to this type of uncertainty. Slingsby et al. (2011) have sought to visualize proximity to adjacent 2001 OAC Supergroup clusters. More widely within GIS, the term uncertainty is used to denote that almost any representation is inherently incomplete. These problems are compounded when GIS representations seek to accommodate change over time (Plewe 2002). In the context of geodemographics, we seek to devise temporal uncertainty indicators in order to accommodate these problems. This allows fast changing neighborhoods to be identified whilst retaining the small level granularity of conventional geodemographic classifications for the others. In the majority of locations, we argue that the 2001 OAC remains of use, avoiding the need for costly and time-consuming updating through ancillary sources that may be of unknown provenance.

2 Change Since 2001

Consistent with the classic filtering theory of urban geography (Hoyt 1939), Sleight (2004) has suggested that population change is of little overall consequence for a geodemographic classification as most areas will continue to house the same social groups over time – even if the identities of the individuals themselves change. Where local scale change does occur, however, annual mid-year population estimates provide one useful indicator of this at the OA level. The broad picture of change based on 2010 mid-year population estimates is shown in Table 3: the UK population increased by an estimated 5.33% between 2001 and 2010, with Northern Ireland experiencing the fastest growth in population, and Scotland the slowest.

Figure 2 illustrates the maximum absolute deviation from the 2001 OA populations for England and Wales over the period 2001 to 2010. The bimodal distribution becomes more pronounced over time and by 2010 every OA in England and Wales had experienced at least


2002
2003
2004
2005
2006
2007
Number of Output Areas
Percentage Change
0
2000
4000
6000
8000
10000
12000
-100
-75
-50
-25
0
25
50
75
100
125
150
175
200 and Over

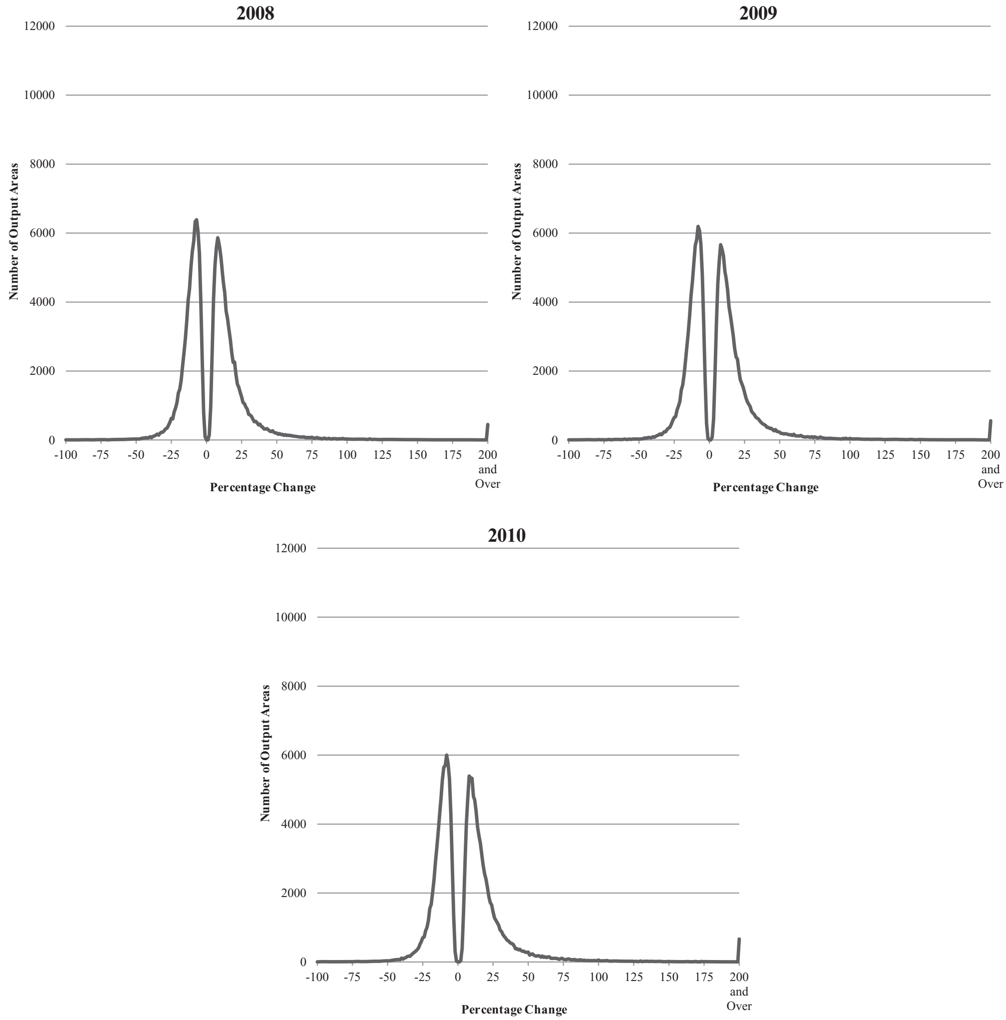


Figure 2 Maximum population change since 2001 in England and Wales from 2002 to 2010

1% population change, with 55.1% of OAs increasing in population. This global figure conceals important local variations: for example, London's population increased by 9.1% over the same period (ONS 2011), but the populations of Westminster and Tower Hamlets, both Boroughs of London, saw population increases of 24.8% and 18.3% respectively, while that of Brent, another London Borough, decreased by 4.8%. Glasgow, in comparison, saw a more modest increase of 2.4% between 2001 and 2010. Still, local changes in population size provide an important indicator of changes in social, economic and demographic conditions at an OA scale. It can be envisaged that in such areas changes should be accommodated in the use of geodemographic classifications. The ability to identify such areas becomes more important during the life of any geodemographic classification as the likelihood of uncertainty in the continuing validity of the original group assignment increases.

3 Monitoring the Emergence of Uncertainty Over Time

Given that population change in the UK is inherently unevenly distributed geographically, there is a need for additional measures that seek to quantify such changes and provide a comprehensive update of the 2001 OAC (and others like it) between censuses. Traditional measures of small area estimation (see Rao 2003), such as regression models (Fay and Herriot 1979), Bayesian methods (Congdon 2004) or M-quantile models (Chambers and Tzavidis 2006, Tzavidis et al. 2010) can offer synthetic estimates of change. The benefits of using such measures, however, are reliant on multiple datasets being made available at the smallest spatial scale and ensuing updates occur on a regular basis. At present, the lack of these resources prevents the realistic utilization of small area measures.

Here we adopt the more straightforward approach of using the limited data sources that are currently available at the OA level to construct temporal uncertainty indicators. This has the benefit of keeping the same geodemographic groups and giving additional information on how likely it is that these groups still represent an area. Of course, this falls short of any ability to reassign areas flagged as uncertain to a different class, or any newly created group to best encapsulate the emergent characteristics of a new group of areas. This situation is likely to be addressed over time if more open data are made available or modeled at the finest level of granularity. Here, our argument is that there are two key factors that impact upon the reliability of the 2001 OAC assignments over time: (1) the extent to which the resident population size is likely to have changed; and (2) (where available) the nature and amount of recorded changes to the dwelling stock. A dwelling is defined as comprising a single household space or several household spaces sharing some facilities (see www.gov.uk), and change in the dwelling stock enumerates the changes in any given area. There is likely to be a strong inter-relationship between changes in population and dwelling stock: for example, dilapidated housing stock might be cleared and replaced with new developments at different residential densities. In other cases there may be no such link – for example, existing housing stock may become occupied at higher residential densities by incomers, or redevelopment may not lead to changes in residential density. In each of these cases, however, we suggest that changes in either or both of these indicators is likely to lead to differences in the demographic characteristics of OAs, along with changes in the numbers of individuals likely to bear these characteristics. While the totality of demographic change is unlikely to be captured by these two measures, we nevertheless suggest that they provide a measure of the reliability of local level demographic estimates and also provide an insight into how safe a geodemographic classification is to use. For areas for which the classification is rendered unusable because of temporal change, it becomes incumbent upon the user to source additional data to augment or replace the classification itself.

Mid-year population estimates are produced on an annual basis separately for England and Wales, Scotland and Northern Ireland. In England and Wales they are produced at the OA level by a single age band for both males and females. This is also true of Scotland, save that they are produced at the Data Zone level. Northern Ireland's output is different in that estimates are produced at the Super Output Area level for four age bands. While coverage of the UK is at varying levels of geography the different measures are useful because in each case they correspond to high granularity census geography. Dwelling stock data, classified by Council Tax band, are made available by the ONS for England and Wales at the OA level, but there are no equivalent open data for Scotland and Northern Ireland. Council Tax levies are based on the capital value of residential property in England, Wales and Scotland. The Valuation Office Agency (VOA) assigns every residential property to a valuation band and there is a clear financial incentive to maintain an up-to-date register. This is available in the public domain for

OAs in England and Wales. The lack of freely available data for Scotland and Northern Ireland creates a disparity in coverage if this data source is being used to form a single temporal uncertainty indicator for the 2001 OAC, with the 21% of UK OAs assigned to Scotland and Northern Ireland not being represented. This current limitation does not have to prevent dwelling stock data from being used solely or to form part of temporal uncertainty indicators for England and Wales. However the disaggregation of total dwelling stock estimates into Council Tax bands, a means of further differentiating change and also a proxy for changes in housing wealth (Harper and Mayhew 2012), is hindered by a change in 2005, when a national revaluation of property in Wales (VOA 2005) rendered them incomparable with their English equivalents. Nevertheless, while the bands are not compatible between the two countries they do allow for general changes in the values of property at the small area level to be seen; for example redevelopment and attendant upgrading of low-cost housing since 2001.

Additional small area change measures might be developed from open data sources in the future. For example, crime data for Great Britain are made available on a monthly basis by easting and northing, and by crime type (see www.police.co.uk) in a form suitable for aggregation to OA level. However, the data do not currently have full UK coverage (the Police Service of Northern Ireland aggregate data to coarser policing areas) and there is a need to research the volatility and reporting bias in small area estimates before developing usable small area measures.

4 Uncertainty and the 2001 Output Area Classification

The limitations of the available data to construct temporal uncertainty indicators of the 2001 OAC mean that at present there is no way to produce a comprehensive indicator of uncertainty at the OA level. Accordingly, we proceed here using the best available indicators at the OA level – annual mid-year population estimates, and annual dwelling stock figures available for England and Wales only; Council Tax band data for England and Wales are not used because of the Welsh revaluation and its implications for data compatibility. Our basic hypothesis is that changes in these indicators of the attributes (dwelling stock) and characteristics (population size) in each OA provide a reliable indicator of the enduring relevance of 2001 OAC to describing its geodemographic characteristics. For each indicator, we record the maximum absolute deviation from the 2001 value over the period 2001 to 2010. This allows us to flag the occurrence of change in circumstances in which an initial increase (or decrease) is subsequently compensated for by a subsequent decrease (or increase) to a value that might suggest that little change had occurred over the entire period. Mid-year population estimates are not available for OAs outside England and Wales, so coarser Data Zones are used for Scotland and Super Output Areas are used for Northern Ireland (in Scotland there are an average 6.5 OAs for every Data Zone and in Northern Ireland 5.6 OAs for every Super Output Area). In addition to using the population and dwelling stock change to form temporal uncertainty indicators, a combination of the two is utilized for England and Wales. The respective change shown by each indicator is standardized using z-scores and brought together to form an overall composite score, where data for each indicator are available (see Table 4).

The reliability of using mid-year population estimates as a change indicator over time is influenced by the methodology used to produce them. There are slight differences in the methods used to calculate the mid-year population estimates in each UK country although the three responsible organizations each use a common cohort component method (ONS 2010) to update the population base. Essentially the 2001 UK Census is used as a population base and

Table 4 Temporal uncertainty indicators

Temporal Uncertainty Indicator	Description	Geographical coverage	Data source
Population	Mid-year population estimates from 2002 to 2010 are used to calculate the maximum absolute deviation from the 2001 UK Census figures. This provides each OA a figure of maximum percentage change in the 2002 to 2010 period.	UK	Yearly mid-year population estimates provided by the ONS for England and Wales, NRS for Scotland and NISRA for Northern Ireland
Dwelling Stock	Dwelling stock counts from 2002 to 2010 are used to calculate the maximum absolute deviation from the 2001 figures. This provides each OA a figure of maximum percentage change in the 2002 to 2010 period.	England and Wales	Yearly dwelling stock by council tax band counts provided by the Valuation Office Agency
Composite	The population and dwelling stock percentage change are each standardised using z-scores. These figures are then added together to form a composite score for each OA.	England and Wales	Yearly mid-year population estimates and yearly dwelling stock by council tax band counts

then each year births are added, deaths subtracted and small area migration estimates are based on available survey and proxy data. In England, Wales and Scotland data on international migration come from the International Passenger Survey (IPS), the Labor Force Survey (LFS) and Home Office data on asylum seekers and their dependants (ONS 2010). In Northern Ireland the inflows are estimated from the list of patients registered with a family doctor and the outflows from the number of people who have de-registered: data from the Irish Quarterly Household Survey (QNHS) are also used, to estimate migration to the Republic of Ireland (NISRA 2011). Further issues associated with migration statistics are discussed in ONS (2011) for England and Wales, in GROS (2010) for Scotland and in Dignan et al. (2010) and IJpelaar et al. (2011) for Northern Ireland.

The release of 2011 UK Census data provides an opportunity to critique the accuracy of current mid-year population estimates. At present the finest level of granularity available for 2011 UK Census data are at local authority level for England and Wales. The ONS has identified that for England and Wales the population of males aged 10 to 19 and 30 to 39 is larger than that suggested by the population estimates for March 2011, while the opposite is true for the male population aged 20 to 29 (ONS 2012b). Amongst other discrepancies, the March 2011 population estimates are too high for the 25 to 29 age group in some university areas (ONS 2012b). The accuracy of the mid-year population estimates is thus likely to be spatially variable, with knock on consequences for any temporal uncertainty indicator that utilizes them. Dwelling stock counts, and their change since 2001 are at present the only viable alternative indicator to the uncertainty of the 2001 OAC. Unlike mid-year population estimates, these provide a definitive count rather than an estimate, so a greater level of certainty may be attached to the figures, although this does not make them more important in evaluating the broader picture of temporal change. Using change in dwelling stocks as a temporal uncertainty indicator provides a general overview of how property in an area has changed and provides an alternative insight to relying on population estimates alone to characterize change at the small area level. Dwelling stock changes are thus a further indicator of stability in small area geodemographic assignment.

In England and Wales a comparison between the two data sources is possible. There are respective merits of using mid-year population estimates and dwelling stock change as temporal uncertainty indicators. While there is undoubtedly a strong inter-relationship between the two, this is not constant across England and Wales. Table 5 presents a confusion matrix of the areas of change as indicated by the two sources. The data for England and Wales have been ranked and divided into deciles for each temporal uncertainty indicator. The OAs that share the same decile for both of the indicators suggest the uncertainty created in these areas derives equally from population and dwelling stock change. For the OAs where the temporal uncertainty indicator deciles do not match, this would suggest that either population or dwelling stock is driving the uncertainty in those areas, but not both. Decile 1 contains the OAs that have experienced the most change from 2001 to 2010 and decile 10 the least. (Because 11% of OAs in England and Wales experienced no change in dwelling stock over the past decade there is no decile 10 for dwelling stock change). The relationship between mid-year population estimates and dwelling stock change shown in Table 5 lacks any strong correspondence, alluding to the conclusion that the two indicators pick up different aspects of change across different areas in England and Wales. This suggests that in areas where it is possible, using the temporal uncertainty indicator that combines both mid-year population estimates and dwelling stock change would be preferable. In part because areas that have experienced change in both population *and* dwelling stock have an increased likelihood of being misrepresented by their current geodemographic assignment.

Table 5 Population and Dwelling Stock temporal uncertainty indicators confusion matrix

	Decile Number	Population temporal uncertainty Indicator (Percentages)										
		Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Total
Dwelling Stock temporal uncertainty Indicator (Percentages)	Decile 1	4.36	1.14	0.80	0.66	0.63	0.54	0.53	0.51	0.47	0.45	10.10
	Decile 2	1.80	1.65	1.17	0.97	0.86	0.82	0.76	0.70	0.66	0.63	10.01
	Decile 3	0.95	1.41	1.25	1.11	1.07	0.97	0.88	0.85	0.79	0.73	10.00
	Decile 4	0.72	1.20	1.19	1.11	1.05	1.03	0.96	0.95	0.91	0.88	10.00
	Decile 5	0.55	0.95	1.10	1.11	1.07	1.09	1.05	1.04	1.03	1.01	9.98
	Decile 6	0.45	0.90	1.04	1.07	1.05	1.09	1.14	1.11	1.16	1.08	10.09
	Decile 7	0.42	0.81	0.93	1.00	1.09	1.07	1.16	1.15	1.17	1.14	9.93
	Decile 8	0.38	0.71	0.90	1.01	1.11	1.10	1.17	1.15	1.21	1.24	9.99
	Decile 9	0.47	1.23	1.66	1.94	2.09	2.28	2.39	2.53	2.56	2.74	19.89
	Decile 10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Total	10.11	10.00	10.04	9.96	10.01	9.99	10.05	9.99	9.96	9.90	100.00

The different aspects of change picked up by population and dwelling stock temporal uncertainty indicators, along with a combined composite measure, provide an indication of geodemographic change in England and Wales over the past decade. Figure 3 displays the distribution of change for each temporal uncertainty indicator, with the additional inclusion of population change for England and Wales, Scotland and Northern Ireland for reference. Table 6 identifies thresholds of change, beyond which the 2001 OAC classification is deemed unreliable. These threshold values, like many decisions in geodemographic classification, are subjective. The thresholds values were based on identifying areas that were within one standard deviation, around 68.2% of the UK's OAs, and classing them as certain. Manual intervention was required to decide upon the final threshold values (see Table 6) to allow for the greatest compatibility between the temporal uncertainty indicators possible, but also to limit the areas classified as uncertain to locations where the more extreme changes in local characteristics have taken place. Overall, the percentage of OAs classed as uncertain by each of the three temporal uncertainty indicators ranges from 21% to 29%.

Figure 4 illustrates how the combinations of the three temporal uncertainty indicators (population change; dwelling stock change; and the composite measure) can be used alongside threshold values. Each map is a cartogram, in which every OA has been rescaled in direct proportion to its 2010 total population, in order to best visualize where change has happened. The population temporal uncertainty indicator shows a large number of OAs in the Greater London area have experienced population change over the threshold value. The remainder of England and Wales has a fairly even distribution of above threshold values, but other urban areas such as Manchester and Birmingham dominate their respective local areas. The dwelling stock temporal uncertainty indicator provides a different picture of change in England and Wales. OAs that have experienced change greater than the threshold values are predominately distributed in the South East and South West of England, where 34% of all OAs in these two regions have experienced dwelling stock change greater than the threshold value. The composite temporal uncertainty indicator has a different geographical distribution again; although, as Table 6 indicates it designates fewer OAs in total as uncertain when compared with the two other temporal uncertainty indicators. Of the population and dwelling stock temporal uncertainty indicators, it is the population measure that has the more even geographical distribution across England and Wales, albeit with higher concentrations of uncertainty in urban areas. Change in the dwelling stock indicator is particularly marked in the South East and South West of England. The composite indicator also suggests greatest incidence of uncertainty in the South East and South West of England, and also concentrations in urban areas across England and Wales.

The three temporal uncertainty indicators identify different areas across England and Wales that have experienced the most change. Table 7 segments these results according to 2001 OAC Supergroup. A third of the OAs identified as having experienced change above the population temporal uncertainty indicator threshold are in the "Typical Traits" and "Multicultural" Supergroups, suggesting that the uncertainty of these two Supergroups is heavily driven by population change. This provides only part of the picture as "Typical Traits" is also influenced by changes to dwelling stock, as over 20% of OAs identified as having above threshold change to dwelling stock are located in this Supergroup. Compared to the 10% figure for the "Multicultural" and the 21% for the "Prospering Suburbs" Supergroups it is clear that different combinations of change drive the uncertainty of geodemographic types to varying extents. The composite temporal uncertainty indicator provides only a slight variation to the distributions seen with the dwelling stock measure. While arguably just an artifact of the threshold values used as the indicator, there is a suggestion that the more extreme change, and therefore

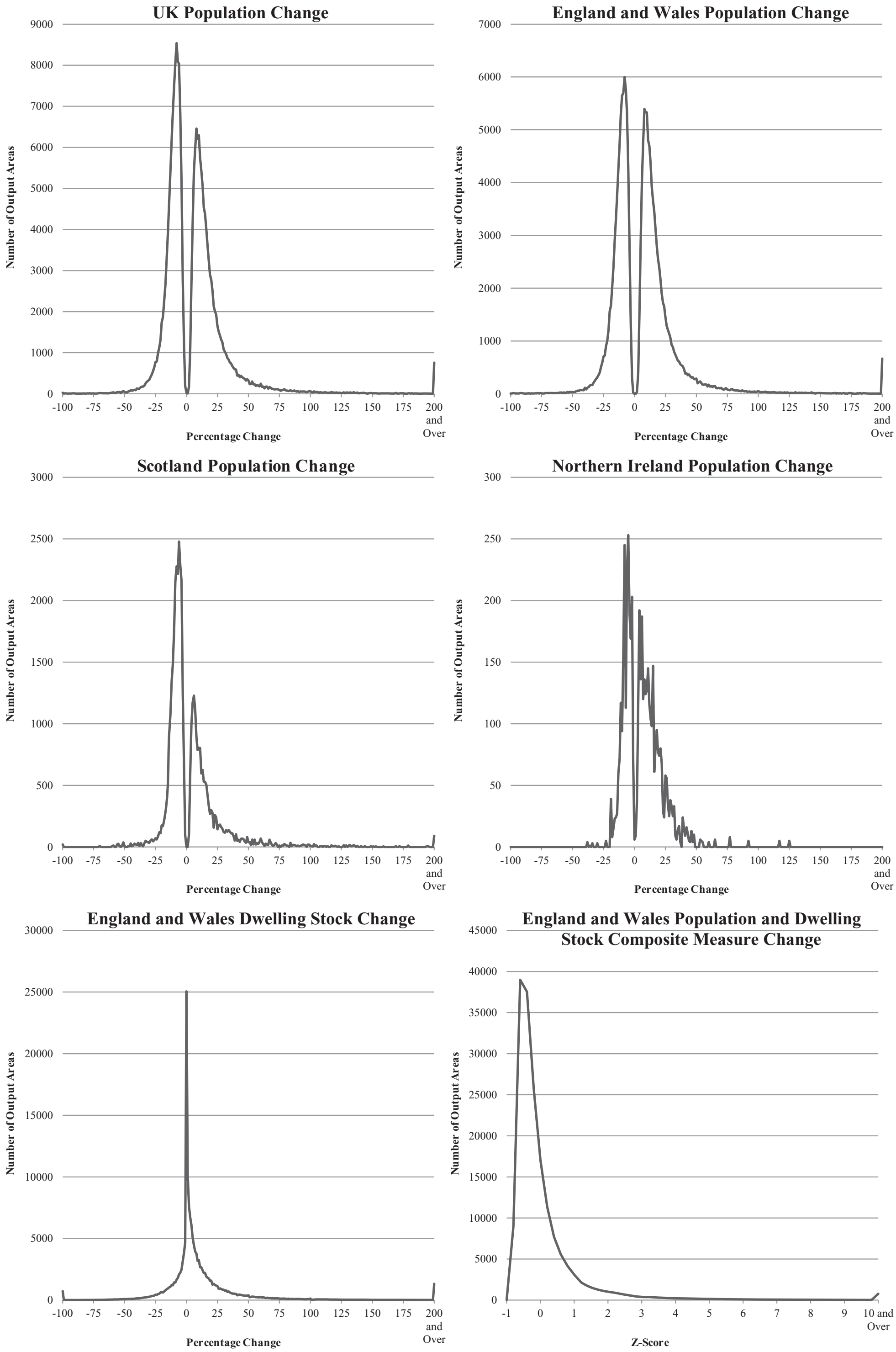


Figure 3 Change distribution in 2010 of temporal uncertainty indicators

Table 6 Threshold distribution of temporal uncertainty indicators

Temporal Uncertainty Indicator	Negative Threshold Value	Positive Threshold Value	Average percentage of OAs Below Threshold	Average percentage of OAs Above Threshold	Below Threshold to Above Threshold Ratio
Population	–15%	20%	71	29	2.5 : 1
Dwelling Stock	–15%	20%	74	26	2.8 : 1
Composite	–0.8	0.4	79	21	3.7 : 1

uncertainty, seen across the 2001 OAC Supergroups is driven more by dwelling stock change than by population change alone.

Table 8 illustrates that across the regions of England and Wales there are variations in the percentage of OAs that have above threshold levels of change for each temporal uncertainty indicator. London for example has 46% of OAs classed as uncertain if using the population temporal uncertainty indicator, but only 19% or 21% if using the dwelling stock or composite temporal uncertainty indicators, respectively. Similar dominance of the population temporal uncertainty indicator at identifying uncertainty can be found in Yorkshire and Humberside along with the East Midlands. Conversely, a different pattern is seen in the South East and South West of England, with the dwelling stock change indicating greater uncertainty than the other two measures. The variability seen in uncertainty picked up by the population and dwelling stock temporal uncertainty indicators between the regions is not repeated for the composite measure. For the population and dwelling stock temporal uncertainty indicators there is a range of 19 and 18%, respectively between the regions in the amount of uncertainty picked up. For the composite temporal uncertainty indicator this is just 5%, indicating this measure has an increased stability across England and Wales, with just over one in five OAs being identified as uncertain using this measure.

Figures 5, 6 and 7 present the composite temporal uncertainty indicator for London's Boroughs (see Figure 8), broken down by 2001 OAC Supergroups. Approximately 22% of London OAs classified as "Multicultural" exceed the threshold, as do a similar percentage of "City Living" neighborhoods. The "Multicultural" and "City Living" Supergroups together comprise over 75% of London's OAs, and only 15% of OAs assigned to any of the other five, more settled, Supergroups exceed the threshold. Figure 5 indicates that the majority of uncertainty in the "Multicultural" group is found in the East of London, in the Boroughs of Tower Hamlets, Newham, Hackney and Barking and Dagenham. While there are pockets of uncertainty found elsewhere around the capital, these areas are in the minority. Figure 6 suggests the City of London and the City of Westminster in the centre of London, is where the greatest uncertainty in the "City Living" assignments are found. The distribution of uncertainty shown in Figure 7 reflects the five Supergroups being displayed. There is no distinct pattern, with no single area having a particularly high concentration of uncertainty, with the areas identified as uncertain being found in isolated pockets situated around the outer boroughs of London. These results reflect the dominance of the "Multicultural" and "City Living" Supergroups in London, with over three-quarters of the capital falling into one or other of these two groups. This dominance in comparison with the other Supergroups means that areas containing a high

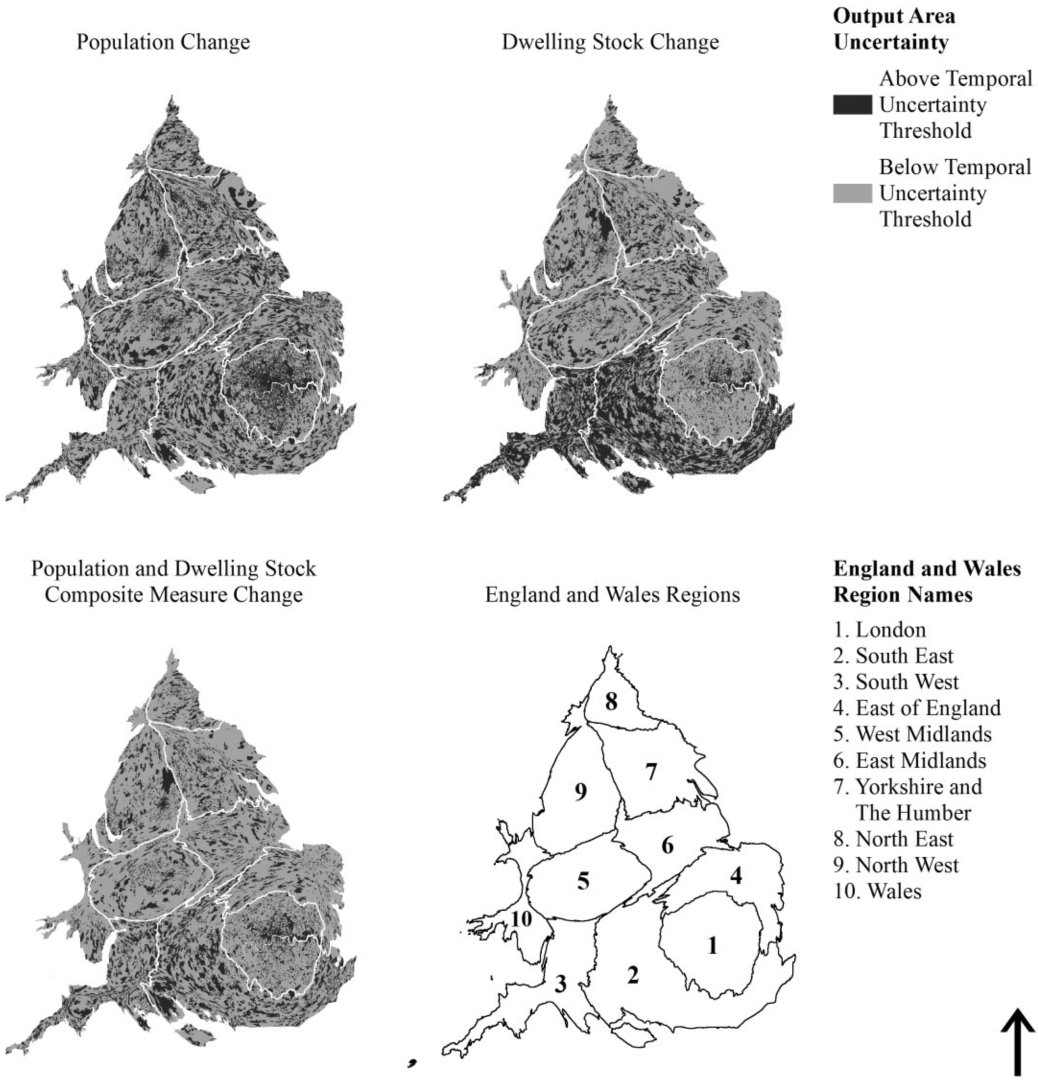


Figure 4 Thresholds of temporal uncertainty indicators in England and Wales, viewed as cartograms
Map data from Office for National Statistics: 2001 UK Census, Output Area Classification 2001, Mid-Year Population Estimates and Valuation Office Agency: Dwelling Stock by Council Tax Band © Crown Copyright. Contains Ordnance Survey Data © Crown Copyright and Database Right 2011

concentration of uncertainty, as identified, have developed over the past decade. This has not been possible for the other Supergroups because of their more sporadic geographically distribution across London.

The only temporal uncertainty indicator with coverage for Scotland and Northern Ireland is that derived from mid-year population estimates. Figure 9 highlights the example of Glasgow and surrounding area for each of the seven Supergroups in the 2001 OAC. "Constrained by Circumstances" is the dominant Supergroup in Glasgow, with 55% of OAs assigned to the group. This dominance does not however translate to an increased propensity

Table 7 Above threshold percentage distribution of the temporal uncertainty indicators by 2001 OAC Supergroup

Temporal Uncertainty Indicator / 2001 OAC Supergroup	Population temporal uncertainty Indicator	Dwelling Stock temporal uncertainty Indicator	Composite temporal uncertainty Indicator
Blue Collar Communities	11	15	15
City Living	13	7	7
Countryside	12	14	12
Prospering Suburbs	14	21	21
Constrained by Circumstances	15	11	11
Typical Traits	17	21	20
Multicultural	17	10	13
TOTAL*	100	100	100

*Figures may not sum exactly due to rounding

Table 8 Above threshold percentage distribution of the temporal uncertainty indicators by regions in England and Wales

Temporal Uncertainty Indicator / Regions in England and Wales	Population temporal uncertainty Indicator	Dwelling Stock temporal uncertainty Indicator	Composite temporal uncertainty Indicator
East of England	29	29	21
East Midlands	30	21	19
London	46	19	21
North East England	27	25	22
North West England	30	29	24
South East England	29	32	20
South West England	30	37	22
Wales	29	29	21
West Midlands	29	23	19
Yorkshire and the Humber	29	19	21

for uncertainty, as only 31% of OAs assigned to this group are classed as uncertain. In addition, there is no distinct patterning to this uncertainty. It does not appear that particular areas of uncertainty have developed over the past decade, unlike, for example, the "Multicultural" or "City Living" groups in London. There is a similar pattern with the other Supergroups in Glasgow, where no distinct concentrations of uncertainty have developed. The exception to this is the "City Living" group where the majority of uncertain areas are located in the centre of Glasgow. This apparent difference between Glasgow and London can in part be explained by the total proportion of OAs that have been classed as uncertain. London's dominate "Multicultural" Supergroup has over half of the OAs assigned to that group classed as "uncertain"

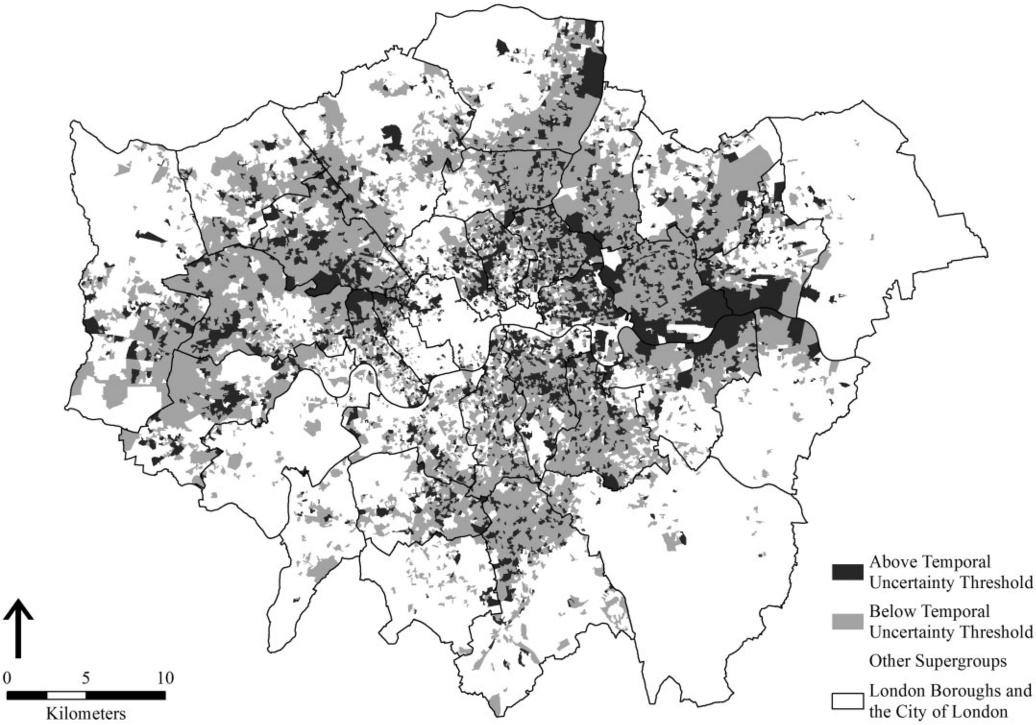


Figure 5 Distribution of 'Multicultural' Supergroup OAs falling above and below the composite temporal uncertainty indicator threshold values
Map data from Office for National Statistics: 2001 UK Census, Output Area Classification 2001, Mid-Year Population Estimates and Valuation Office Agency: Dwelling Stock by Council Tax Band © Crown Copyright. Contains Ordnance Survey Data © Crown Copyright and Database Right 2011

when using the population temporal uncertainty indicator (unlike the 31% previously identified for Glasgow's dominant Supergroup). In terms of overall population change, London is a more uncertain city than Glasgow, having 46%, compared with 27%, of OAs above the uncertainty threshold. A full national coverage of these and results for other parts of the UK is available at http://www.opendataprofiler.com.

Analysis of how uncertainty varies between any parts of the UK can only be undertaken using the population temporal uncertainty indicator. At this level the change in the distributions of the 2001 OAC Supergroups between 2002 and 2010 is shown in Figure 10. The limited change in 2002 increases steadily through to 2010, with the "Blue Collar Communities" and "Prospering Suburbs" Supergroups experiencing limited change relative to the "Multicultural" and "City Living" groups. It is evident that neighborhoods assigned to different geodemographic groups have differing propensities to change. Table 9 presents the percentage changes for each of the 2001 OAC Supergroups between 2002 and 2010, broken down into the constituent parts of the UK. The changes in England and Wales are concentrated towards the higher population change categories, while those in Scotland and Northern Ireland tend to be smaller in magnitude. These results suggest that OAs in England and Wales are proportionally more likely to have changed since 2001. This overall change can be further sub-divided by 2001

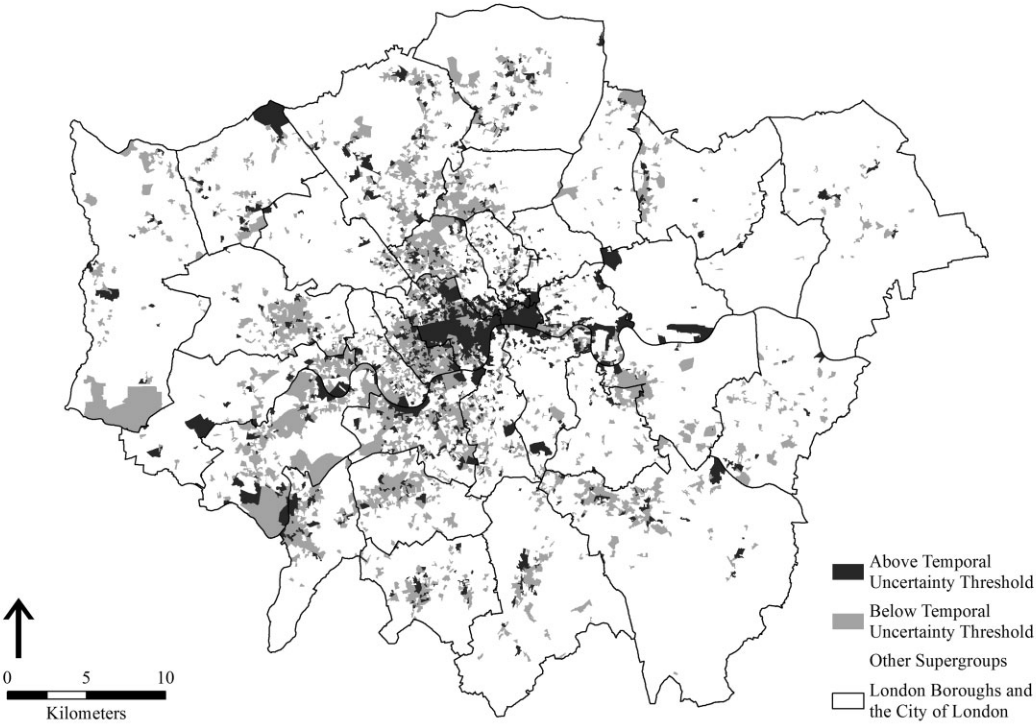


Figure 6 Distribution of 'City Living' Supergroup OAs falling above and below the composite temporal uncertainty indicator threshold values
Map data from Office for National Statistics: 2001 UK Census, Output Area Classification 2001, Mid-Year Population Estimates and Valuation Office Agency: Dwelling Stock by Council Tax Band © Crown Copyright. Contains Ordnance Survey Data © Crown Copyright and Database Right 2011

OAC Supergroup to accommodate both the effects of location and geodemographic characteristics to determine the level of uncertainty associated with use of the classification.

5 Conclusions

Geodemographics continue to be widely used as local area discriminators, across a wide spectrum of business and service planning applications. Commercial systems remain popular, despite concerns (particularly in public service applications) about their closed and "black box" nature. We believe that this arises in no small part because such solutions are known to be frequently updated, *inter alia* using commercial data sources that are not widely available. Many users are concerned that the "best" solutions require the "best" data, and "frequently updated" is often taken as a surrogate measure for "best", despite the fact that the provenance of most commercial sources is largely unknown. There are good reasons, however, to suggest stability in population and settlement structure – classically espoused in Hoyt's (1939) notion of filtering in urban structure, whereby the social, economic and demographic structure of neighborhoods remains stable over time, even if the identities of the residents themselves turn over much more rapidly. This article has investigated the extent to which these notions of sta-

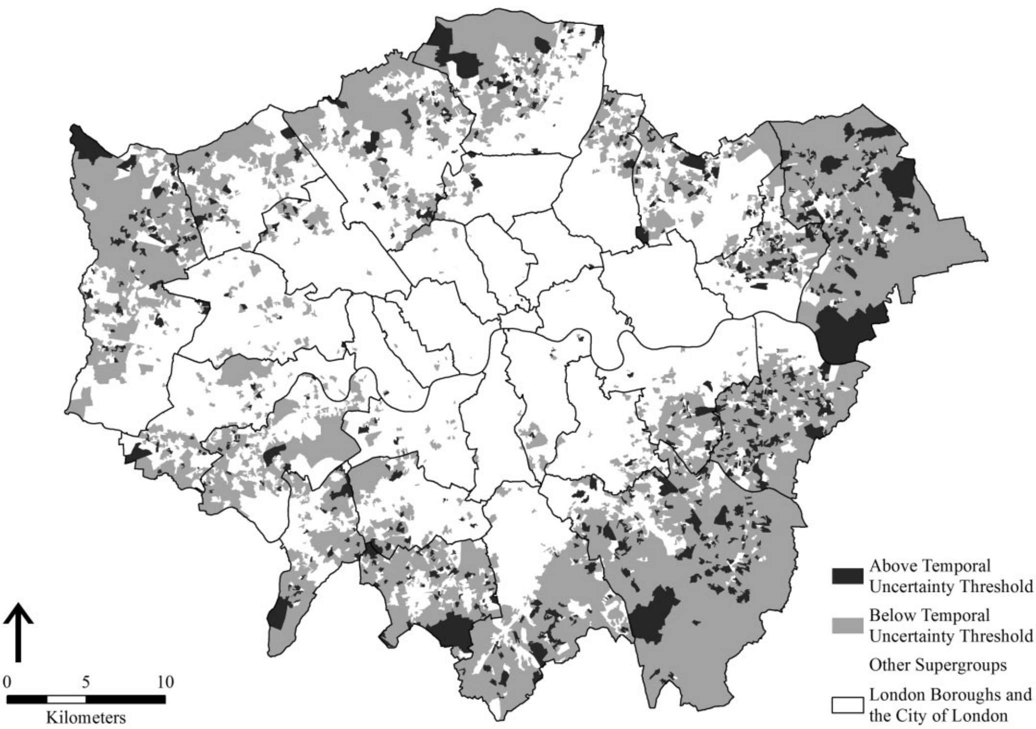


Figure 7 Distribution of 'Blue Collar Communities', 'Countryside', 'Prospering Suburbs', 'Constrained by Circumstances' and 'Typical Traits' Supergroups OAs falling above and below the composite temporal uncertainty indicator threshold values
Map data from Office for National Statistics: 2001 UK Census, Output Area Classification 2001, Mid-Year Population Estimates and Valuation Office Agency: Dwelling Stock by Council Tax Band © Crown Copyright. Contains Ordnance Survey Data © Crown Copyright and Database Right 2011

bility and change play out in geodemographic terms, using the 2001 OAC. In practice, the results and the associated website http://www.opendataprofiler.com provide empirical evidence of the likely stability or otherwise of the 2001 OAC Supergroups in different parts of the country. A practical implication of this is that users of the 2001 OAC can have increased confidence in the use of the classification in areas where our analysis suggests that change has been much more muted.

As such, our analysis suggests areas and target groups for which the frequent updating of commercial geodemographic classifications may be unnecessary. Our use of mid-year population estimates and dwelling stock data to construct multiple temporal uncertainty indicators provides a reliable means of gauging the stability or otherwise of neighborhood conditions. The conclusion from this is that while a large number of small areas in the UK do experience change over time, this change is not uniform in either amount or distribution. Compared to using the 2001 OAC in isolation, the advantage of knowing which areas may no longer resemble their initial classification designation becomes clear. The user is then aware of the need to investigate such areas using alternative data sources in order to better understand the current population and dwelling dynamics and make more informed decisions.

The creation of temporal uncertainty indicators does, in part, address some of the perceived inadequacies of the 2001 OAC when compared to commercial alternatives, and high-

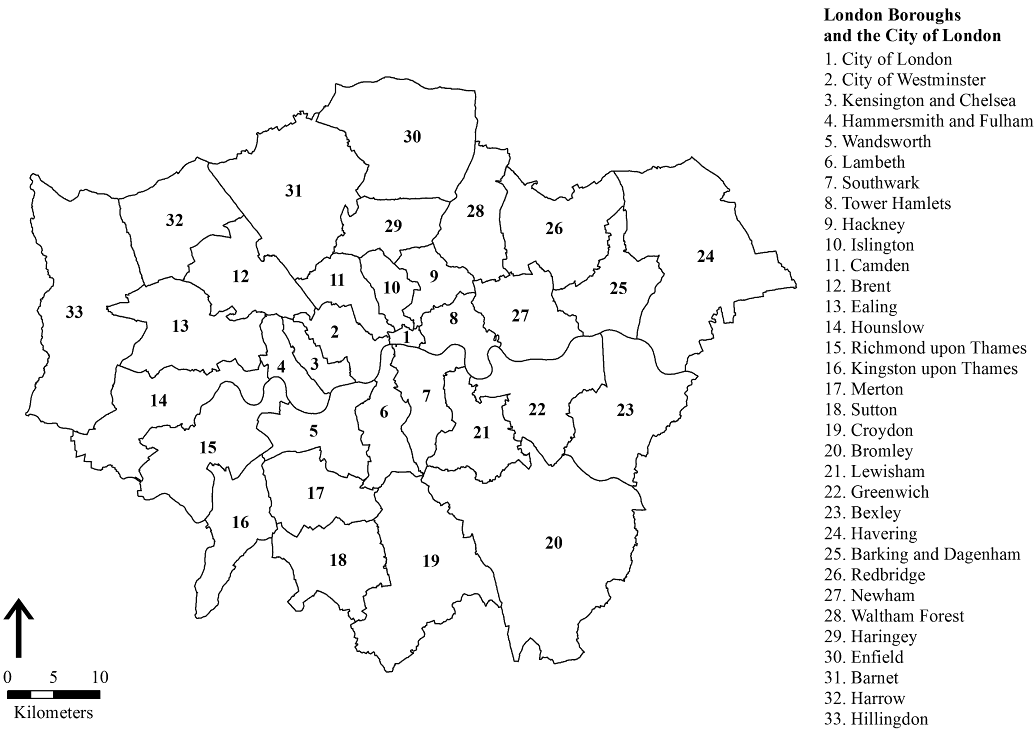


Figure 8 London Boroughs and the City of London
Contains Ordnance Survey Data © Crown Copyright and Database Right 2011

lights stability across the UK and England and Wales. The three temporal uncertainty indicators reveal that only one in four OAs have experienced change greater than the defined threshold limits. Despite these OAs being grouped as uncertain the extent of change in the socio-economic characteristics will vary, to the extent it would seem unlikely that *all* of these OAs would require reclassification. It is of course important to consider the geographic variance in what a significant change in population or dwelling stock would be. Longley et al. (2011) suggest that a large proportion of the British population have remained settled for at least 600 years with the possible exception of urban conurbations, such as London, Birmingham and Manchester. Our analysis also provides an insightful analysis of the likely geodemographic breakdown of change in the UK over the last decade, using subjectively defined thresholds to identify significant change to either population, dwelling stock or a combination of the two. We also identify the regional variation in change.

These findings need to be tempered with the qualification that mid-year population estimates are themselves inherently uncertain, and that the indicators that we have used are less comprehensive in scope and application than the 41 census variables that underpin the 2001 OAC. There is also the associated issue that the greater the estimated population change, the greater the uncertainty associated with the estimate of it. Data issues in Scotland and Northern Ireland further compound these qualifications, where updating is only possible at higher levels of granularity. The problem of data-mismatch between countries in the UK is unlikely to be resolved in the near future, and new open data sources are not likely to be released at neighborhood levels of granularity either.

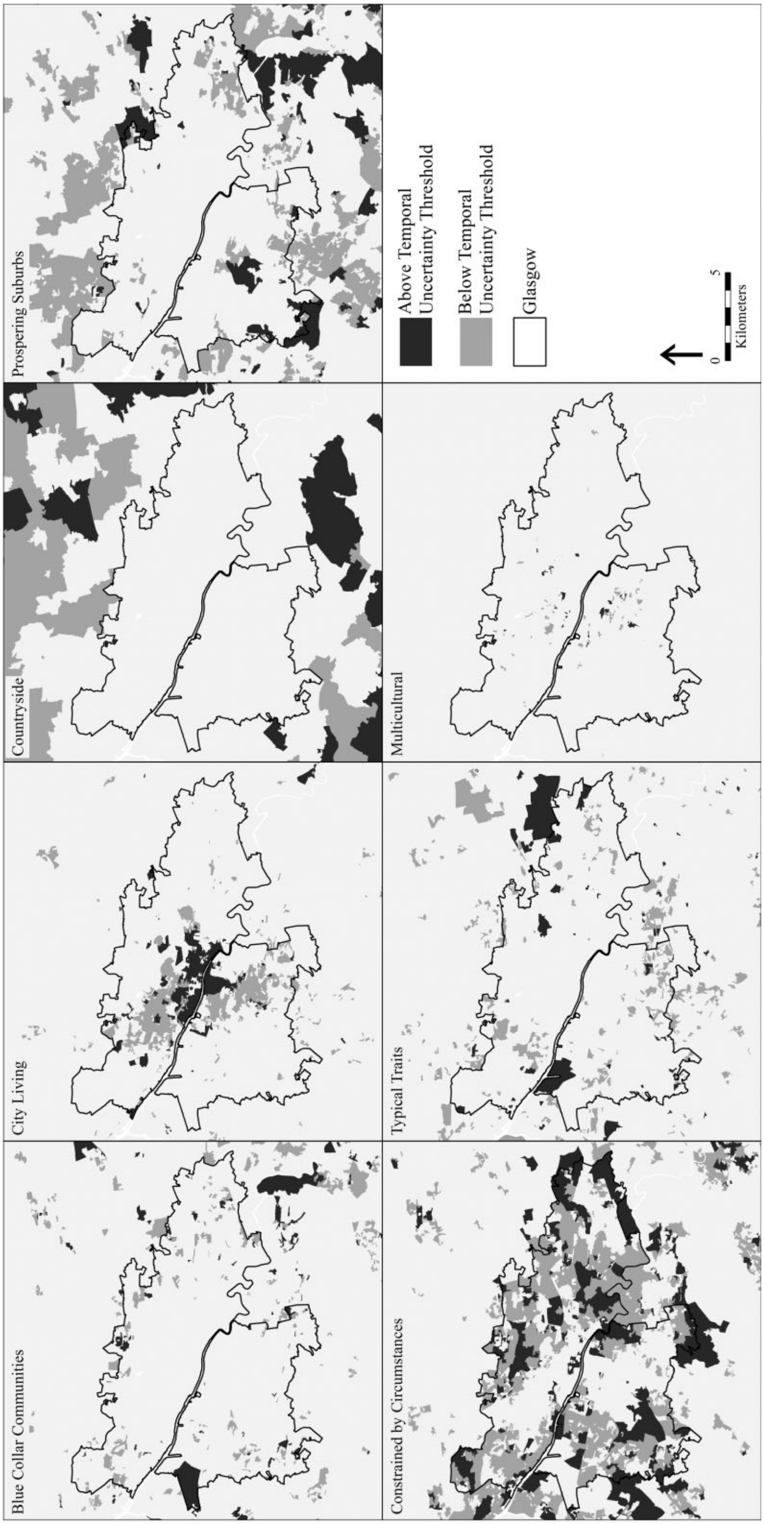
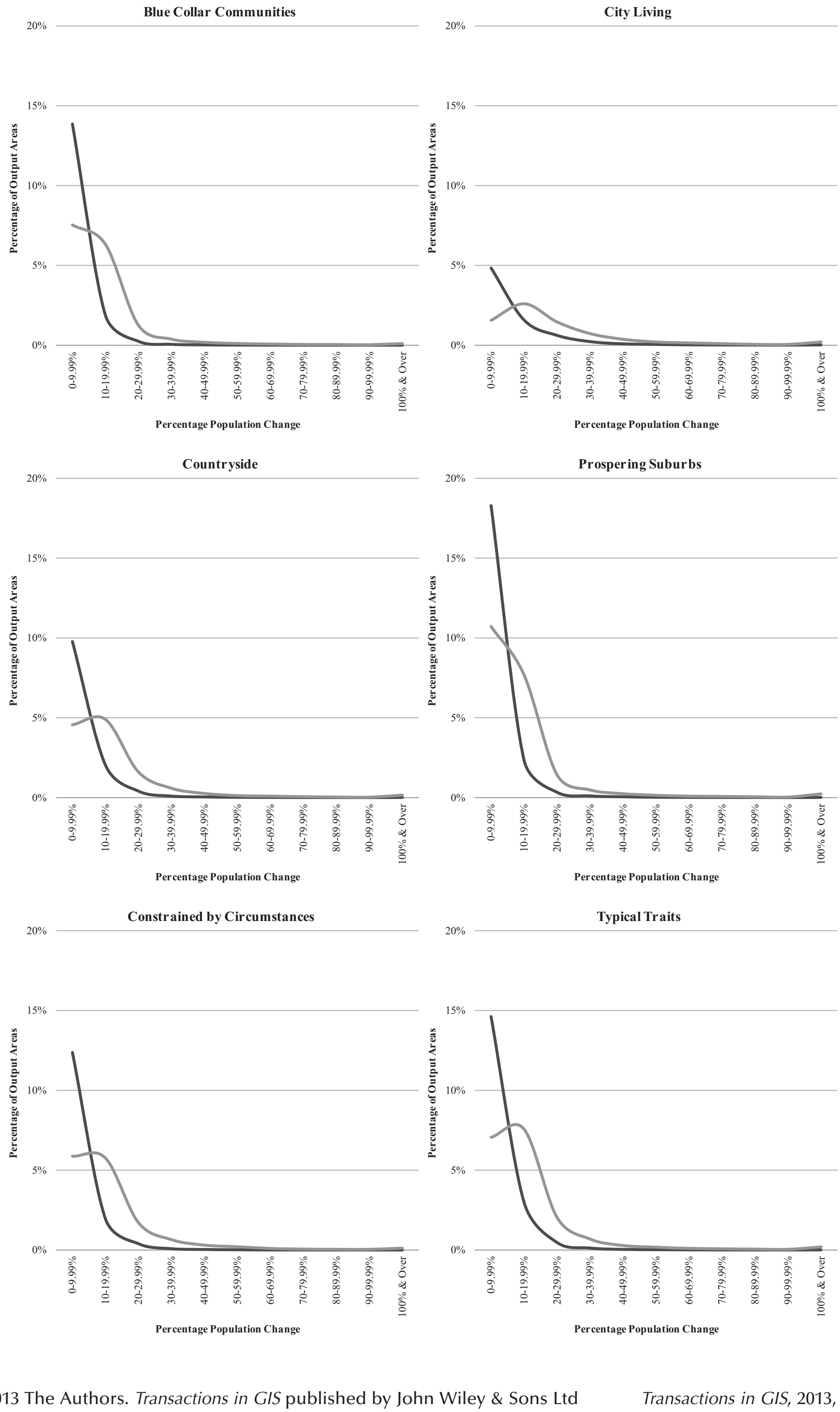


Figure 9 2001 OAC Supergroups in the Greater Glasgow region falling above and below the population temporal uncertainty indicator threshold values
Map data from Office for National Statistics: 2001 UK Census, Output Area Classification 2001 and National Records of Scotland: Mid-Year Population Estimates © Crown Copyright. Contains Ordnance Survey Data © Crown Copyright and Database Right 2011

Table 9 Percentage distribution change of the population temporal uncertainty indicator between 2002 and 2010 by 2001 OAC Supergroup for England and Wales (EW) and Scotland and Northern Ireland (SNI)

		2001 OAC Supergroup													
		Blue Collar Communities		City Living		Countryside		Prospering Suburbs		Constrained by Circumstances		Typical Traits		Multicultural	
		EW	SNI	EW	SNI	EW	SNI	EW	SNI	EW	SNI	EW	SNI	EW	SNI
Percentage Population Change	0.00 to 9.99	−47	−42	−78	−54	−54	−51	−43	−36	−59	−47	−54	−36	−70	−67
	10.00 to 19.99	173	3108	26	1173	105	2778	195	968	105	1208	138	1823	44	591
	20.00 to 29.99	341	4040	104	3742	251	2176	274	1788	241	1374	319	1400	135	1000
	30.00 to 39.99	479	100	167	100	411	4900	281	1344	409	100	424	100	179	100
	40.00 to 49.99	695	1300	281	100	402	6100	322	1150	507	12350	586	100	244	100
	50.00 to 59.99	785	100	246	100	423	100	427	1700	600	100	728	3300	436	100
	60.00 to 69.99	1671	100	419	100	470	100	413	100	1850	100	900	100	478	100
	70.00 to 79.99	2375	100	359	100	836	100	518	100	867	100	1656	100	848	0
	80.00 to 89.99	3650	100	494	100	644	100	389	100	3050	100	1683	100	636	100
	90.00 to 99.99	2850	100	457	100	1050	100	600	100	1900	100	2800	100	738	100
	100.00 and Over	5900	100	929	100	1195	100	1381	100	9100	100	2271	100	1204	100


Blue Collar Communities
City Living
Countryside
Prospering Suburbs
Constrained by Circumstances
Typical Traits
Percentage of Output Areas
Percentage Population Change
0%
5%
10%
15%
20%
0-9.99%
10-19.99%
20-29.99%
30-39.99%
40-49.99%
50-59.99%
60-69.99%
70-79.99%
80-89.99%
90-99.99%
100% & Over

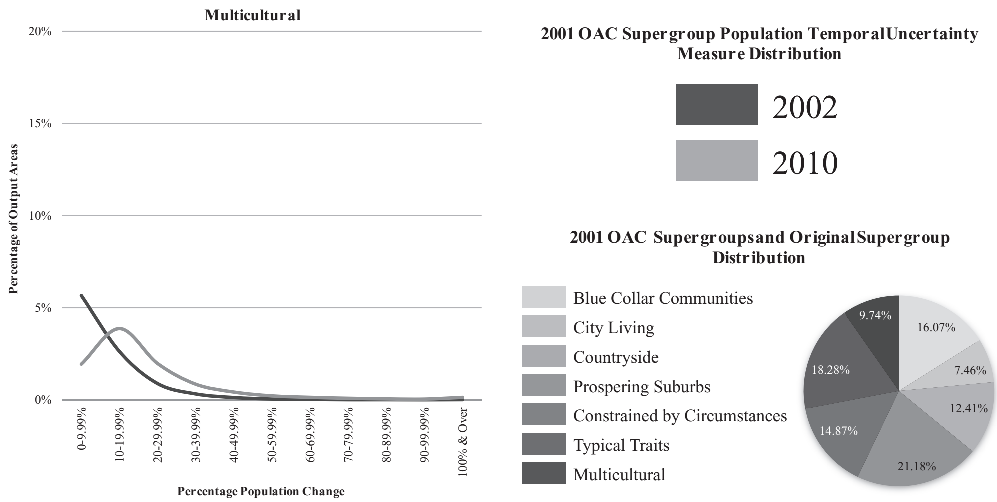


Figure 10 UK distribution of the population temporal uncertainty indicator by 2001 OAC Supergroups in 2002 and 2010

Looking to the futre, we see this work as of value in providing a framework for updating geodemographic classifications based largely or wholly upon 2011 UK Census data. The mapping on the http://www.opendataprofiler.com site provides a readily intelligible means of understanding the likely reliability of the 2001 OAC, and similar mapping might be developed as a means of representing the likely obsolescence of the data underpinning 2011 classifications across the UK. In particular these uncertainty indicators, and any additional ones that may develop from the increased provision of fine level open data, can be developed and applied to the new 2011 OAC when it is released. Moreover, we anticipate extensions of this work in modeling small area change in the other variables that are integral to geodemographic classifications. As increasing amounts of relevant open data become available, so improved methodologies may be devised in order to update classifications, and indeed identify the point at which an entire classification needs to be re-engineered. The use of emerging open data in this way in conjunction with geodemographics is a valuable direction for future research to take. The analysis reported here is of course itself uncertain, not least in the assumptions that are made in linking total population and dwelling stock data to a wider range of population characteristics: but the underpinning methodology is open and transparent and, as such, offers clear benefits over reliance upon costly data sources of unknown provenance.

References

Adnan M, Longley P, Singleton A, and Brunson C 2010 Towards real-time geodemographics: Clustering algorithm performance for large multidimensional spatial databases. *Transactions in GIS* 14: 283–97

CACI 2009 Acorn User Guide. WWW document, http://www.caci.co.uk/acorn2009/CACI.htm

Chambers R and Tzavidis N 2006 M-quantile models for small area estimation. *Biometrika* 93: 255–68

Congdon P 2010 *Applied Bayesian Hierarchical Methods*. Boca Raton, FL, Chapman and Hall/CRC

Dignan T, IJpelaar J, Marshall D, and Watson C 2010 *Small Area Population Estimates for Northern Ireland (2008)*. Belfast, Northern Ireland, NISRA Occasional Paper No. 30

Experian 2010 Optimise the Value of Your Customers and Locations, Now and in the Future. WWW document, http://www.experian.co.uk/assets/business-strategies/brochures/mosaic-uk-2009-brochure-jun10.pdf

Fay R E and Herriot R A 1979 Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association* 74: 269–77

GROS 2010 Mid-Year Population Estimates for Scotland: Methodology Guide. WWW document, http://www.gro-scotland.gov.uk/files2/stats/population-estimates/mid-year-pop-est-methodology.pdf

Harper G and Mayhew L 2012 Applications of population counts based on administrative data at local level. *Applied Spatial Analysis and Policy* 5: 183–209

Hoyt H 1939 *The Structure and Growth of Residential Neighborhoods in American Cities*. Washington D.C., Federal Housing Administration

IJpelaar J, Marshall D, Paul S, and Moylan K 2011 *Quality Report/User Guide: Northern Ireland Population Estimates*. Belfast, Northern Ireland, NISRA Occasional Paper No. 32.

Longley P A 2005 Urban studies. In Kempf-Leonard K (ed) *Encyclopaedia of Social Measurement*. Oxford, UK, Elsevier: 921–26

Longley P A, Cheshire J A, and Mateos P 2011 Creating a regional geography of Britain through the spatial analysis of surnames. *Geoforum* 42: 506–16

NISRA 2011 How are Population Estimates Created? The Methodology. WWW document, http://www.nisra.gov.uk/archive/demography/population/midyear/mye_methodology.pdf

ONS 2010 Mid-year Population Estimates Short Methods Guide. WWW document, http://www.ons.gov.uk/ons/guide-method/method-quality/specific/population-and-migration/pop-ests/mid-year-population-estimates-short-methods-guide.pdf

ONS 2011 Methodology Note on Production of Super Output Area Population Estimates. WWW document, http://www.ons.gov.uk/ons/guide-method/method-quality/specific/population-and-migration/pop-ests/methodology-note-on-production-of-super-output-area-population-estimates.pdf

ONS 2012a User Engagement on a New United Kingdom Output Area Classification: Summary of Responses. WWW document, http://www.ons.gov.uk/ons/guide-method/geography/products/area-classifications/ns-area-classifications/new-uk-output-area-classification/user-engagement-on-uk-output-area-classification.pdf

ONS 2012b Explaining the Difference between the 2011 Census Estimates and the Rolled-Forward Population Estimates. WWW document, http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-data/2011-first-release/first-release--quality-assurance-and-methodology-papers/differences-between-2011-census-est-and-rolled-forward-pop-est.pdf

Openshaw S 1989 Making geodemographics more sophisticated. *Journal of the Market Research Society* 31: 111–31

Openshaw S 1995 Marketing spatial analysis: A review of prospects and technologies. In Longley P A and Clarke G P (eds) *GIS for Business and Service Planning*. New York, John Wiley and Sons: 227–46

Plewe B 2002 The nature of uncertainty in historical geographic information. *Transactions in GIS* 6: 431–56

Rao J N K 2003 *Small Area Estimation*. New York, John Wiley and Sons

Shelton N, Birkin M, and Dorling D 2006 Where not to live: A geo-demographic classification of mortality for England and Wales, 1981–2000. *Health and Place* 12: 557–64

Sleight P 2004 *Targeting Customers. How to Use Geodemographics and Lifestyle Data in Your Business*. Henley-on-Thames, UK, WARC

Slingsby A, Dykes J, and Wood J 2011 Exploring uncertainty in geodemographics with interactive graphics. *IEEE Transactions on Visualization and Computer Graphics* 17: 2384–91

Tzavidis N, Marchetti S, and Chambers R 2010 Robust estimation of small area means and quantiles. *Australian and New Zealand Journal of Statistics* 52: 167–86

Vickers D and Rees P 2007 Creating the UK National Statistics 2001 output areas classification. *Journal of the Royal Statistical Society* 170: 379–403

Vickers D, Rees P, and Birkin M 2005 *Creating the National Classification of Census Output Areas: Data, Methods and Results*. Leeds, University of Leeds, School of Geography Working Paper No. 05/2 (available at http://www.geog.leeds.ac.uk/wpapers/05-2.pdf)

VOA 2005 Council Tax Revaluation Wales 2005: An Overview. WWW document, http://www.voa.gov.uk/corporate/CouncilTax/cTWalesOverview.html