

Residential neighbourhood classification: an environmentally enhanced approach

Nigel Walford¹ and Richard Armitage²

1. Department of Geography and Geology, Kingston University, Penrhyn Road, Kingston upon Thames. KT1 2EE

Tel: +44 (0)20 8547 7512

Email: N.Walford@kingston.ac.uk

2. School of Science, Engineering and Environment, University of Salford, Salford. M4 5WT

Tel: +44 (0)161 295 5646

Email: R.P.Armitage@salford.ac.uk

Abstract

National small area classifications in Britain were first produced over 40 years ago using statistics from 1971 Population Census and have now become a regular feature of governmental, academic and commercial analysis of census information. These classifications aim to encapsulate the aggregate demographic and socio-economic character of small areas by means of a simple thumbnail description. However, these pen portraits often also refer to the environmental nature of the different types of area where people live, employing terms such as 'leafy suburb', 'industrial hinterland' or 'agricultural heartland'. This paper reports on research that aims to determine whether a set of environmental (land use) indicators are capable of discriminating between areas in a way that matches a 'standard' area classification derived from multivariate analysis of demographic and socio-economic statistics. The research assesses the impact of adding a set of environmental (land use) variables to a collection of Census variables on area classification using k-means clustering in two contrasting case study local authorities. The results reveal that clustering with and without the addition of land use variables produce partially overlapping (coincident) classifications of the small areas and certain of the land use variables are aligned with some area types.

Key words:

Geodemographics; residential environment; area classification; census statistics

1.0 Introduction

Where, or at what addresses, do people with particular combinations of demographic and socio-economic attributes live or, to put the question another way, do people with certain combinations of characteristics live in similar places? For nearly fifty years successive British Population Censuses have provided data for spatial units that have been analysed to tell us a lot about the characteristics of people living in relatively small areas (Harris, Sleight and Webber, 2005; Gale et al., 2016), but these accounts have rarely been extended to include discussion about the physical environment of these areas (Alexiou et al., 2016). Reference has been made to household environment, defined as its occupied dwelling unit, for which information about the number of rooms and type of residential property (detached, semi-detached, terraced, apartment dwelling, etc.), and access to or use of a discrete set of housing amenities (including sole or shared use of a bathroom/WC and central heating. Although differences in these housing characteristics do indicate something of the external physical environment of a household's dwelling unit, in many respects they are more strongly related to the occupants' socio-economic and to some extent demographic condition

(Alexiou et al., 2016; Singleton and Longley, 2015) than to the residential neighbourhood they inhabit. For example, low density occupancy of multi-roomed detached dwellings with access to multiple amenities is likely to indicate enhanced wealth compared with high density terraced housing with fewer rooms and sharing or lacking use of housing amenities.

These complex mixes of household and individual characteristics have traditionally been encapsulated by classificatory analyses (Webber and Craig 1976; Webber and Craig 1978; Craig 1985; Hodge and Monk 1991; Wallace and Denham 1996; Vickers and Rees, 2007; and Gale, 2014), including those developed for commercial purposes (e.g. Experian's MOSAIC and CACI's ACORN Systems), which incorporate little if any input data relating to peoples' lived environment. However, the thumbnail sketches penned to describe some of the 'neighbourhood types' frequently employ such inherently environmental terms as 'leafy suburb', 'inner city tenement', 'agricultural village' or 'spacious countryside' (Gale, 2014; Gale et al., 2016). These classifications are regarded as a satisfactory and sufficient means of finding out what types of people or household live in different places, which may be useful for commercial marketing and public planning and resource allocation purposes. However, arguably they are less successful for exploring and understanding the lived-experiences of people with respect to the physical environment of these places (Alexiou et al. 2016). Figure 1 highlights this issue with two Output Areas from the same local authority that were both classified as Prospering Suburbs (Supergroup 4) in the 2011 national area classification, yet contemporaneous aerial photographic images reveal that their residents experienced seemingly quite different physical environments. Area 1 had mixed land cover containing residential and industrial land, whereas area 2 was generally residential with some open space and trees.

[INSERT FIGURE 1]

A number of the 'social disorders' of British society are contentiously viewed as part of a cycle of deprivation and disadvantage (Rutter and Madge, 1976; Welshman, 2008). One element of this cycle relates to the type of area in which people live, very often expressed in terms of access to housing and standards of accommodation together with the socio-economic characteristics of the people living there (Orla, 2012). Indicators such as the number of persons per room, the availability of amenities or the number of dwellings per unit area reflect the condition of the housing itself, but the quality of the environment where people live and its interrelationship with housing conditions have received relatively little attention, although the potential of green open spaces to provide opportunities for exercise has recently been advanced in respect of policies to address health and wellbeing issues (Mitchell and Popham 2017; van de Berg et., 2015; WHO 2016; Wolch 2014). There is a range of environmental indicators that could be calculated including the proximity and juxtaposition of housing to other landscape features (e.g. industrial facilities, transport routes, heritage buildings, allotment gardens, parkland, woodland, fields and open space). These could be incorporated into an analysis of people's living space in a systematic fashion. The features and spaciousness of the environment in which people live may be as important as the physical condition of the housing itself in contributing to their sense of well-being.

The aim of this paper is to explore the feasibility and potential impact of incorporating environmental variables within neighbourhood classification and to compare the results obtained with those arising from the typical 'census only' approach. As an exploratory investigation it does not attempt to replicate a national, UK-wide geodemographic classification, but rather uses two case study local authorities with contrasting and complimentary characteristics within which the method and variables used in a recent national classification have been applied in preparation for a subsequent more extensive trial. It also introduces relatively simple environmental variables in the form of land use

area, whereas further extension will incorporate landscape metrics and visibility of environmental features in relation to residential building heights. The following section briefly reviews the study areas and examines the data and classificatory method used. The results of applying this method with and without environmental variables enable the impact of including the latter to be assessed by statistical and cartographic means in relation to whether areal units are in the same cluster and the extent to which cluster descriptions differ. The implications of the research in respect of its future direction and the potential for extending the scope of the next generation of neighbourhood classifications following the forthcoming 2021 Population Census are considered in the discussion and conclusion section.

2.0 Study area, data sources and methods

2.1 Study Area

Two English local authority (LA) districts, Colchester Borough Council (CBC) in north east Essex and Salford Metropolitan District (SMD) located to the west of Manchester City centre (see Figure 2) were chosen for testing the feasibility and impact of incorporating environmental variables into neighbourhood classification. A focus on two LAs at this stage of the investigation enabled some validation of the results in respect of the land use environmental variables and the outcomes of the cluster analysis. CBC is a local authority of some 33,300 ha containing a moderate sized provincial town at its core surrounded by three small towns (Tiptree, West Mersea and Wivenhoe), villages, hamlets and open countryside. Its population grew from 142,500 in 1991 to 173,600 in 2011 on account of local employment growth and it being a commutable distance from London. Approximately 66 percent of the district's resident population live in the core urban area leaving 15 percent in the three small towns and 19 percent in comparatively sparsely settled villages and hamlets. A wide range of housing types is represented within the main built-up area and the smaller settlements, including terraced, semi-detached and detached property and flats from various periods. SMD incorporates the City of Salford and the towns of Eccles, Worsley, Swinton, Walkden, Little Hulton, and Irlam. Salford's status as a city dates back to 1926, and the current Metropolitan District to 1974. In total SMD covers an area of 9,719 hectares. The population of Salford was 233,900 in 2011, showing a growth of around 17,000 since 2001, but still down on the population in 1991 which was 241,542. Salford is characterised by a range of housing types, including tower blocks and terraces in the eastern part of the SMD, and more semi-detached and detached housing to the west. While generally built up, or "urban", Salford is characterised by a number of areas of formal and informal greenspace.

[INSERT FIGURE 2]

These LAs were selected on the basis that they provided a contrast between northern and southern England, between relatively 'metropolitan urban' and 'provincial rural' localities, and because across both districts they included sufficient numbers of output areas classified in the eight Supergroups comprising the upper tier in the hierarchy of the 2011 national Output Area Classification (OAC) (see Table 1) (Gale, 2014; Gale et al., 2016). This was considered important in order to test the impact of adding environmental variables across the range of Supergroups. There were 1388 OAs in the two LAs: the most prevalent OAC Supergroups in Colchester were Urbanites and Suburbanites, which reflects its position as a medium sized provincial town within the settlement hierarchy, whereas Salford's dominant Supergroups were Constrained City Dwellers and Hard Pressed Living, together accounting for some 45 percent of the total. Combining the OAs in Colchester and Salford and comparing with the national distribution of the OAC Supergroups reveals that the study areas and national percentages were very similar for some Supergroups (Multicultural Metropolitans and Suburbanites), whereas in other cases they were very different (Rural Residents, Urbanites and Hard Pressed living). Overall the percentages are sufficiently similar to allow the two districts taken together to be considered numerically representative of the national profile of OAC Supergroups. However, this review

of the study areas' in respect of the OAC should only be regarded as contextual and indicative of their situation within a classification produced by analysing data nationally for all output areas. Our approach was to replicate the variable transformation and statistical techniques methods used in that analysis, first by using the same set of census attributes for the two study areas, to produce a "local OAC", and secondly by combining these census data with a selection of environmental variables.

[INSERT TABLE 1]

2.2 Data sources

The method chosen for testing the impact of adding environmental variables started from the procedure and analytical techniques used in the national 2011 OAC (Gale, 2014). As noted, OAC was selected as the methodology due to it being publicly available (Singleton and Longley, 2009), unlike the methodologies underpinning commercially produced classifications. Full details of the OAC methodology and the data used are freely available (<http://www.opengeodemographics.com/>) and these will be briefly reviewed before outlining the procedure for obtaining a set of environmental variables. Although selection of the study areas was based on comparison with the national OAC 2011 Supergroups, it would not have been appropriate simply to use the national OAC classification results for the two case study areas because these were produced in the context of analysing census statistics for all output areas across the UK (Singleton and Longley 2015). However, correspondence with OAC 2011 was achieved by replicating the analytical methods for output areas in the case study LAs, first using only the census counts and second with the addition of environmental variables. The OAC was based on 60 variables selected from a larger set of 2011 census counts that were determined following a series of preliminary experimental analyses with 167 variables (Gale, 2014). The 60 census variables (see Table 2), which covered five domains (demographics, housing, socio-economics, household composition and employment), were pre-processed in three ways before being subjected to k-means clustering. The raw count data were converted to percentages using appropriate denominators, then they were transformed using the inverse hyperbolic sine method to address the issues of outliers and variation in the spread of data values, and finally they were standardised using the range standardisation procedure to deal with the problem differences in the scale and magnitude between the variables, which can impact on clustering outcomes (ONS, 2015). Most neighbourhood classification systems are based on some form of cluster analysis. The conceptual starting point for cluster analysis is that the set of variables with respect to which the individual entities have been measured or quantified defines a multidimensional statistical space. In recent years K-means clustering has become the de facto standard approach, although earlier national classifications in the UK involved two-stages with principal components analysis being applied to the initial set of variables in order to generate a new set of orthogonal uncorrelated components (Craig, 1985a; Charlton et al., 1985; Hodge and Monk, 1991; Wallace and Denham, 1994). The scores or values of these components were calculated before cluster analyses, typically using hierarchical clustering Ward's error sum of squares as the measure of similarity for grouping individual spatial units. The K-means cluster analyses were carried out in SPSS (version 24) specifying a solution with eight clusters (matching the number of supergroups in OAC) and a maximum of 10 iterations.

[INSERT TABLE 2]

The characteristics of the physical environment can be measured in many ways. Research has shown that access to and the presence of greenspace can be a significant factor in the physical and mental wellbeing of urban residents (Ambrey and Fleming 2013; Ekkel and de Vries, 2017; Mitchell and Popham 2007, Wood et al, 2017). Deprivation has been linked to standards of housing and the physical environment (Kuffer et al., 2017). Land use can be used as a measure of the nature and quality of the physical environment (Krekel, Kolbe and Wüstermann 2016). A number of

different land use classifications were examined before the Generalised Land Use Database (GLUD) (Department for Communities and Local Government, 2009) classification was adopted. The GLUD classification comprises 10 classes of land use and was designed to classify the Ordnance Survey's MasterMap® topographical feature layer. The 10 classes are: Domestic Buildings, Domestic Gardens, Non-Domestic Buildings, Roads, Paths, Rail, Greenspace, Water, Other and Unclassified. GLUD was selected because of the potential to separate buildings into domestic and non-domestic, which other classification such as the National Land Use Database's (NLUD) land use classification does not. As GLUD is designed to be used with OS MasterMap® it cannot be regarded as Open, but the intention at this stage was to test the concept of including environmental variables in a neighbourhood classification rather than to derive a fully Open product.

The creation of the land use classification was based on the 2006 GLUD descriptors (Department for Communities and Local Government, 2009) for each class, which classifies MasterMap® features based on a combination of their Theme, Make and Descriptive Group attributes. A simple attribute-based selection combined the three MasterMap® attributes to identify the GLUD classes. However, in order to separate out the domestic and non-domestic buildings an additional three step approach was used. The first step involved classifying any buildings adjacent to an area of hardstanding of more than 300m² as non-domestic. Ideally the second step would use Address-Point® to identify buildings with a business address, but as this product was not available for this project, a substitute approach using adjacency to domestic gardens (one of the classes in the GLUD) was employed to identify domestic buildings. The third step involves identifying all buildings with a footprint larger than 1000 m² and classifying those as non-domestic. In the two study areas, no Unclassified land parcels were identified and so nine land use classes were used in the subsequent analysis. Salford and Colchester were processed separately, using the 2011 version of the OS MasterMap® (obtained from Digimap), and once the land use data had been created manual editing was used to correct misclassified polygons based on the authors' knowledge of the study areas and field checking. Due to the lack of availability of Address-Point®, a particular error that occurred was the misclassification of residential tower blocks (which have a large footprint and are usually not adjacent to gardens) as non-domestic buildings.

Analysis of the study areas shows that greenspace is the dominant land use in both LAs, although to a greater extent in Colchester than Salford (see Table 3). Gardens occupied the second highest proportion of land area, although the figure in Salford was more than double the one for Colchester. Water was also important in Colchester on account of the district having a coastline with the North Sea, associated estuaries and a large reservoir in the south of the Borough. Roads accounted for relatively more of Salford's area on account of its somewhat more urban setting. The distribution of these land use vector polygons across the two areas in Figure 3 emphasises the dominance of greenspace in CBC with a scattered mosaic of gardens away from the main urban areas. Greenspace prevailed to a similar extent in the south-west of Salford, although the greater concentration of road, rail and water routes accompanied by domestic/non-domestic buildings and gardens over much of the remainder of the authority contrasts with Colchester.

[INSERT TABLE 3]

[INSERT FIGURE 3]

The vector land use maps were intersected with the digital boundaries of the 2011 Output Areas for the two LAs. Output Areas were designed to maximise population homogeneity derived from socio-economic characteristics (Martin et al. 2001), and as such may not be regarded as appropriate for representing the spatial variation in environmental characteristics. However, it was felt that as many other Neighbourhood Classification are based on the Output Area

as the mapping unit, its use here would facilitate comparisons to be made. Problems with creating land use/land cover data for OAs are reported by Alexiou et al. (2016), who noted that in 2005 the Department for Communities and Local Government had issues when experimenting with the development of generalised land use data based on OAs. In those experiments the centroids of features were allocated to Output Areas. Here the actual area of the feature is used, with the areas of whole or parts of features being allocated to an OA. For each Output Area the proportions of the nine land use types were calculated using the total area of the OA as the denominator. This approach clearly does not take into account the potential spatial arrangement of the land use types within an OA, but it does provide an indication as to composition of the physical environment in the OA, and allows the problem demonstrated in Figure 1 to be investigated. It is the intention to investigate spatial configuration as a development of this work.

2.3 Examining the variability in the physical environment

The variability in the physical environment characteristics was examined using ordination. While ordination is usually used for data reduction it can also be applied as an exploratory method to examine the variability in a set of data. Here the nature of the variation in the land use variables was analysed using principal components analysis. The results (Table 4) showed that the first three PCA axes representing 88% of the variability in the environmental data. The first PCA axis, accounting for 65% (eigenvalue 0.1066) of the variability, was characterised by variation from OAs dominated by the Greenspace class to ones that might be defined as residential (dominated by Domestic Buildings and Gardens) (Figure 4 Table 4). The second axis, accounting for 18% (eigenvalue 0.03), is characterised by variation between “green” OAs (dominated by Greenspace, Gardens and Domestic Building) and “grey” OAs (ones containing more impermeable surfaces) (Table 4). The PCA analysis indicates that OAs in the study areas can be characterised in terms of degrees of physical characteristics of habitation, and therefore there is potential to exploit a combined analysis of the census and land use variables.

[INSERT TABLE 4]

[INSERT FIGURE 4]

2.4 Creation of combined classification

The land use variables were prepared, transformed and standardised in the same way as with the census variables. This pre-processing of the land use data was carried out to provide consistency across the full set of variables when k-means clustering was performed. It was also the case that each of the land use variables displayed similarly right-skewed distributions as observed in many of the census variables, which constitutes one of the main reasons for their transformation. The combined set of demographic, socio-economic and environmental variables were then analysed using the k-mean clustering methods described above to produce a second area classification. The following section first compares the membership of the two sets of clusters produced by the cluster analyses. The second part compares the characteristics of the clusters in these analyses. The OAs were categorised on a binary or dichotomous basis according to whether the census only and the census plus land use k-means clustering allocated them to the same or different ‘Supergroups’. OAs clustered together or apart were respectively assigned the values 1 (White shading) and 0 (Black shading). This shading helps to emphasise where tracts of OAs were allocated to the same or different groups by the separate cluster analyses.

3.0 Results

The area classifications produced with and without land use variables were compared in two ways: to determine the extent to which individual output areas across the two LAs were allocated to the same or different clusters in k-means

cluster analyses; and whether the environmental variables changed the characterisation of the clusters. The comparisons were carried out at the top level of the K-means clustering with eight clusters, the same number as in the national 2011 OAC, rather than drilling down to the Group and Sub-group levels.

3.1 Comparison of cluster membership

Over two-thirds of the 1388 OAs across the two districts were clustered together in six of the eight K-means clusters when the two cluster analyses were carried out. The census only clusters are identified by the digits 1 to 8 and the census plus land use clusters by the letters A to H. Starting from the census only clusters in the columns of Table 5, the rows of the census plus land use clusters have been ordered to maximise the counts along the diagonal. These are the OAs that were clustered together in both k-means cluster analyses. Over 90 percent of the OAs in census only clusters 1, 3 and 8 remained together and paired with census plus land use clusters A, E and B; and the figure was over 75 percent in clusters 4 and 6 (linked with C and F). Only 60.0 percent remained together in census only cluster 2 (with H), but this lower percentage may be regarded as an artefact of the small number of OAs (5) with the movement of two of these units accounting for a 40.0 percent change in the total. The greatest disruption to cluster membership occurred in census only cluster 5 with substantial numbers distributed to census plus land use clusters C, D and F. These changes of cluster membership are shown even more starkly in Figure 5 where each square represents one OA: images in the left column show the OAs in the eight census only groups and the coloured squares with letters in the righthand column denote the OAs that moved to a different group as a results of adding in the land use variables. It is clear that census only group 5 is the one most impacted by the addition of land use variables.

[INSERT TABLE 5]

[INSERT FIGURE 5]

3.2 Comparison of cluster characteristics

Both sets of clusters produced by the k-means cluster analyses need to be described independently of the national OAC 2011 classification, despite some of the national Supergroups being well represented (see Table 1), because they relate to the subset of 1388 OAs in the two LAs. One of the now standard ways of visualising the relative importance of the different variables in a neighbourhood classification is by means of radar or radial charts, although these can be difficult to interpret if there are substantial numbers of variables. Figure 6 employs this device but only includes those variables where the difference between the census only and census plus land cover cluster standardised means were greater than +0.15 or less than -0.15 from the overall standardised mean. These upper and lower thresholds impacted on the census variables for clusters 2 and 5 and for the remainder adding land use variables into the k-means only produced a change in the standardised mean within the range +0.15 to -0.15. Several of the land use variables featured beyond this range in the census plus land use clustering. Using the paired clusters from the two k means clustering given in Tables 5, Table 6 shows the effect of including the land use variables. Cluster pairs 2, H and 5, D (census and census plus land use respectively) had difference in census variables beyond the range +0.15 to -0.15, where the other six cluster pairs only had a difference in standardised mean in respect of some of the land use variables. The addition of land use variables had the most impact on census cluster 5 in respect of influencing the census variables where greater difference in standardised mean occurred.

[INSERT FIGURE 6]

[INSERT TABLE 6]

Figure 7 compares the spatial distribution of the census only and census plus land use neighbourhood classifications

for Colchester and Salford. The same shading schemes have been used for both sets of maps without taking into account the pairings reflected along the diagonal of Table 5. Both local authorities had reasonably extensive swathes of land covered by a few of the groups and there is some evidence of spatial 'clumping'. Differences in the essential character of the two local authorities means that the same groups were unlikely to dominate the patterns produced by either of the K-means cluster analyses. For example, there were substantial regions covered by census only group 1 in both Colchester and Salford; whereas census plus land use group F covered a large area of Colchester and group B spread widely across Salford. However, mixing up of cluster membership arising from the addition of environmental variables means that a simple visual comparison of the results of the two cluster analyses is misleading.

[INSERT FIGURE 7]

Figure 8 helps to clarify the differences in cluster membership produced by the K-means clustering of the two sets of variables. Unshaded OAs are those which were clustered together in the same group along the diagonal in Table 5, whereas those shaded black were clustered in different groups. Visual inspection of the spatial pattern produced reveals some clumping of OAs in these categories. This binary categorisation of the OAs has also been examined using the Mann Whitney non-parametric test in respect of four variables: total population, area (ha), and scores (values) on the first and second components of the PCA outlined earlier (Figure 4) (Table 4). The first component distinguished between the OAs in respect of whether the green space was of a domestic or communal nature, whereas component two showed a similar separation between green and non-green space. The statistical results (Table 7) indicate that the means for the together and separate groups of OAs were similar in respect of area and scores on components 1 and 2, but there was a significant difference ($P=0.010$) for total population.

[INSERT FIGURE 8]

[INSERT TABLE 7]

4.0 Discussion and Conclusions

This paper has investigated the feasibility of including environmental (land use) variables with 'standard' demographic and socio-economic census counts in neighbourhood classifications and assessed the impact of adding the land use variables in comparison with results based solely on census data input. The comparative analysis has been based on replicating the data pre-processing and statistical methods used in a previous national area classification, in particular the 2011 Output Area Classification (Gale et al., 2017). By basing the neighbourhood classifications on both types of the data, it may be possible to provide a more realistic, evidence-based assessment of differentiation in the nature of the areas in which people live. This could potentially contribute to our understanding of the influence of environment and land use on people's lives, their health and wellbeing and interactions with the places they inhabit. Although the analysis presented in this paper focuses on two contrasting local authorities, the increased availability of small format digital aerial imagery and desktop image processing makes derivation of simple environmental indicators across large areas a realistic possibility.

The results indicate that adding the land use variables had a differential impact on the census only clusters, with substantial numbers of OAs being clustered together under both analyses, in some cases (census only clusters 1, 3 and 8) in excess of 90% of the OAs in the census only clusters were clustered together again when land use variables were included. In contrast two of the census only clusters experienced dispersal of their OAs into several clusters

when land use variables were added. Although one of these (census only cluster 2) had only five OAs, the other census only cluster 5 had 166 OAs spread unequally across four census plus land use clusters (C, D, E and F). Carrying out a statistical analysis on the difference in total population, area and scores on principal components 1 and 2 in relation to whether OAs were clustered together or separately by the two clusterings suggests that difference was significant only in the case of total population,

The analysis has incorporated environmental indicators in the form of area of nine types of land use into the area classification in respect of the relative proportions of different land uses derived from the Generalised Land Use Database in conjunction with the Ordnance Survey's MasterMap® topographical feature layer. There is support from the literature that land use constitutes an important source of variation in people's perception of the environmental character of urban areas (Dennis et al. 2018; Palmer, 2004). However, arguably there are other environment features and landscape metrics (Alexiou et al 2016; Kuethe, 2012) not reflected in land use on its own that could be taken into account in order to augment the analyses undertaken here. For example, consideration could be given to the intervisibility, proximity and adjacency of 'desirable' and 'objectionable' environmental features, although this would potentially introduce subjective judgements over which features are desirable and which are not. Further development of the research presented here could investigate whether differences between the classifications are reflected in contrasts between the health and wellbeing of people living in the output areas, for example to see if the census plus land use classification discriminates better between the health outcomes of individuals.

Acknowledgements

Office for National Statistics, 2011 Census: Aggregate data (England and Wales) [computer file]. UK Data Service Census Support. Downloaded from: <http://infuse.mimas.ac.uk> . These data are licensed under the terms of the Open Government Licence (<http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2>). Office for National Statistics, 2011 Census: Aggregate data (England and Wales) [computer file]. UK Data Service Census Support. Downloaded from: <http://infuse.mimas.ac.uk> . These data are licensed under the terms of the Open Government Licence (<http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2>).

This work is based on data provided through EDINA UKBORDERS with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown. It also uses data provided under license by the Ordnance Survey.

References

- Alexiou, A., Singleton, A. and Longley, P.A. (2016) A classification of multidimensional Open Data for urban morphology. *Built Environment*, 42: 382-395. <https://doi.org/10.2148/benv.42.3.382>.
- Ambrey, C.L. Fleming, C.M. (2013) Public green space and life satisfaction in urban Australia. *Urban Studies*, 51: 1290-1321. <https://doi.org/10.1177/0042098013494417>.
- Andersson, R. and BråmÅ, Å (2004) Selective migration in Swedish distressed neighbourhoods: can area-based urban policies counteract segregation processes? *Housing Studies*, 19: 517-539. <https://doi.org/10.1080/0267303042000221945>
- Department for Communities and Local Government (2009) *Generalised Land Use Database – Update 2006*. Communities and Local Government Publications. ISBN: 978 1 4098 1119 0

- Charlton, M., Openshaw, S. and Wymer, C. (1985) Some new classifications of census Enumeration Districts in Britain: a poor man's ACORN, *Journal of Economic and Social Measurement*, 13: 69-96. Doi: <http://dx.doi.org/10.1080/13658816.2011.554838>
- Craig, J. (1985) *1981 Socio-economic Classification of Local and Health Authorities of Great Britain*, SMPS 48, London: HMSO.
- Dennis, M., Barlow, D., Cavan, G., Cook, P., Gilchrist, A., Handley, J., James, P., Thompson, J., Tzoulas, K., Wheater, P. and Lindley, S. (2018) Mapping urban green infrastructure: a novel landscape-based approach to incorporating land-use and land-cover in the mapping of human-dominated systems. *Land*, 7: 1-25. <https://doi.org/10.3390/land7010017>
- Ekkel, E.D. and de Vries, S. (2017) Nearby green space and human health: Evaluating accessibility metrics. *Landscape and Urban Planning*, 157, 214-220. Doi: <https://doi.org/10.1016/j.landurbplan.2016.06.008>
- Doyle, O. (2012) *Breaking the cycle of deprivation: an experimental evaluation of an early childhood intervention*, Working Paper Series, No. 12/17, University College Dublin, UCD Centre for Economic Research, Dublin Available at: <http://hdl.handle.net/10419/72225>. [Accessed: 19 July 2019].
- Gale, C.G. (2014) *Creating an open geodemographic classification using the UK census of the population*. Doctoral thesis, University College London, Department of Geography, London, UK.
- Gale C.G, Singleton, A. D., Bates, A.G. and Longley P.A. (2016) Creating the 2011 area classification for output areas (2011 OAC). *Journal of Spatial Information Science*, 12, 1-27. Doi: <http://dx.doi.org/10.5311/JOSIS.2016.12.232>
- Harris, R., Sleight, P. and Webber, R. (2005) *Geodemographics: GIS and Neighbourhood Targeting*. Wiley-Blackwell.
- Hodge, I. and Monk, S. (1991) *In Search of a Rural Economy: Patterns and Differentiation in Non- Metropolitan England*. Land Economy Monograph No.20, University of Cambridge.
- Kuethe, T.H. (2012) Spatial Fragmentation and the Value of Residential Housing. *Land Economics*, 88: 16.27. Doi: <https://doi.org/10.3368/le.88.1.16>
- Kuffer, M., Pfeffer, K., Sliuzas, R., Baud, I. and van Maarseveen, M. (2017) Capturing the Diversity of Deprived Areas with Image-Based Features: The Case of Mumbai. *Remote Sensing*, 9, 384-407. Doi: <https://doi.org/10.3390/rs9040384>
- Martin, D., Nolan, A. and Trammer, M. (2001) The application of zone-design methodology in the 2001 UK Census. *Environment and Planning A*, 33: 1949-1962.
- Mitchell, R. and Popham, F. (2007) Greenspace, urbanity and health: relationships in England. *Journal of Epidemiological Community Health*, 61: 681–683. Doi: <http://dx.doi.org/10.1136/jech.2006.053553>
- National Statistics (2004) National Statistics 2001 Area Classification for Statistical Wards. http://www.statistics.gov.uk/about/methodology_by_theme/area_classification/wards/default.asp
- Palmer, J.F. (2004) Using spatial metrics to predict scenic perception in a changing landscape: Dennis, Massachusetts. *Landscape and Urban Planning*, 69: 201-218. Doi: <https://doi.org/10.1016/j.landurbplan.2003.08.010>
- Rutter, M. and Madge, N. (1976) *Cycles of Disadvantage: A Review of Research*. Heinemann: London.

- Singleton, A.D. and Longley, P.A. (2015) The internal structure of Greater London: a comparison of national and regional geodemographic models. *Geo*, 2: 69-87. Doi: <http://dx.doi.org/10.1002/geo2.7>
- Singleton, A.D. and Longley, P.A. (2009) Geodemographics, visualisation, and social networks in applied geography. *Applied Geography*, 29: 289-298. <https://doi.org/10.1016/j.apgeog.2008.10.006>.
- Van de Berg, M., Wendel-Vos, W., van Poppel, M., Kemper, H., van Mechelen, W. and Maas, J. (2015) Health benefits of green spaces in the living environment: a systematic review of epidemiological studies. *Urban forestry and Urban Greening*, 14, 806-816. Doi: <https://doi.org/10.1016/j.ufug.2015.07.008>
- Vickers, D. and Rees, P. (2007) Creating the UK National Statistics 2001 output area classification. *Journal of the Royal Statistical Society: Series A*, 170: 379-403. Doi: <https://doi.org/10.1111/j.1467-985X.2007.00466.x>
- Wallace, M. and Denham, C. (1994) *The ONS Classification of Local and Health Authorities of Great Britain*, SMPS 59, London: HMSO.
- Webber, R. and Craig, J. (1976) Which local authorities are alike? *Population Trends*, 5:13-19.
- Webber, R. and Craig, J. (1978) *A Socio-economic Classification of Local and Health Authorities*, SMPS 35, London: HMSO.
- Welshman, J. (2008) The cycle of deprivation: myths and misconceptions. *Children and Society*, 22:75-85. Doi: <https://doi.org/10.1111/j.1099-0860.2007.00088.x>
- WHO (2016) *Urban green spaces and health: a review of evidence*. Copenhagen: WHO Regional Office for Europe,
- Wolch, J.R., Byrne, J. and Newell, J.P. (2014) Urban green space, public health, and environmental justice: The challenge of making cities 'just green enough'. *Landscape and Urban Planning*, 125, 234-244. Doi: <https://doi.org/10.1016/j.landurbplan.2014.01.017>
- Wood, L., Hooper, P., Foster, S. and Bull, F. (2017) Public green spaces and positive mental health – investigating the relationship between access, quantity and types of parks and mental wellbeing. *Health and Place*, 48, 63-71. Doi: <https://doi.org/10.1016/j.healthplace.2017.09.002>

Output Area Classification (OAC) Supergroups	Colchester	Salford	Colchester and Salford combined	UK OAC
Rural residents	8.2	0.3	3.9	11.8
Cosmopolitans	6.3	8.1	7.9	5.7
Ethnicity Central	1.4	5.2	3.9	5.1
Multicultural Metropolitans	7.9	10.3	10.0	10.1
Urbanities	30.7	15.5	23.6	16.7
Suburbanites	27.9	15.4	22.3	20.2
Constrained City Dwellers	7.5	20.4	16.1	11.7
Hard Pressed Living	10.1	24.8	12.3	18.9
	100%	100%	100%	100%

Table 1: Representation of Colchester and Salford output areas in the eight 2011 Output Area Classification (OAC) Supergroups.

Note: Figures are percentages of output areas in each local authority, for the two authorities combined and for the UK overall in OAC 2011 Groups.

Source: www.opengeodemographics.com/

Census counts and land use variable identifiers and descriptions

k001	% Persons aged 0 to 4	k036	% Persons providing unpaid care
k002	% Persons aged 5 to 14	k037	% Persons aged over 16 whose highest level of qualification is Level 1, Level 2 or Apprenticeship
k003	% Persons aged 25 to 44	k038	% Persons aged over 16 whose highest level of qualification is Level 3 qualifications
k004	% Persons aged 45 to 64	k039	% Persons aged over 16 whose highest level of qualification is Level 4 qualifications and above
k005	% Persons aged 65 to 89	k040	% Persons aged over 16 who are schoolchildren or full-time students
k006	% Persons aged 90 and over	k041	% Households with two or more cars or vans
k007	Number of persons per hectare	k042	% Persons aged between 16 and 74 who use public transport to get to work
k008	% Persons living in a communal establishment	k043	% Persons aged between 16 and 74 who use private transport to get to work
k009	% Persons aged over 16 who are single	k044	% Persons aged between 16 and 74 who walk, cycle or use an alternative method to get to work
k010	% Persons aged over 16 who are married or in a registered same-sex civil partnership	k045	% Persons aged between 16 and 74 who are unemployed
k011	% Persons aged over 16 who are divorced or separated	k046	% Employed persons aged between 16 and 74 who work part-time
k012	% Persons who are white	k047	% Employed persons aged between 16 and 74 who work full-time
k013	% Persons who have mixed ethnicity or are from multiple ethnic groups	k048	% Employed persons aged between 16 and 74 who work in the agriculture, forestry or fishing industries
k014	% Persons who are Asian/Asian British: Indian	k049	% Employed persons aged between 16 and 74 who work in the mining, quarrying or construction industries
k015	% Persons who are Asian/Asian British: Pakistani	k050	% Employed persons aged between 16 and 74 who work in the manufacturing industry
k016	% Persons who are Asian/Asian British: Bangladeshi	k051	% Employed persons aged between 16 and 74 who work in the energy, water or air conditioning supply industries

k017	% Persons who are Asian/Asian British: Chinese and Other	k052	% Employed persons aged between 16 and 74 who work in the wholesale and retail trade; repair of motor vehicles and motor cycles industries
k018	% Persons who are Black/African/Caribbean/Black British	k053	% Employed persons aged between 16 and 74 who work in the transport or storage industries
k019	% Persons who are Arab or from other ethnic groups	k054	% Employed persons aged between 16 and 74 who work in the accommodation or food service activities industries
k020	% Persons whose country of birth is the United Kingdom or Ireland	k055	% Employed persons aged between 16 and 74 who work in the information and communication or professional, scientific and technical activities industries
k021	% Persons whose country of birth is in the old EU (pre-2004 accession countries)	k056	% Employed persons aged between 16 and 74 who work in the financial, insurance or real estate industries
k022	% Persons whose country of birth is in the new EU (post 2004 accession countries)	k057	% Employed persons aged between 16 and 74 who work in the administrative or support service activities industries
k023	% Persons whose main language is not English and they cannot speak English well or at all	k058	% Employed persons aged between 16 and 74 who work in the in public administration or defence; compulsory social security industries
k024	% Households with no children	k059	% Employed persons aged between 16 and 74 who work in the education sector
k025	% Households with non-dependent children	k060	% Employed persons aged between 16 and 74 who work in the human health and social work activities industries
k026	% Households with full-time students	Land use variables	
k027	% Households who live in a detached house or bungalow	k061	% Domestic
k028	% Households who live in a semi-detached house or bungalow	k062	% Gardens
k029	% Households who live in a terrace or end-terrace house	k063	% Greenspace
k030	% Households who live in a flat	k064	% Non-domestic
k031	% Households who own or have shared ownership of property	k065	% Other
k032	% Households who are private renting	k066	% Paths
k033	% Households who are social renting	k067	% Rail
k034	% Households who have one fewer or less rooms than required	k068	% Roads

k035 Individuals day-to-day activities limited a lot or a little (Standardised Illness Ratio) k069 % Water

Table 2: Code identifiers and descriptions of census and land use variable.

	Colchester and		
	Colchester	Salford	Salford Combined
Domestic Buildings	1.5	5.3	2.5
Gardens	6.4	14.1	8.4
Greenspace	77.4	59.5	72.8
Non-Domestic Buildings	0.8	2.1	1.1
Other	4.1	7.3	4.9
Paths	0.2	0.6	0.3
Rail	0.2	0.5	0.3
Roads	3.5	9.2	4.9
Water	5.9	1.4	4.7
	100.0%	100.0%	100.0%

Table 3: Comparison of Generalised Land Use Database land use categories in Colchester and Salford areas.

Note: Figures are percentages of each local authority and their combined area in the nine Generalised Land Use Database categories.

	PCA 1	PCA 2	PCA 3
Domestic Buildings	0.71937	0.23206	-0.22123
Gardens	0.69764	0.65855	0.26819
Greenspace	-0.98678	0.14865	-0.00667
Non-Domestic Buildings	0.13416	-0.49188	-0.08295
Other	0.17737	-0.87484	0.35532
Paths	0.10776	-0.10169	-0.16067
Rail	0.17791	-0.24072	0.00720
Roads	0.73439	-0.00796	-0.63560
Water	-0.03338	-0.22839	0.05024

Table 4: Component loadings for the first three PCA axes expressed as correlations.

	Census only	Total

	1	2	3	4	5	6	7	8	
A	93.1	0	0	0	0	3.5	0	0	308
H	0	60.0	0	0	0	0	0	0	3
E	0	0	94.7	0	0	0	0.6	4.0	78
C	0	0	0	80.0	18.7	0	0	3.0	58
D	6.0	20.0	0	0	33.7	2.2	23.0	1.0	160
F	0.6	20.0	0	20.0	45.2	75.5	0.9	0	364
G	0.3	0	0	0	2.4	18.8	68.1	1.0	297
B	0	0	5.3	0	0	0	7.4	91.1	120
Total	317	5	76	30	166	367	326	101	1388

Table 5: Comparison of Colchester and Salford output areas K-means cluster membership (using eight 'Supergroups') (percentages).

Matched clusters	Variables with difference in standardised means beyond +0.15 to -0.15 range
1. Census A. Census and environment	Land use variables - domestic, gardens, greenspace, non-domestic, other, paths and roads Census variables - nil
2. Census H. Census and environment	Land use variables - domestic, gardens, greenspace, non-domestic, other, paths and roads Census variables - % Persons aged 65 to 89%, % Persons who are Black/African/Caribbean/Black British, % Persons aged between 16 and 74 who use private transport to get to work, % Employed persons aged between 16 and 74 who work in the information and communication or professional, scientific and technical activities industries and % Employed persons aged between 16 and 74 who work in the human health and social work activities industries
3. Census E. Census and environment	Land use variables - domestic, gardens, greenspace, non-domestic, other, paths and roads Census variables - nil
4. Census C. Census and environment	Land use variables - domestic, gardens, greenspace, non-domestic, other, paths, roads and water Census variables - nil
5. Census D. Census and environment	Land uses - domestic, gardens, greenspace, non-domestic, paths and roads Census variables - % Persons who have mixed ethnicity or are from multiple ethnic groups, % Persons who are Asian/Asian British: Chinese and Other, % Households who live in a detached house or bungalow, % Households who are private renting, % Persons aged over 16 whose highest level of qualification is Level 4 qualifications and above, % Persons aged between 16 and 74 who are unemployed, % Employed persons aged between 16 and 74 who work in the information and communication or professional, scientific and technical activities industries, % Employed persons aged between 16 and 74 who work in the financial, insurance or real estate industries and % Employed persons aged between 16 and 74 who work in the education sector.
6. Census F. Census and environment	Land use variables - domestic, gardens, greenspace, non-domestic, other, paths and roads Census variables - nil
7. Census G. Census and environment	Land use variables -domestic, gardens, greenspace, non-domestic, other, paths, roads and water Census variables - nil
8. Census B. Census and environment	Land use variables - domestic, gardens, greenspace, non-domestic, other, paths and roads Census variables - nil

Table 6: Impact of adding land use variables to k means cluster analysis of Colchester and Salford output areas according to the pairings shown in Table 5.

	Area (ha)	Population	PCA1 Score	PCA2 Score	Total
Clustered separately	45.12	213.1	-0.0760	-0.0218	347
Clustered together	27.44	230.5	0.0253	0.0073	1041
Significance (Mann Whitney)	0.058	0.010	0.187	0.796	

Table 7: Comparison of census only and census plus land use k-means clustering outcome (together or separate) according to total population, area and environmental ordination.



(a) Presence of large industrial development in Output Area



(b) Low rise buildings and greenspaces in Output Area

Figure 1: Two output Areas in Salford classified as Prosperous Suburbs but with aerial photography indicating obvious differences in their physical environments.

Sources: Aerial photographs -Landmap (2014): Landmap: Optical Earth Observation Collection. NERC Earth Observation Data Centre, 2019. Boundary data - Office for National Statistics (2011): Census aggregate data (Edition: May 2011). UK Data Service. DOI: <http://dx.doi.org/10.5257/census/aggregate-2001-2>



Figure 2: Location of Colchester and Salford Local Authorities in England.

Source: Ordnance Survey, Edina. © Crown copyright and/or database right 2019 OS.

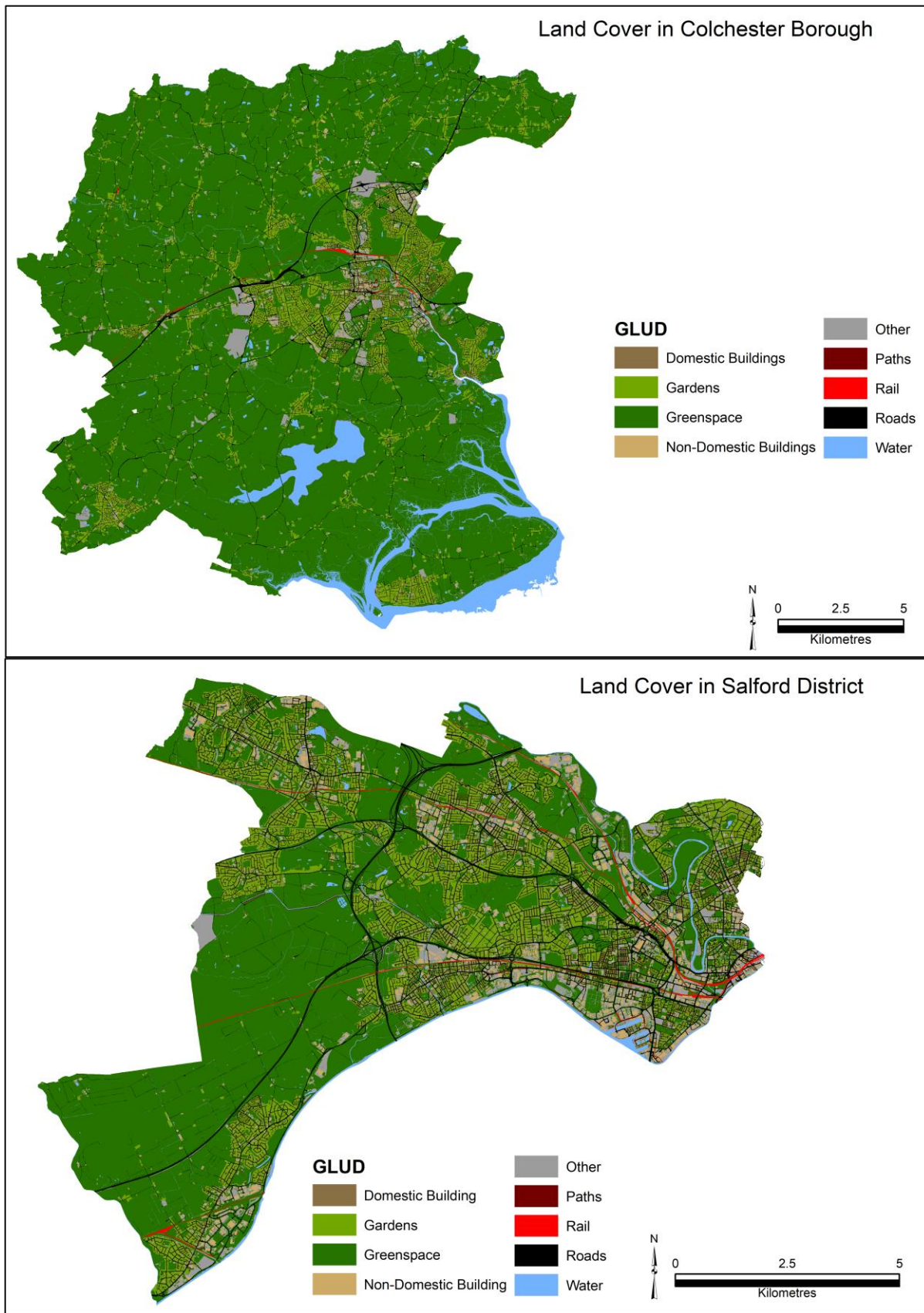


Figure 3: Vector land use polygons in Colchester and Salford Local Authorities derived from the Generalised Land Use Database.

Source: GLUD, Department for Communities and Local Government, 2009; Ordnance Survey, Edina. © Crown copyright and/or database right 2019 OS. Note: output area boundaries are not shown in order to maximise the visibility of small land use parcels.

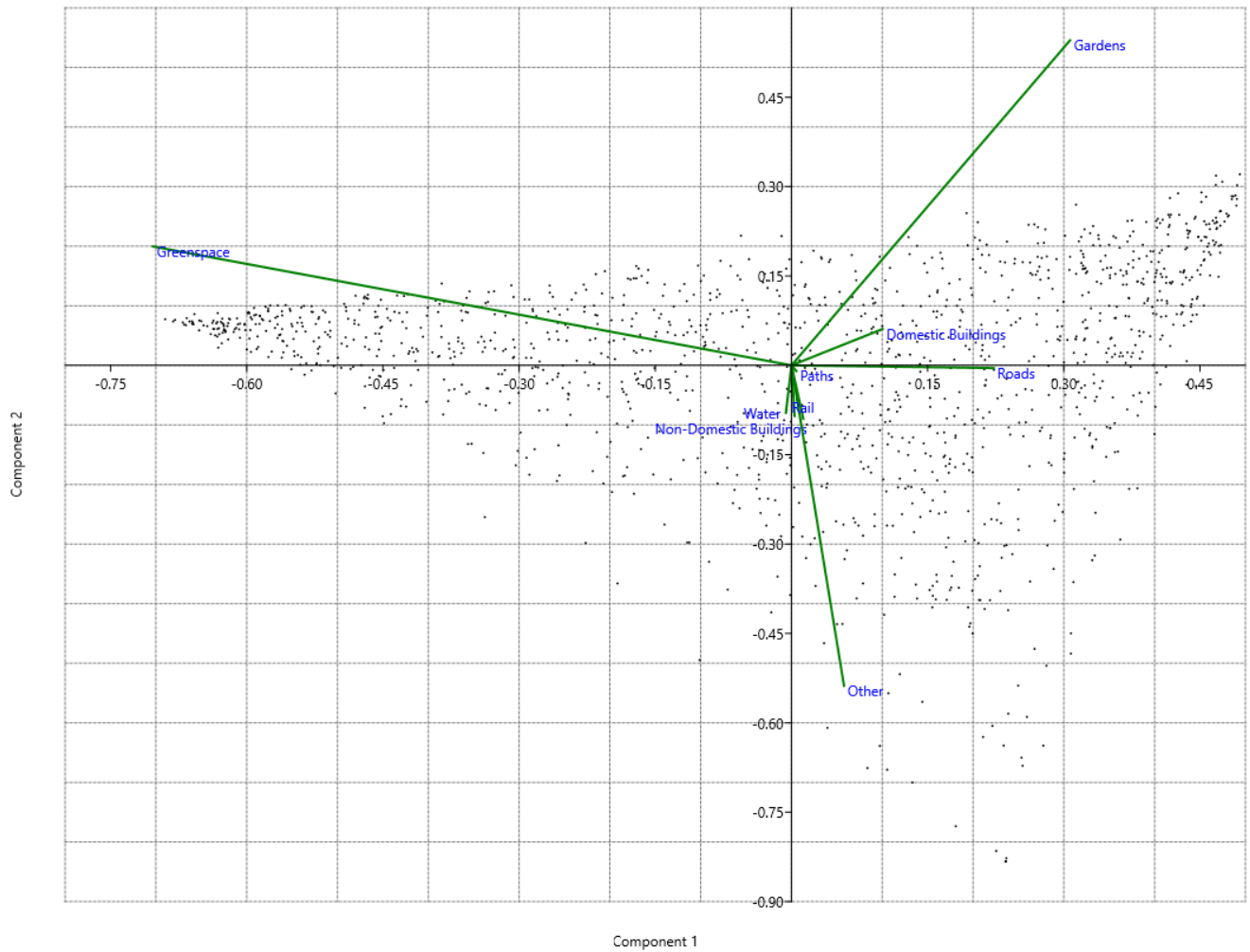


Figure 4: PCA biplot showing the first and second axes. The green lines represent the loading of each variable against the two axes. The points represent the OAs against each axis.

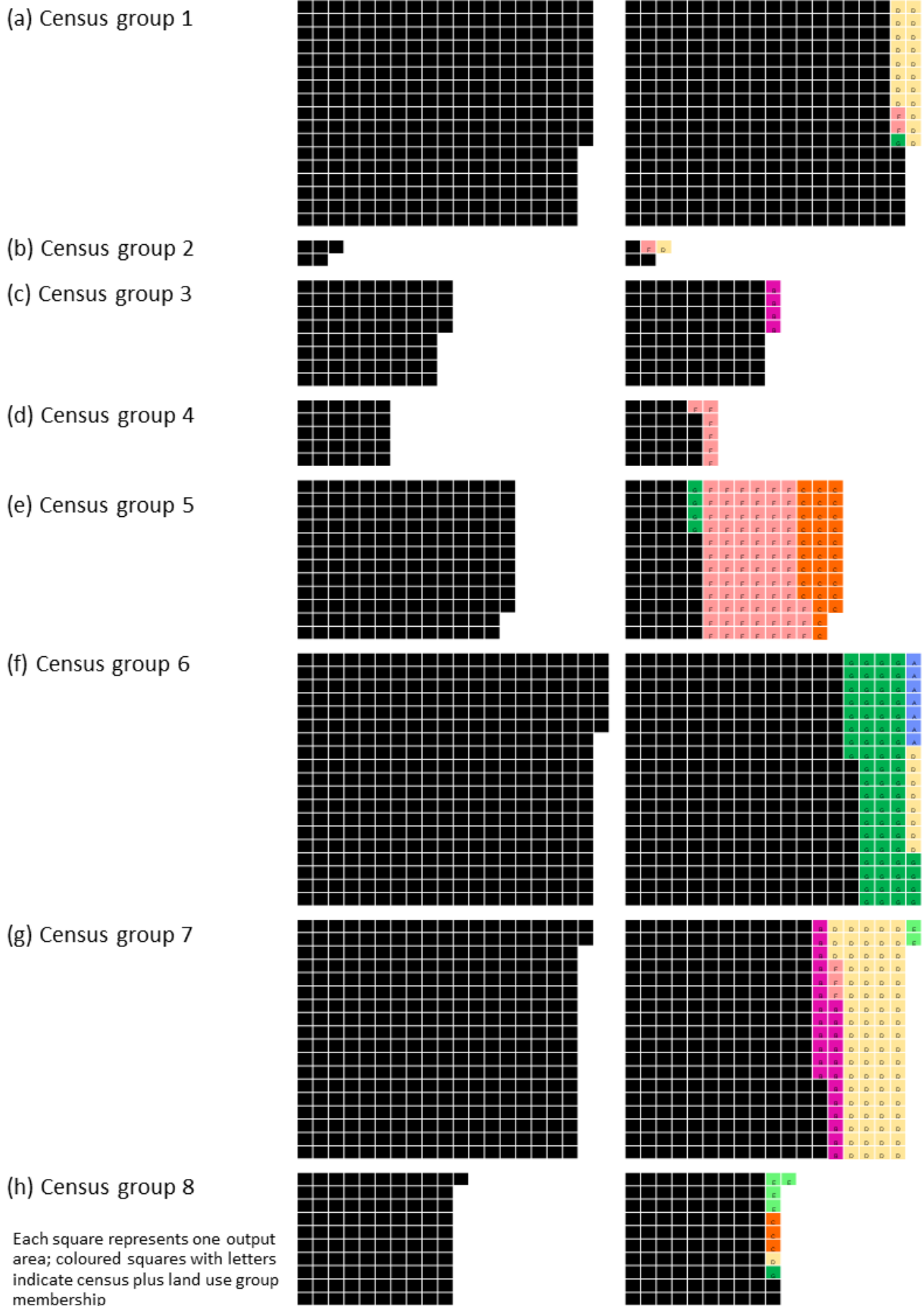
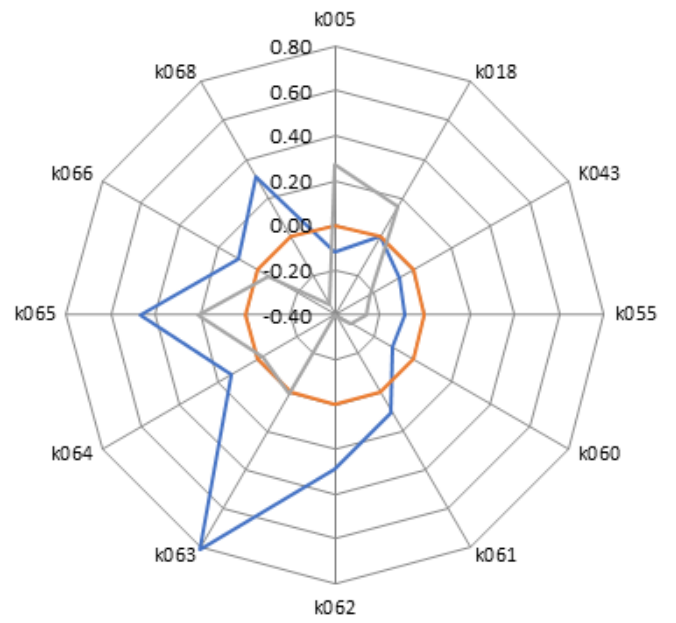
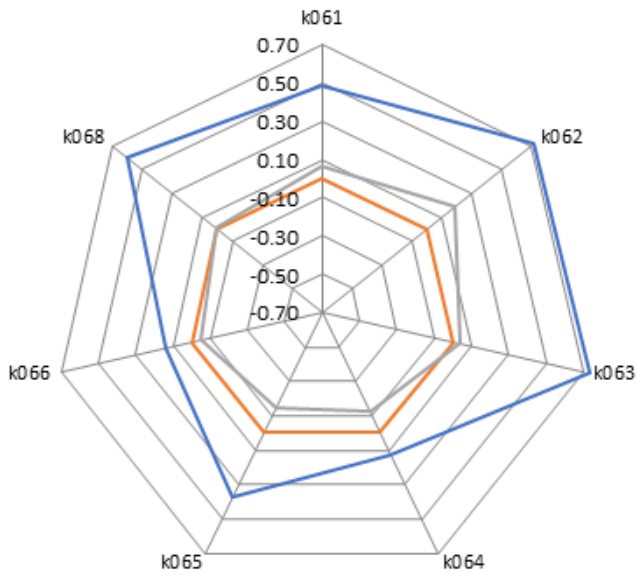


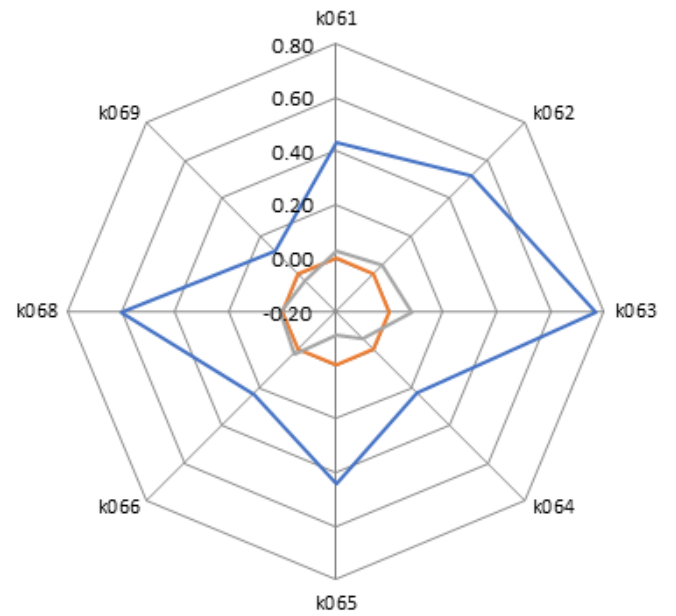
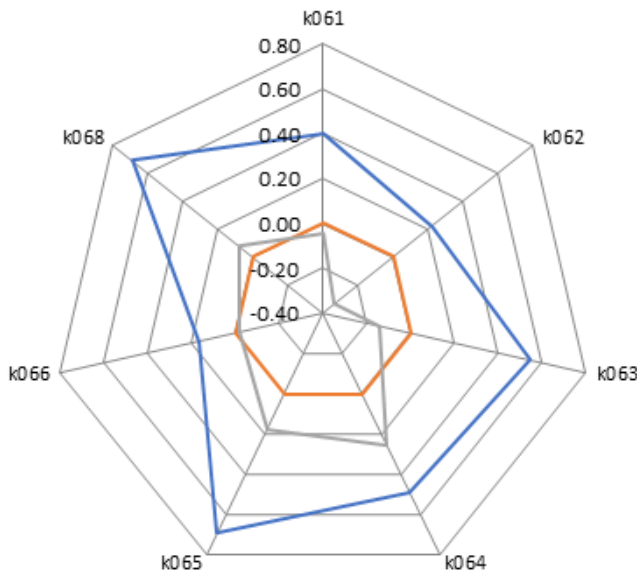
Figure 5: Representation of impact on cluster analysis groups of adding land use variables to census variables across Colchester and Salford districts.

Note: Each square represents one OA: images on the left represent OAs in census only clusters and the coloured squares in those on the right denote OAs that were clustered differently when land used variables were added.



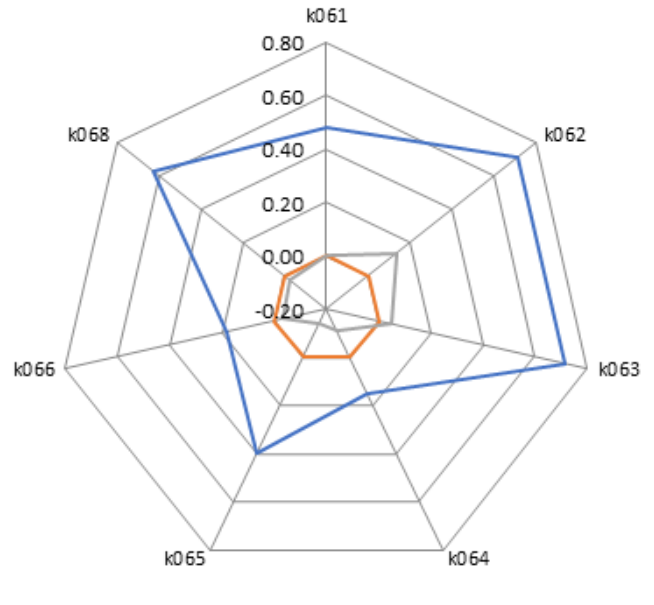
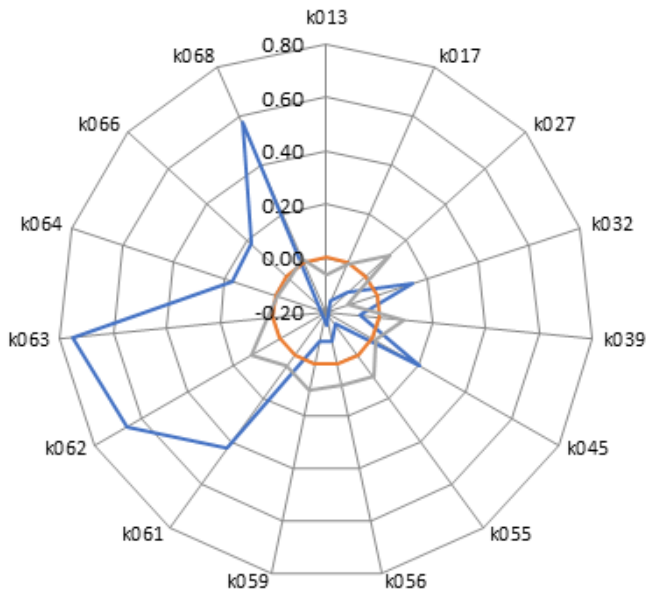
— Groups census only 1
 — Overall mean
 — Groups census and environment A

— Groups census only 2
 — Overall mean
 — Groups census and environment H



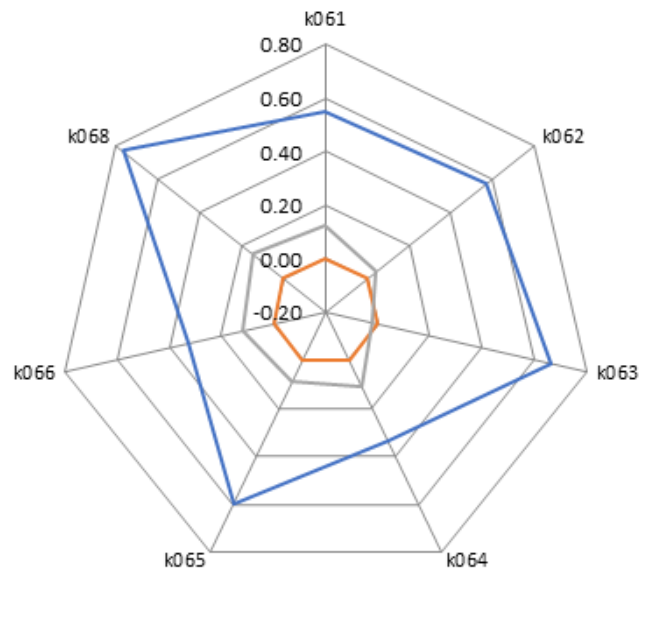
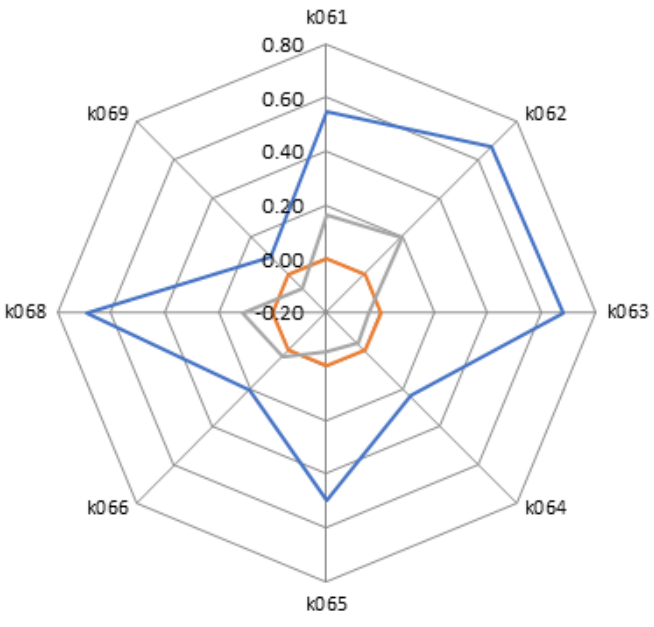
— Groups census only 3
 — Overall mean
 — Groups census and environment E

— Groups census only 4
 — Overall mean
 — Groups census and environment C



— Groups census only 5
 — Overall mean
 — Groups census and environment D

— Groups census only 6
 — Overall mean
 — Groups census and environment F

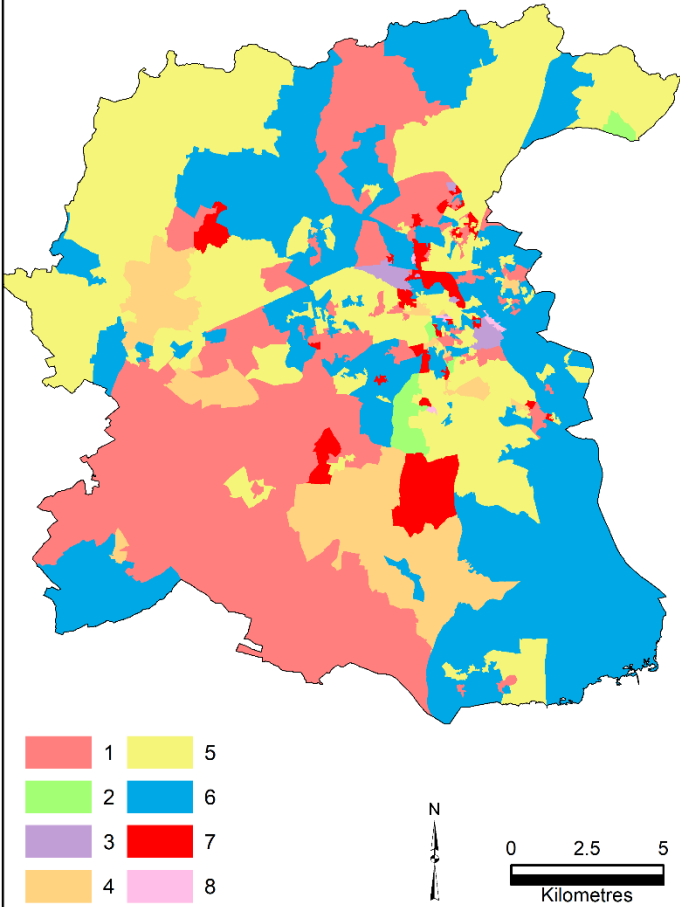


— Groups census only 7
 — Overall mean
 — Groups census and environment G

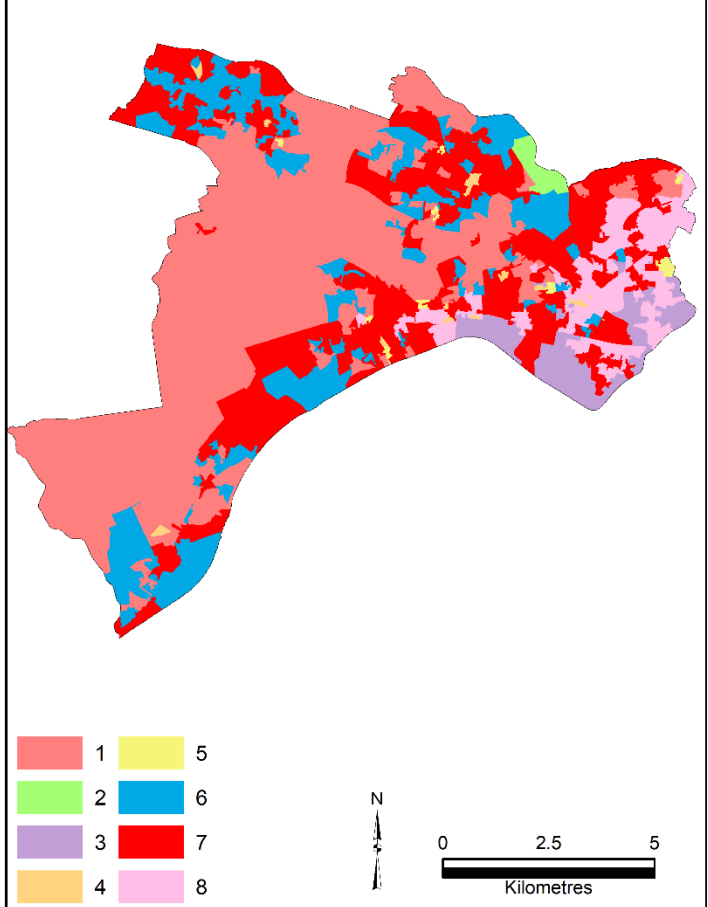
— Groups census only 8
 — Overall mean
 — Groups census and environment B

Figure 6: Radar charts of census only and census plus land use classifications in Colchester and Salford districts.

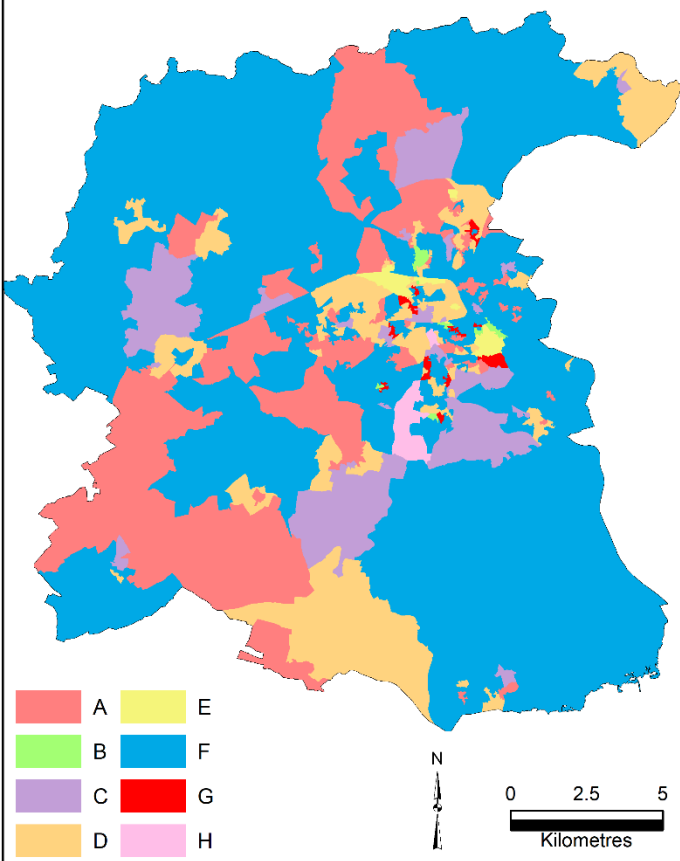
Colchester groups based on census variables only



Salford groups based on census variables only



Colchester groups based on census and environment variables



Salford groups based on census and environment variables

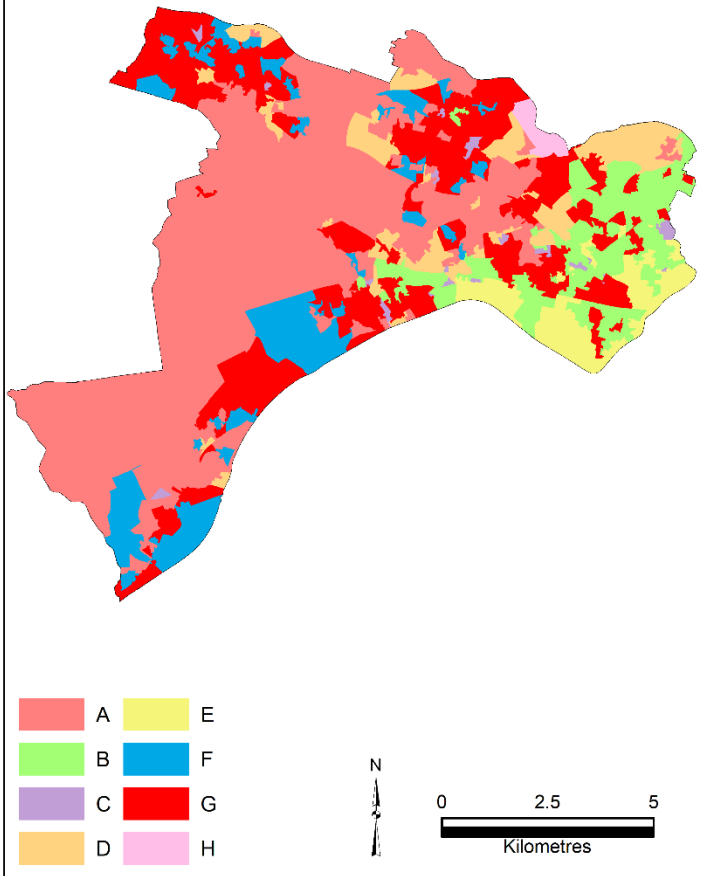
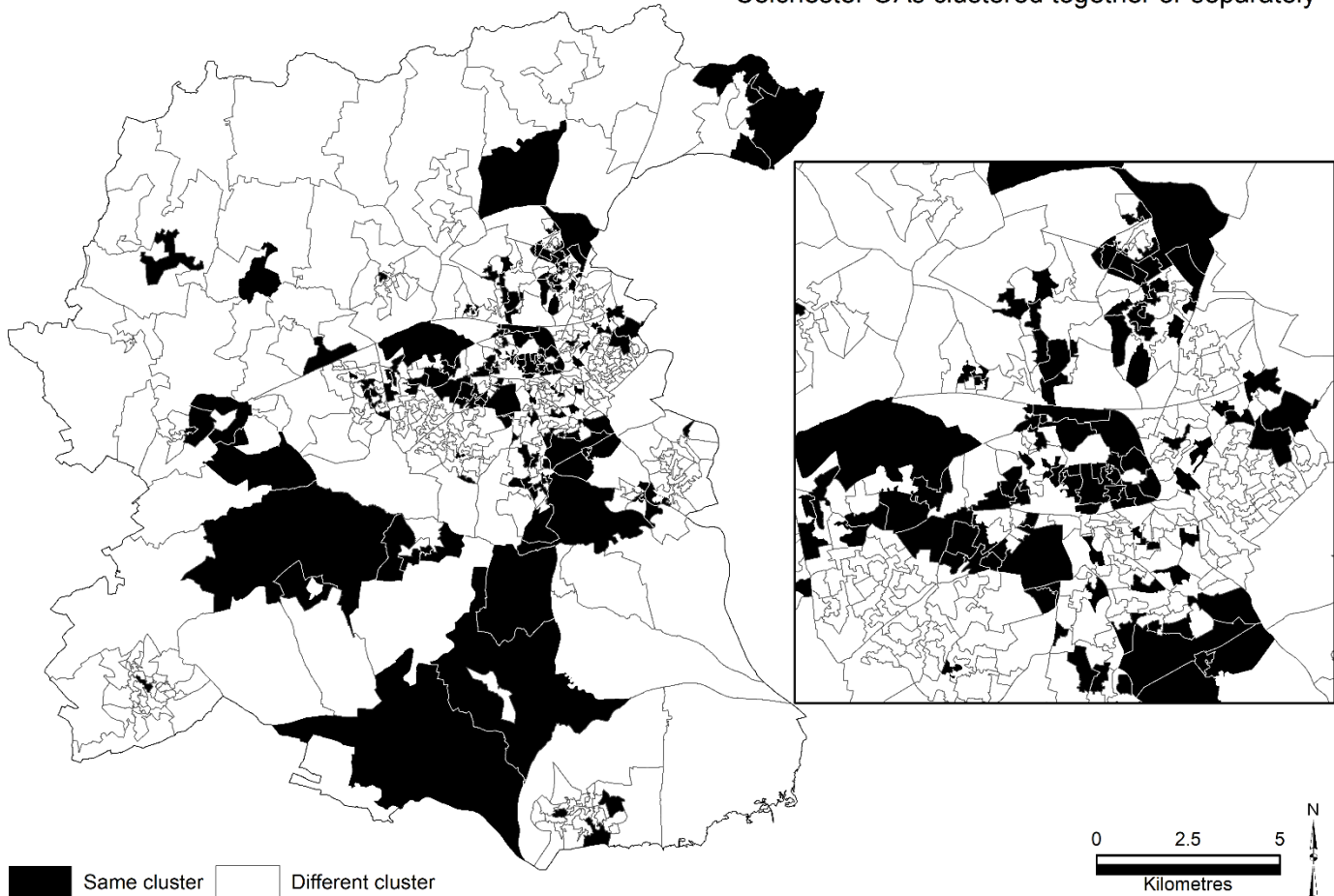


Figure 7: Comparison of census only and census plus land use classifications in Colchester and Salford districts.
Source: Ordnance Survey, Edina. © Crown copyright and/or database right 2019 OS

Colchester OAs clustered together or separately



Salford OAs clustered together or separately

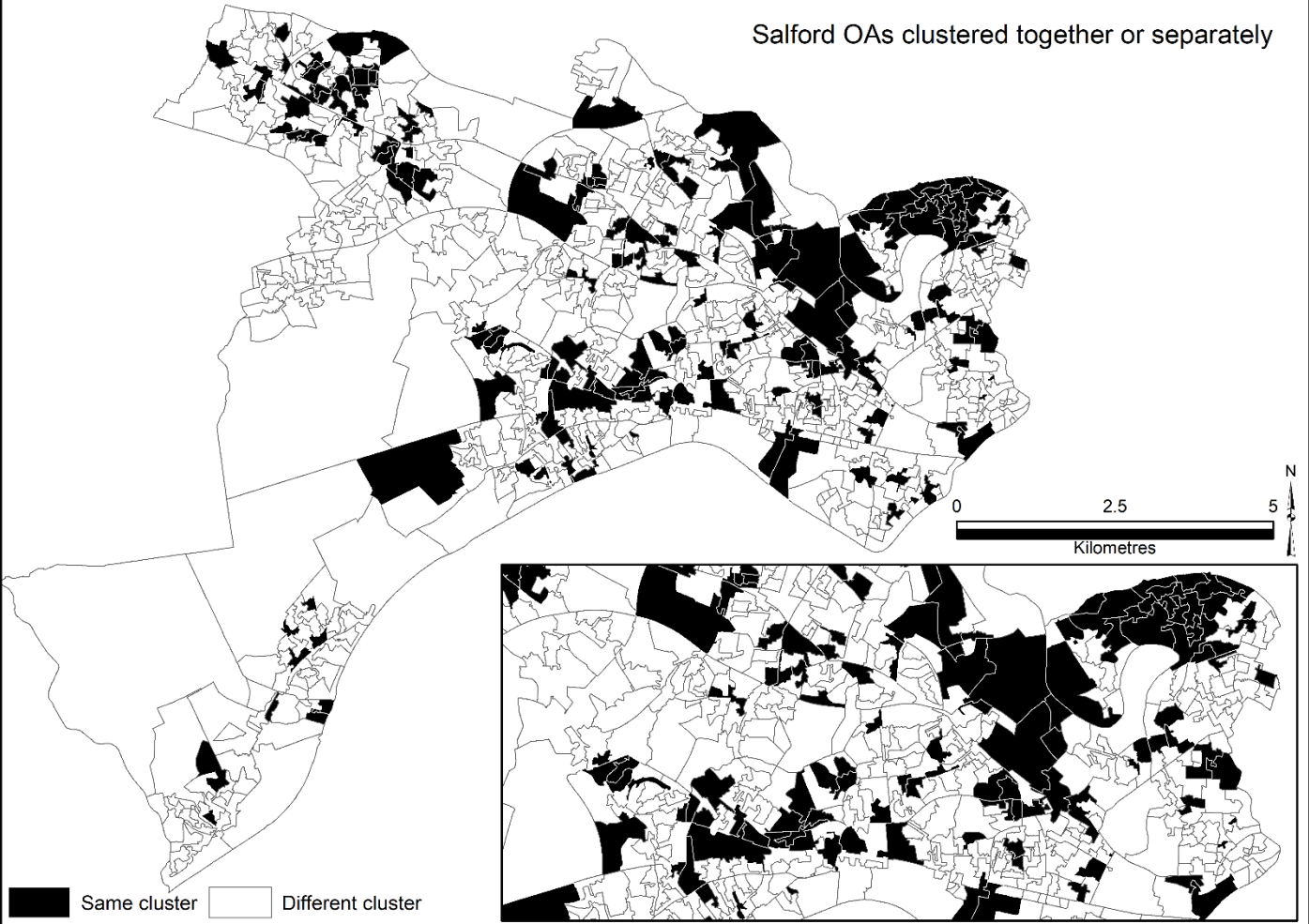


Figure 8: Comparison of output areas classified together and apart in the census only and census plus land use classifications in Colchester and Salford districts.

Source: Ordnance Survey, Edina. © Crown copyright and/or database right 2019 OS