Population, Space and Place

WILEY

RESEARCH ARTICLE OPEN ACCESS

Address- Versus Postcode-Based Building Blocks for the Creation of Small Area Statistical Geographies: Some Lessons From the 1911 British Population Census

Nigel Walford

Department of Geography, Geology and the Environment, School of Built Environment and Geography, Kingston University, Kingston upon Thames, UK

Correspondence: Nigel Walford (N.Walford@kingston.ac.uk)

Received: 5 June 2024 | **Revised:** 15 May 2025 | **Accepted:** 19 May 2025

Funding: British Academy

Keywords: automated zone design | censuses | postal geography | Re-zoning

ABSTRACT

Postal geography is increasingly used to define spatial units for data collection and statistical reporting in British population censuses. There has been a shift from defining a single set of spatial units for each census to creating areas with greater consistency across censuses using artificial polygons around collections of postcoded addresses. However, the question remains whether creating new zones for the purpose of reporting aggregate statistics can feasibly be carried out by starting from individual addresses rather than postcode units, while at the same time preserving the confidentiality of households and individuals. Although the issues could be approached by using modelled data derived from modern censuses, the approach adopted here has been to use genuine household and individual level data from the 1911 Census for addresses that have been georeferenced in a selection of local authorities. Testing the difference between address- and postcode-based new zones was carried out by aggregating from two sets of Thiessen polygon building blocks, one for individual 1911 census addresses and the other for these addresses grouped within modern unit postcode boundaries. The sets of new zones were assessed by measures of target population attainment, homogeneity and shape compactness and in respect of how they allocated different percentages of the addresses to the same spatial unit. People and households were distributed at lower densities across new zones that occurred in relatively unpopulated areas, although even in more densely inhabited parts of the study areas some people and households were colocated in the same new zones produced from the two starting sets of different Thiessen polygons.

1 | Introduction

The geospatial characteristics and statistical content of the data produced from the 1981, 1991, 2001 and 2011 British censuses changed significantly over this 30 year period. Some of these changes can be attributed to the completion of a nationwide postal geography in 1974, although some larger cities had postal areas earlier including a set of 10 districts in London based on the compass points introduced in January 1857 (Raper, Rhind and Shepherd 1992). The lowest tier of British postal geography is referred to as the Unit Postcode (or postcode unit) and these were originally defined as groups of addresses with a single grid reference to the most south westerly corner of the 1 km grid square in which they were located. Although boundaries for higher levels in the postal geography (post town, postal district and postcode sector) were defined, they were not drawn around the groups of addresses in Postcode Units, which allowed degree of flexibility. Such postcode unit boundaries were eventually created, almost as a by-product of the process introduced to define the lowest tier of spatial units used for reporting statistics from the Population Census in 2001. Nevertheless, there was a close connection, but not initially synchronisation, between census and postal geographies from the 1970s and 1980s.

Postcode units are regarded as having too few households and individuals for them to act as the basic spatial unit for releasing aggregate demographic and socioeconomic data, apart from with respect to simple, relatively uncontentious head counts. The following section outlines the developing connections between British census and postal geographies. It also reviews the automated zoning system devised to make spatially bounded postcode units (i.e., polygons) central to the geographies developed for reporting statistical information from the 2001 and 2011 British censuses. This involved developing an algorithm to search interactively for optimal combinations of postcode units that satisfied certain population and household thresholds and other criteria related to nesting within statutory boundaries (Martin 1998a, 1998b; and 2002). The work developing this tool involved access to individual and household level data, nevertheless the comparison of the zones created from postcode unit polygons with other building blocks across a substantial area has not to the author's knowledge been attempted, Norman et (2023: 9) carried out work related to this issue from the perspective of balancing "confidentiality protection and … reassurance to the research community" using synthetic data.

The present article uses historical census data released under the 100 year data accessibility procedures to inform the contemporary debate about spatial demography. It georeferences the data for individuals and households in the 1911 British Population Census for a group of local authorities in the former counties of London and Middlesex and applies the modern postcode unit geography to these addresses. The same source data are thus georeferenced in two ways as individual address points and as areas containing groups of addresses within modern unit postcodes, referred to as pseudo postcodes. The data and methods section provides some background information on the study areas and application of the AZTool (Martin et al. 2001; Cockings and Martin 2005; Flowerdew et al. 2008). New zones created by applying this tool to both address and (pseudo) postcode unit Thiessen polygons are compared in respect of satisfying population and household thresholds, homogeneity and compactness measures (Martin 2004). Differences between the two sets of new zones are examined by quantifying the extent to which addresses are located together and whether aggregate statistics for the zones are significantly different. Finally, the discussion and conclusion section considers the implication of these results for defining contemporary statistical reporting zones in an era when the traditional census enumeration may become a thing of the past.

2 | Postcodes, Census Geographies and Automated Zoning

There are nearly 30 million addresses within 1.8 million postcode units in the UK giving an average of 17 addresses each (Royal Mail n.d.). Their structure provides a degree of flexibility allowing new addresses to be added and expired ones to be removed relatively easily. The initial aim of the postcode system was to streamline and automate the collection, sorting and delivery of items sent through the postal service, but this role was soon augmented by those who saw the potential of using postcodes as geospatial references linked to the characteristics and movement of people in the population. One of the earliest uses was for Regional Highway Traffic Modelling using the postcode as a means of determining grid references for the origins and destinations of traffic flows in conjunction with transport surveys in the 1970s (Duncan 1978). However, one of their main non-postal uses in academic research, policy implementation and commercial applications has been as a way of grouping together households for the purpose of describing their demographic and socioeconomic characteristics. People's growing familiarity with their postcode since they were finalised nationally in 1974, achieved in part by media campaigns encouraging their use and the franking of envelopes with the reminder "Remember to use the postcode" (UK Meter Franking 2017). This increased familiarity led commercial and public sector organisations from the 1980s onwards to treat the postcode as a key piece of information to assist with logistical and marketing aspects of their operations especially in respect of supplying goods and services to the public. The predictive capacity of postcodes in respect of the characteristics of people and households has been widely acknowledged (Webber and Burrows 2018). These developments were also linked to growing interest in joining postal and census geographies.

The addresses allocated to individual postcode units have remained relatively stable and unchanging since their introduction in comparison with the re-districting carried out in respect of electoral and census geographies, in the latter case until 2001. It is not the intention to suggest that postcodes are unchanging, but an address assigned to a postcode when those identifiers for the whole country were finalised in 1974 is likely to have the same postcode now over 50 years later. Addresses created during this period, as a consequence of residential development that is not interspersed amongst existing properties, are likely to have been assigned a new postcode. In contrast changes in respect of census geography occurred to a substantial proportion of the smallest spatial units (enumeration districts up to 1991) due to these being substantially redrawn for each enumeration and in the case of electoral geography because both central and local government areas are periodically redefined to reflect changes in the distribution of the population and decisions concerning political representation (Norman et al. 2003; Walford and Hayles 2012). The small area census geography of England and Wales in 1991 included 14.9% of enumeration districts that were identical in 1981 and 1971 accounting for 25.5% of the land area, 30.1% of 1981 enumeration districts were used in 1991 (34.8% of land area) (Walford and Hayles 2012).

Figure 1 illustrates changes to the geo-referencing of postal addresses and postcodes and the connection with census statistical output geographies since nationwide postcodes were introduced. The first part (Figure 1a) represents the situation in the 1970s and 1980s when only the National Grid reference of the most south western corner of the 1 km^2 was captured, and the decennial census geography was defined by reference to physical features in the landscape (e.g., centre lines of water courses and roads). Some postcodes had all their addresses in a

FIGURE 1 | Historical sequence of changes in digital representation of unit postcodes and small census geography 1970s to 2000s in England and Wales (hypothetical Illustrative example). (a) Addresses for buildings in unit postcodes without boundaries, 1970s and 1980s (referencing to SW corner of 1 km grid sq). (b) Buildings, address points and 1970s/1980s EDs (addresses and buildings in unit postcode split by ED emphasised – (see (c)). (c) Buildings, address points with 1990s enumeration districts (addresses and buildings in two unit postcodes split by EDs emphasised). (d) Unit postcode and output area type boundaries (postcode and census boundaries harmonised) from 2001 onwards. Source: Ordnance Survey, Edina. © Crown copyright and/or database right 2022 OS; Ordnance Survey data © Crown copyright and database right 2011.

single census enumeration district, whereas in other cases they were distributed into two, three or possibly more enumeration districts (Figure 1b). The capture of grid references for all addresses was part of the Ordnance Survey's digitisation programme of its topographic mapping in the 1990s and despite locating individual addresses within a unit postcode it was completed too late to overcome the issue of addresses in a single postcode often being distributed among a number of 1991 Census EDs (Figure 1c). This difficulty was only remedied after a radical change in the way in census geographies were defined came about for the 2001 census, which has been continued in subsequent enumerations (Martin 1998a, 1998b; and 2002). This involved the formation of Thiessen polygons around the addresses in a unit postcode clipped to topographic features (Martin et al. 2001) and interactive amalgamation of these units to produce output zones that satisfied certain criteria. The transition to a method for designing the decennial census geography by aggregating geospatial building blocks in 2001 (Figure 7d) was extended in 2011 to produce a separate set of workplace zones based on the working rather than the residential population of unit postcodes. The imperative for identifiable, visible boundaries of census small areas to be visible diminished further when postal return and online completion were introduced for the censuses held in 2011 and 2021 respectively (Cockings et al. 2011).

It is generally acknowledged that two important advantages of the new method for defining census geography are the seamless connection with unit postcode polygons and the opportunity for retaining a higher proportion of the Output Areas between successive censuses. Only some "2.6% of 2001 output areas were changed from the 2011 Census" in England and Wales (Tait 2012), which represents a significantly higher level of consistency than between previous pairs of enumerations (see above). Despite these changes in the nature and definition of small area census geographies, and general acceptance that spatial units based on postcodes are preferable, the statistics for these areas remain in aggregate form. One of the main reasons is to ensure that confidentiality thresholds are satisfied to prevent the inadvertent release of confidential information or its derivation through a process of differencing by overlaying and combining overlapping sets of zones (Duke-Williams and Rees 1998). However, this suggests that at least one important question remains unanswered, namely whether the geospatial units used for reporting demographic and socioeconomic statistics from the population census would be drawn differently or could even be improved if more was known for certain about the characteristics of people living at individual addresses at the time these areas were defined. This article addresses this question by extending the application of an automated zoning tool developed and used in the UK to produce output area zones for the 2001, 2011, and 2021 censuses to aggregate Thiessen polygons created around individual addresses. It builds on the work of Cockings et al. (2011: 2406) which recognised the lack of "readily available datasets which provide the small-area distribution of all people and households" by using geocoded household and individual level records from the 1911 census, which are available under the 100 year data release procedures. Cockings et al. (2011) aimed to quantify the extent to which output areas might need to re-defined between the 2001 and 2011 censuses, whereas here the concern is with investigating if such output zones produced from address level data might offer a greater degree of homogeneity, compactness and respect for population and household thresholds than can be achieved when starting from unit postcode polygons.

The desire to define statistical and census geographies that are useful for different types of user, including academic researchers, policy analysts, resource planners and commercial marketeers, lies at the centre of the changes that have occurred over recent decades. However, in tandem with these developments was recognition that the release of data about households and individuals collected at the level of their addresses but aggregated to higher spatial units raises the Modifiable Areal Unit Problem (MAUP), highlighted in respect of British census geography by Openshaw (1977). There are two elements to the MAUP: the potential for variations in analytical results arising from moving up or down the spatial scale from smaller to larger units (or vice versa) (see e.g., Duque et al. 2007; Dumedah et al. 2008); and the essentially arbitrary nature of any zoning system that allocates point features (households and individuals located at addresses) to an infinite number of possible outcomes (Flowerdew 2011). Openshaw's (1977) Automated Zoning Procedure addressed the MAUP by a process of merging and remerging spatial building blocks to produce an optimal set of output zones that satisfied design criteria. The AZP constituted the foundation for further work resulting in the creation of the Automated Zone Matching software (Martin 1998a and 1998b) and an automated method for creating the 2001 census output areas (Martin 2002). Further enhancement led to the AZTool software at the University of Southampton (Cockings et al. 2011, 2013), which includes a utility to prepare. aat and. pat files without the need to use Arc Info. Despite several studies applying these automated design principles in different contexts and countries (Cockings and Martin 2005; Haynes et al. 2007; Flowerdew et al. 2008; Haynes et al. 2007; Ralphs and Ang 2009; and Sabel et al. 2013), it remains the case that the majority use polygonal building blocks around groups of addresses (i.e. previously aggregated data) as their input rather than around the addresses themselves. The principal research question explored here is whether creating new zones from historical census records, similar to the output areas now available for modern British censuses, can provide an indication of whether even more homogenous robust, zones could be produced if the starting point was individual grid-referenced addresses rather than groups of addresses in postcode units.

3 | Data and Methods

The study areas comprised ten of the local authorities that existed in the historical counties of London and Middlesex at the time of the 1911 Population Census, five in each (see Figure 2). The ten study areas were chosen on the basis that they represented places with contrasting population densities, different stages of urban and suburban development and, in the case of the Rural District of South Mimms in the north of Middlesex, an area largely composed of farmland and open countryside in 1911 with scattered settlements and a number of farms and estates. This contrasts with, for example, Deptford an industrial and residential area on the south bank of the Thames towards the south-eastern edge of London. These areas experienced different histories of population change during the first decade of the twentieth century (see Table 1): for example, population and household totals declined by 26.9 and 25.2% respectively in the City of London but only by 7.3 and 2.8% in Hampton Wick (Middlesex). Deptford and Hammersmith in London County represented local authorities respectively with virtually static and substantially growing populations and household totals during this decade. Apart from Hampton Wick the other local authorities in Middlesex increased their population and household numbers. These changes in the numbers of persons and households were reflected in the difference in population and household density figures over the decade 1901–1911. Increased population density occurred in Hammersmith and Wembley, whereas greater household density was relatively modest across most areas. Overall, the selection of study areas was intended to reflect low, moderate and substantial population growth.

A method of geocoding the census records for all occupied addresses was originally developed for six of these areas and reapplied to increase the range of areas for the present purpose. The geocoded addresses were matched with original census enumeration records, their transcriptions and the Integrated Census Microdata to form the basis of the present analysis (Walford 2019). Data files were created for each study area that contained the address grid references together with a selection of demographic and socioeconomic variables similar to those

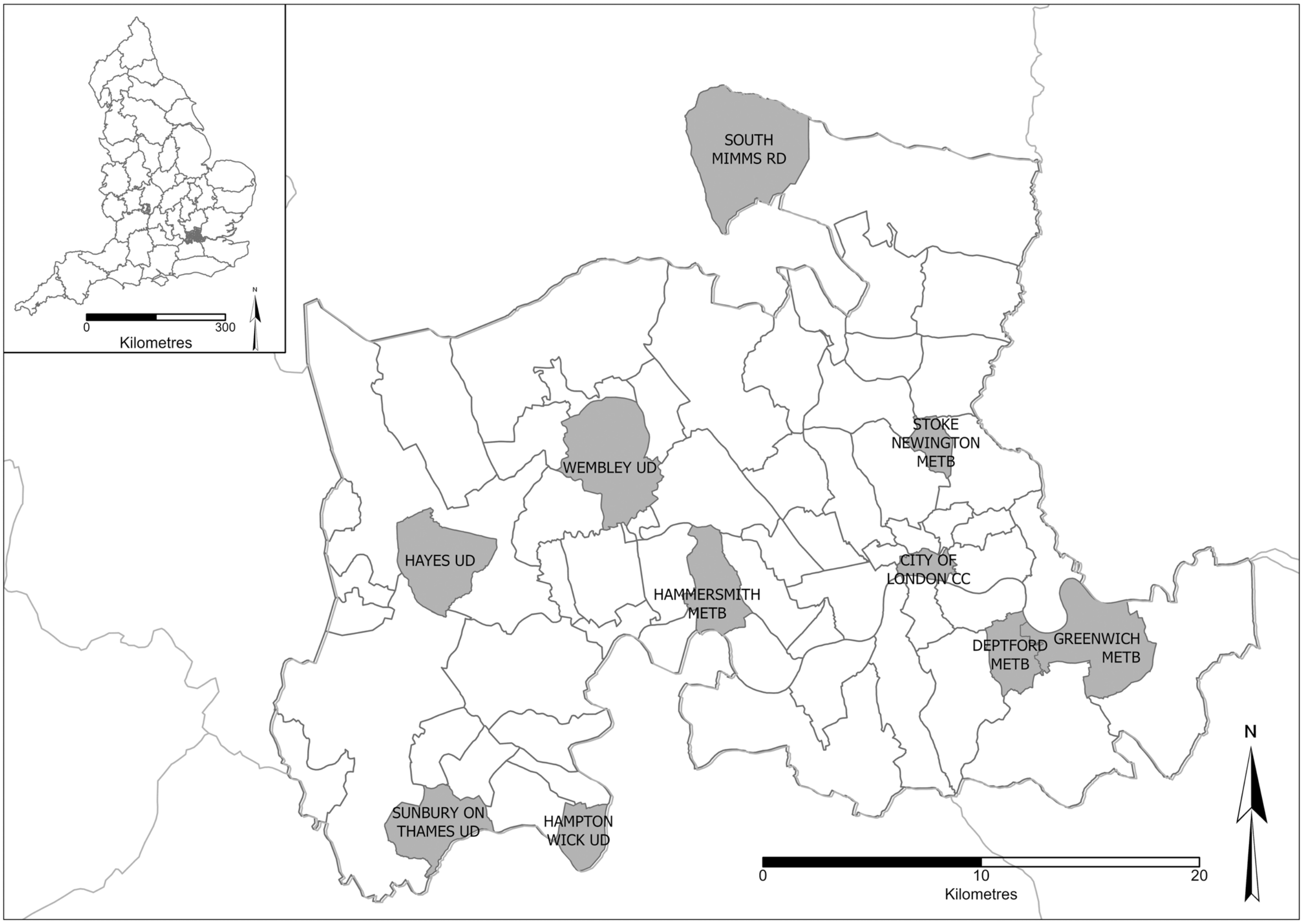


FIGURE 2 | Location of ten London and Middlesex local authorities in 1911. Source: Ordnance Survey, Edina. © Crown copyright and/or database right 2022 OS.

TABLE 1 | Percentage change in population and household numbers and density per hectare 1901–11 in study areas and historic counties of London and Middlesex.

	Population change 1901–1911 (%)	Household change 1901–1911 (%)	Difference in population density 1901–1911 (%)	Difference in household density 1901–1911 (%)
City of London	−26.9	−25.2	−26.6	−4.9
Deptford MB	−0.8	1.7	−1.4	0.7
Greenwich	0.2	5.0	0.1	0.6
Hammersmith	8.3	10.3	10.0	2.3
Stoke Newington UD	−1.1	7.8	−1.7	2.5
County of London	−1.6	0.5	−9.8	−1.8
Hampton Wick UD	−7.3	−2.8	−0.4	0.0
Hayes UD	64.3	51.1	1.2	0.2
South Mimms RD	5.0	11.6	0.0	0.0
Sunbury on Thames UD	1.4	4.7	0.1	0.0
Wembley MB/UD	136.7	168.9	3.3	0.8
County of Middlesex	44.9	55.1	8.0	2.0

Abbreviations: MB, Metropolitan Borough; RD, Rural District; UD, Urban District.

used in research concerned with defining and assessing the need for amendment of modern census output geographies (Martin 1998a, 1998b, 2002, 2010; Cockings et al. 2011). The variables used were persons and households per address, and accommodation type.

The AZT algorithms start from an initial random aggregation of the individual building blocks and proceed by means of a series of iterations to swap these units according to user-defined criteria in respect of target population threshold, homogeneity and shape compactness (Martin 2003; Sabel et al. 2013). The AZTool software [version 1.0.3.0] (http://www.geodata.soton.ac.uk/software/AZTool/) was used to create two sets of optimal new zones starting from two different sets of 'building blocks'. The first were Thiessen polygons within each study area generated around the individual address points and the second was Thiessen polygons created around those groups of addresses that shared the same modern unit postcode, the latter mimicking recent practice in respect of British census geography. It would not have been appropriate to use the existing modern postcode unit polygons themselves for three main reasons: first, substantial numbers of modern postcodes in the study areas covered uninhabited land in 1911; second, they are based on the current addresses only some of which were occupied or existed in 1911; and third, some modern postcode unit polygons were split between historical local authorities. The AZT allows specific constraints to be imposed, for example prohibiting donuts (one output area surrounded by another) and requiring new zones to be entirely within a higher level geography (Ralphs and Ang 2009). The latter was not applicable in this case as the 10 study areas were treated as discrete, isolated local authorities, whereas if resources had permitted geocoding all 1911 Census addresses across the historic counties of London and Middlesex the boundaries of the 66 'old' coal authorities in these could have been used as spatial constraints. The output from using the AZTool is a tract or new zone definition (. csv) file that be joined to the input shapefile enabling aggregation of the building blocks and their accompanying statistical data to create the new zone polygons.

The results of re-zoning are generally assessed with respect to three parameters, the successful application of population and household thresholds, the degree of homogeneity and the compactness of zone shape. Minimum and maximum population and household totals are usually used to maintain confidentiality and are typically determined by multiplying the household target by the estimated average household size (e.g., 2.5). However, these thresholds are more than simple quantitative limits as they also reflect judgement about public acceptability in relation to the release of statistics. The population and household thresholds ("minimum of 40 households and 100 residents" (Hayes et al. 2018: 74)) used by the British census authorities in the 2001 Census were reapplied in 2011 (Cockings et al. 2011), and have been reused in the present analysis of historical census data despite the passage of time and the lower population and household levels in the first decade of the twentieth century to replicate current accepted practice. Application of the recently used thresholds might be regarded as unnecessary, given that the release of address level data records from British censuses over 100 years ago effectively removes the need to protect confidentiality, nevertheless their use made the analysis comparable with modern census zone building. Re-zoning from an existing set of polygons invariably means that the frequency distributions of the population and household totals for the new set of zones generated by the process have statistical dispersion, and a successful outcome may be measured by a smaller standard deviation indicating they are more tightly clustered around the mean.

A measure of homogeneity was an important feature of the zone creation system introduced for the 2001 British Census as the former enumeration districts were criticised for having insufficient regard for socioeconomic and demographic homogeneity. In the present case such a measure also serves as a means of comparing the zones created by aggregation of postcode units with those from addresses. The inter-area correlation (IAC) measure was used, which is the correlation of a given variable between different people living in same areal unit (Tranmer and Steel 1998; Martin et al. 2001; Cockings et al. 2011). Stronger positive correlation indicates greater homogeneity. It is customary to limit homogeneity assessment to a small number of variables, for example accommodation type and tenure (Cockings et al. 2011) as proxies for the homogeneity of the social built environment. The 1911 Census recorded information about type of building occupied, which was used to derive a building type code (Higgs et al. 2013) including categories such as private house, inn, hotel, prison, school, and shop. This code was combined with visual inspection of the historical Ordnance Survey topographic maps once the addresses had been geocoded to produce a fivefold classification of dwelling type that reflects the situation in modern censuses. The descriptive terms used were detached house or bungalow, semi-detached house or bungalow, mid- or end-of-terrace dwelling, flat, and other including barges or canal boats, vans, shops, tents, and homeless. Housing tenure, the second variable typically used in IAC assessment, was not included in the 1911 census. Although an imperfect substitute for household tenure, the number of separate households at an address, as an indicator of multi-occupancy and rental tenure, was also assessed using AIC. The 1911 Census schedule included the following instruction to occupiers: "If a house be let or sub-let to two or more occupiers each occupier of a part of the house must fill up a Schedule for his [sic] part of the house. Boarders are not considered as separate occupiers" (Higgs et al. 2013: 132). The majority of addresses across the set of 10 local authorities were occupied by one household (at least 52% in all areas), but a substantial minority in the London areas (between 11.4 and 47.2%) accommodated two to four households and a small proportion were inhabited by larger numbers ranging between 0.3 and 3.8%. In all areas there were some addresses for communal establishments and institutions (e.g., hospitals, prison, etc.). Shape compactness was measured by the mean shape score defined as the perimeter^2/area with lower values indicating more compact (circular) areas and larger ones less compactness (linear areas) (Cockings and Martin 2005; Haynes et al. 2007).

The workflow for preparing the geocoded 1911 census address point data and with attributes for population and household totals, accommodation type, and number of households at address, and creating two sets of Thiessen polygons for input to the AZTool is illustrated in Figure 3. The starting point in each

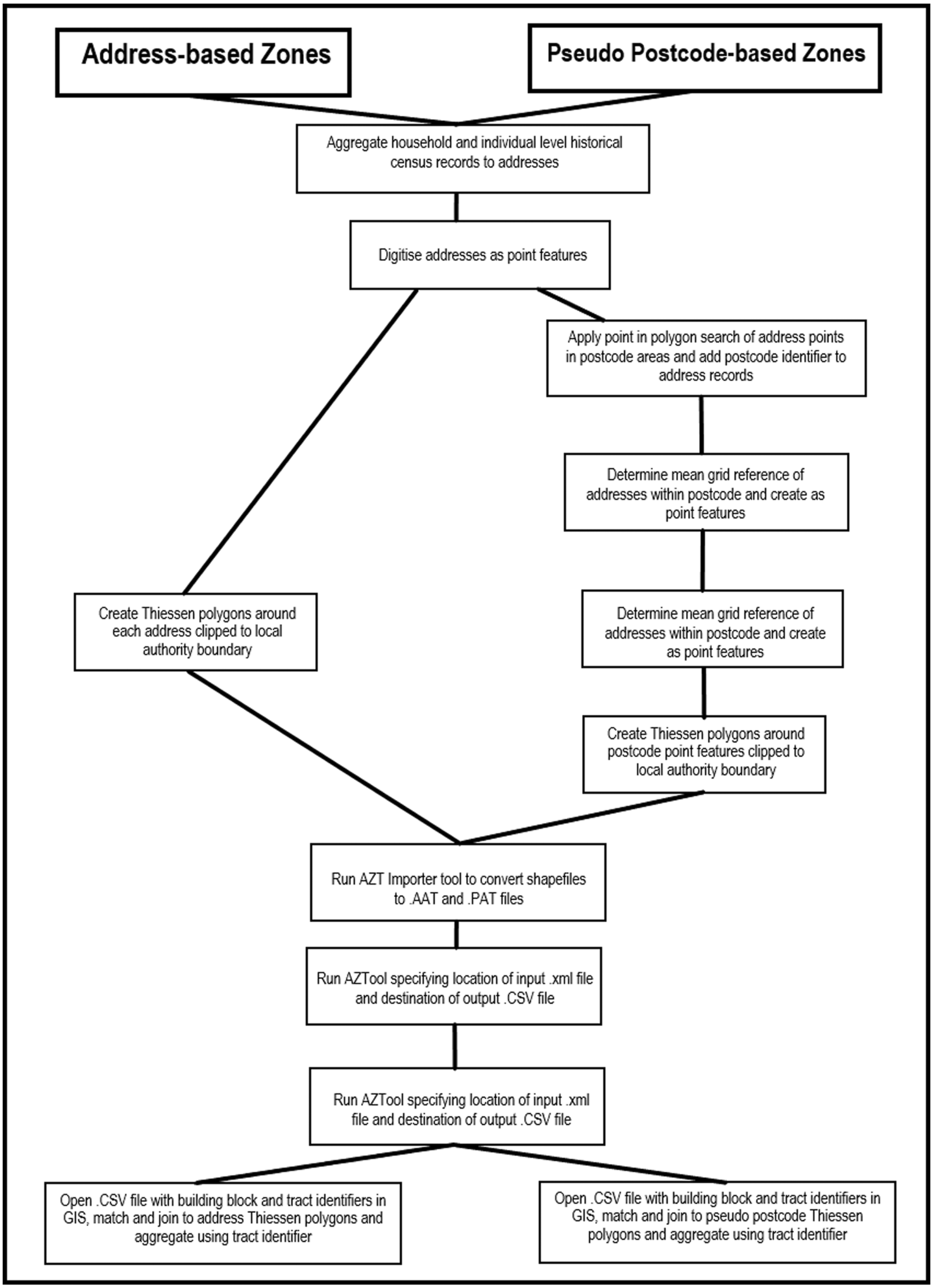


FIGURE 3 | Workflow for creating Thiessen polygons created around individual address points and around the groups of addresses sharing identical unit postcodes.

case is the geocoded address with the associated census attribute data held in shapefile and. csv formats. The key difference in the method for creating the two sets of output areas was that to create the pseudo postcode polygons addresses were first grouped according their modern postcode and the mean or average grid reference for the group was determined. This was used for generating the Thiessen polygon, hereafter referred to as the pseudo postcode-based polygons or zones. This was not

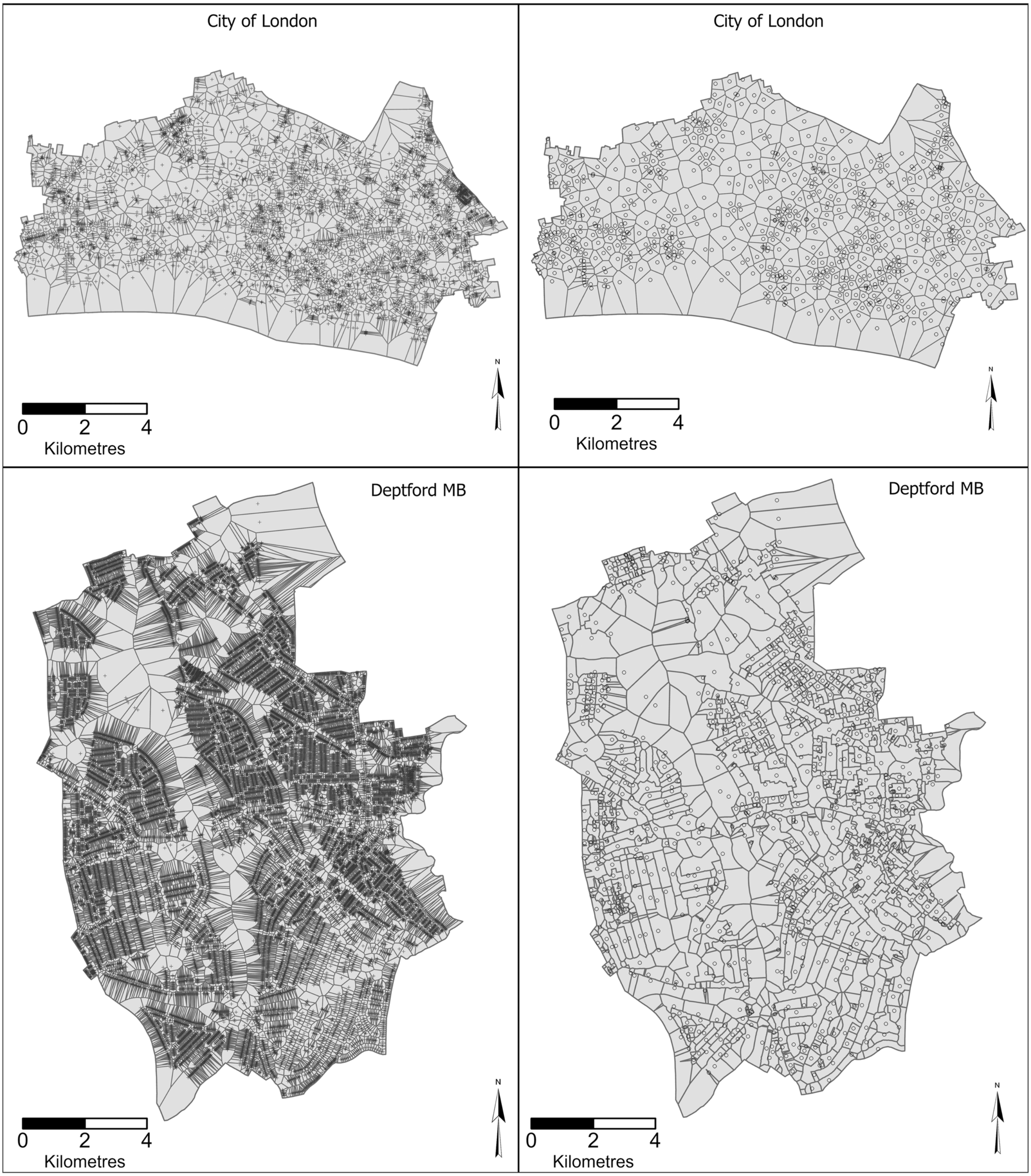


FIGURE 4 | Thiessen polygons created around individual address points and around the mean grid point of groups of addresses sharing identical unit postcodes for each of the ten study areas in the Counties of London and Middlesex in 1911. Source: Ordnance Survey, Edina. © Crown copyright and/or database right 2022 OS.

required for producing Thiessen polygons around individual addresses, subsequently referred to as address-based polygons or zones. The AZTool requires the IAC variables to be spread across separate fields corresponding the accommodation types (detached, semi-detached, terraced, flat and other) and the number of households at address (1, 2, 3, 4 and 5 +). The storage location of the. aat and. pat files together with parameters, including lower and upper population thresholds of 100 and 625 (40 and 250 households), homogeneity and shape compactness measures, for the AZTool were specified in a. xml file. These parameters were unweighted and given equal importance. The next stage in the workflow involved running

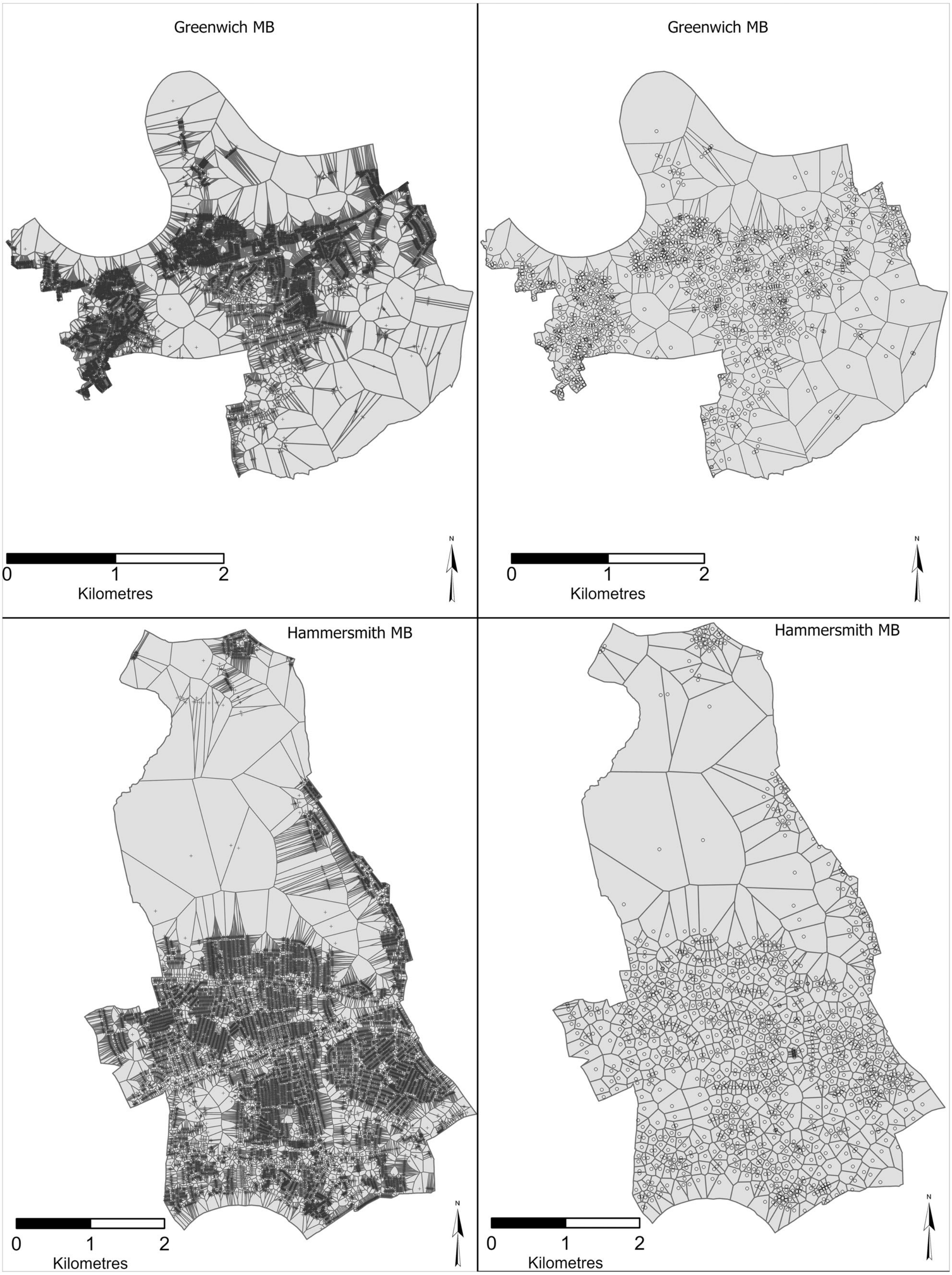


FIGURE 4 | (Continued)

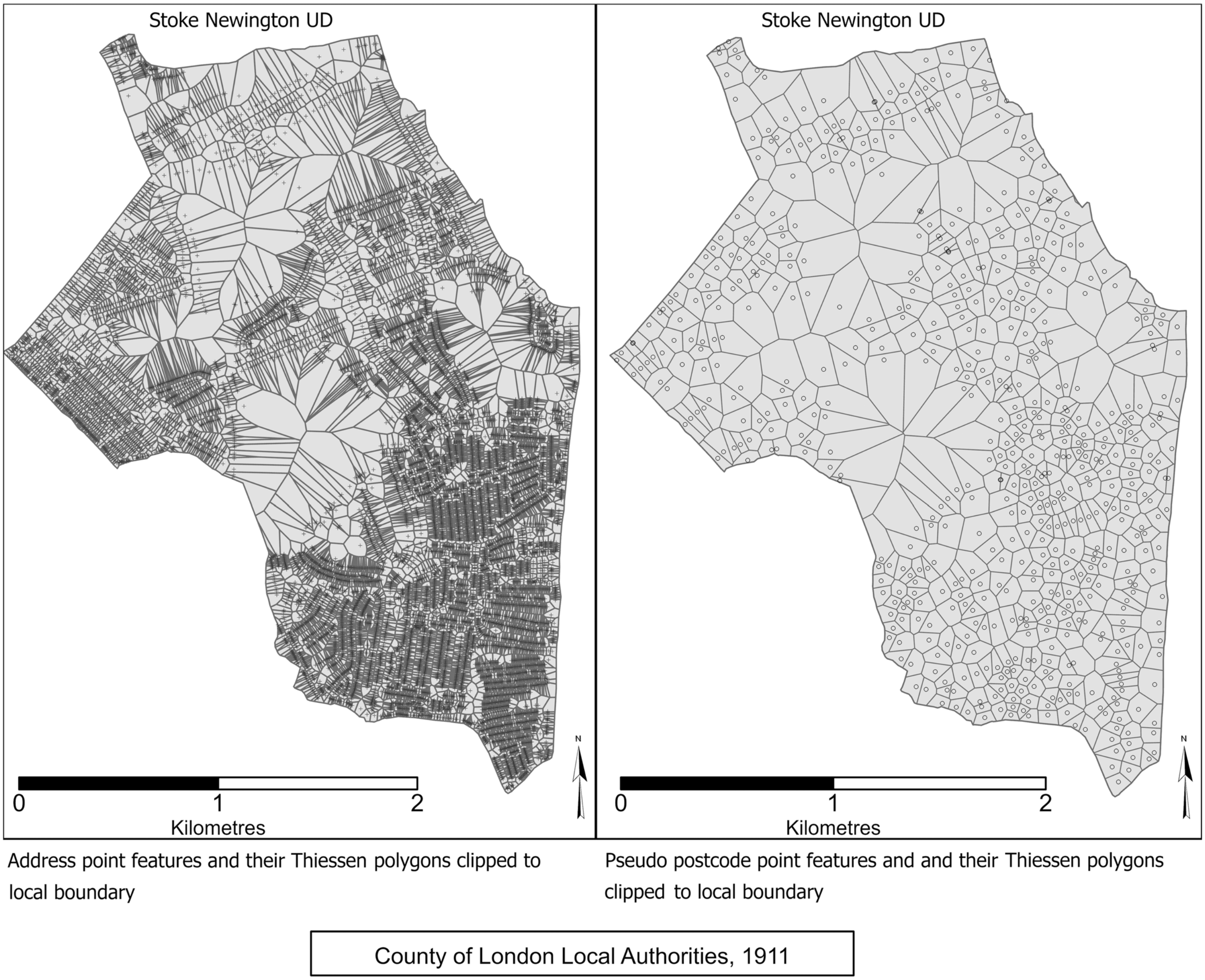


FIGURE 4 | (Continued)

the AZTool to determine the membership of two sets of output zones or tracts, one for each of the address and pseudo postcode Thiessen polygons (building blocks). The. csv output file from each application of the AZTool was joined to the corresponding input Thiessen shapefile and used as the basis for combining the building blocks to the new zones. An analysis of their characteristics in terms of satisfying population and household thresholds, homogeneity and shape compactness measures form the basis of assessing whether consideration should be given to using individual addresses rather than postcodes for defining modern output areas.

Figure 4 illustrates the address and postcode point features together with the Thiessen polygons created around them for each area. The individual parts of Figure 4 also show the modern unit postcode boundaries that contained any addresses in 1911 (i.e. excluding land where subsequent residential, industrial and other development generated new addresses and therefore postcodes). The modern postcode areas were not expected to match the historical versions because they were based in different sets of addresses, nevertheless there appears to be a degree on congruence in some places. The Thiessen polygons were constrained to fit within the historical local authority boundary, whereas the modern postcode boundaries were obviously not limited in this way and in some cases cross over into adjacent areas not included in this analysis. The Thiessen polygons around the 1911 addresses and those forming the historical postcodes reveal differences between the more densely populated areas of the City of London, Deptford and Hammersmith where addresses occurred across much of the land and the scattered populations of the Middlesex study areas, where there were a few concentrations of addresses in towns and villages together with people scattered in isolated farms and hamlets. However, Deptford and Hammersmith and to some extent the City of London also included some land the was relatively unpopulated and covered by transport infrastructure, parkland, open space and similar uses that had not be built over. These variations in the distribution of addresses contribute to the differences in the shape and size of the Thiessen polygons. Nationally such variation in the density of residential addresses remains in place today even if the study areas themselves now include substantially more residential

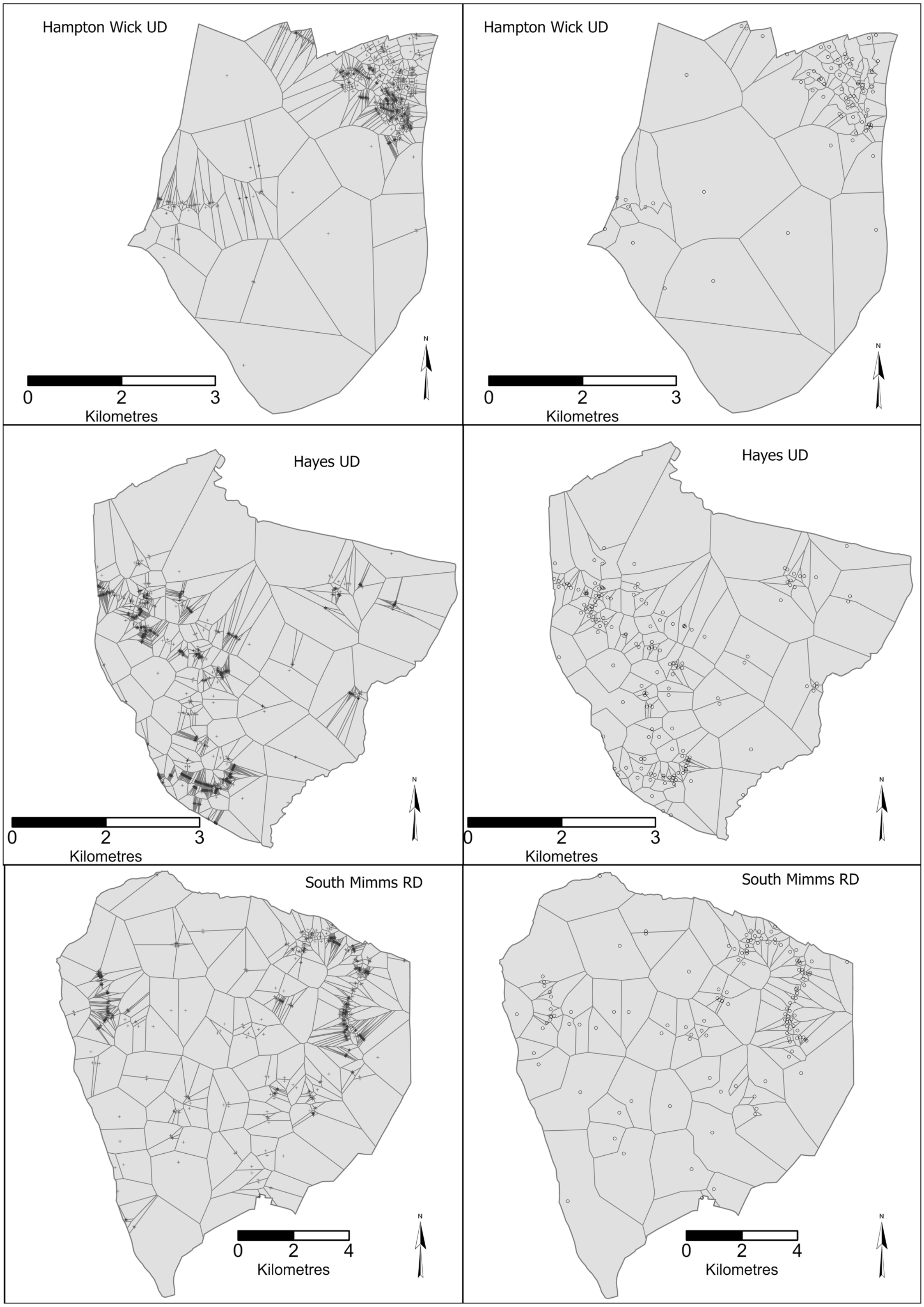


FIGURE 4 | (Continued)

FIGURE 4 | (Continued)

addresses. The City of London, Deptford, Greenwich and Hammersmith each included two or three addresses for large communal establishments, such as hospitals, prisons and workhouses, where the total number of individuals present exceeded the maximum population threshold. These were excluded when applying the AZTool, but they were added to the list of new zone identifiers so their Thiessen polygons were treated as whole areas when the other address-based or pseudo postcode-based ones were aggregated.

4 | Results

This section presents the results of aggregating the historical 1911 census records to both sets of new zones (i.e., based on the address and pseudo postcode Thiessen polygons). The results are examined in respect of the three indicators outlined previously (population and household thresholds, homogeneity, and compactness of zone shape). One criterion for assessing the relative success of the two methods

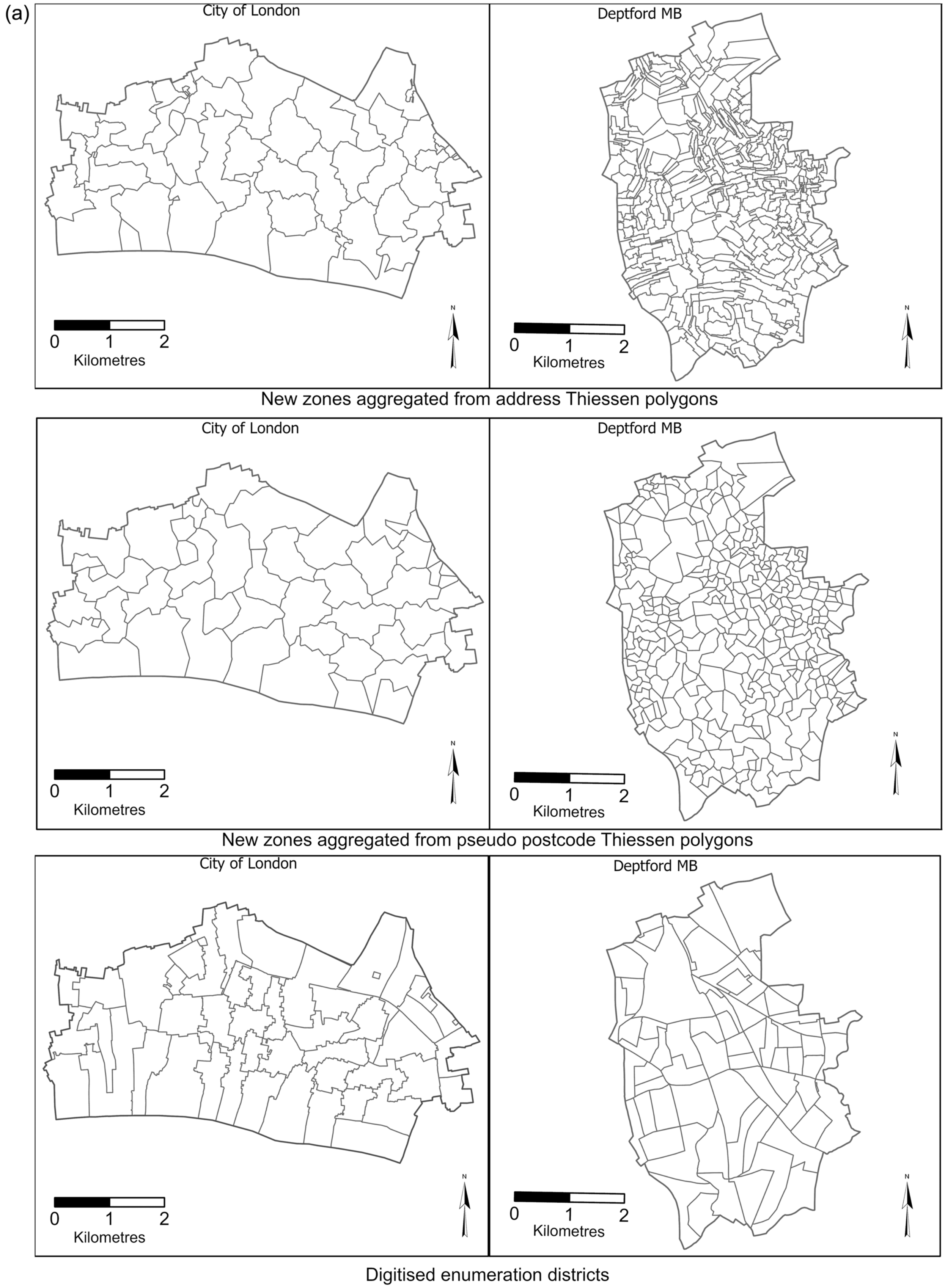
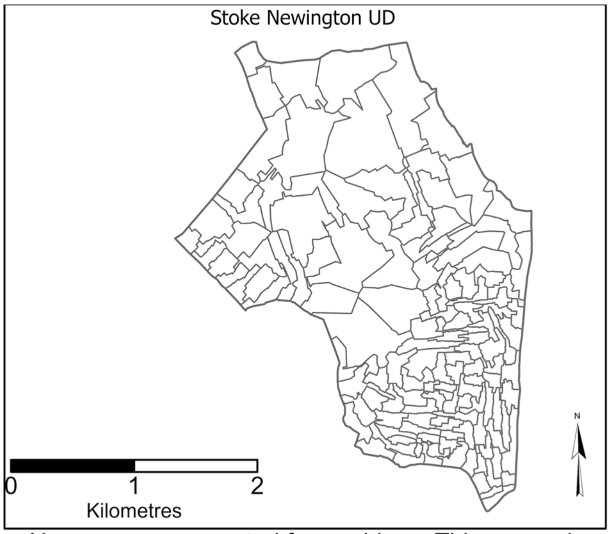
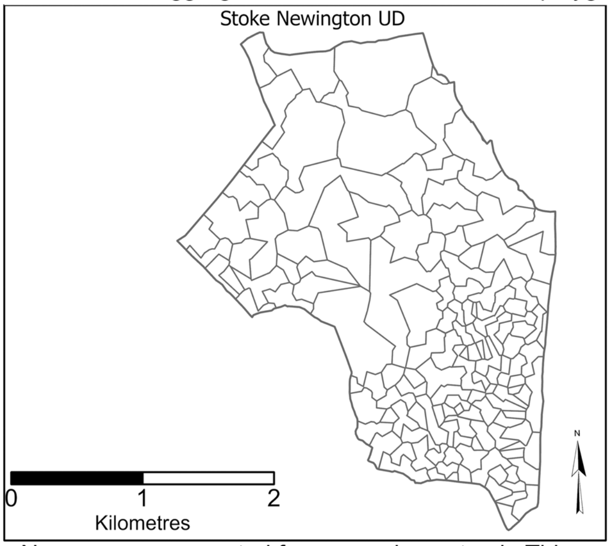


FIGURE 5 | New zones (tracts) created using output from the AZTool to aggregate building blocks (Thiessen polygons) around individual address and pseudo postcode mean points for each of the study areas together with 1911 enumeration district boundaries. (a) London areas. (b) Middlesex areas. Source: Ordnance Survey, Edina. © Crown copyright and/or database right 2022 OS.

FIGURE 5 | (Continued)



New zones aggregated from address Thiessen polygons



New zones aggregated from pseudo postcode Thiessen polygons

Digitised enumeration districts

FIGURE 5 | (Continued)

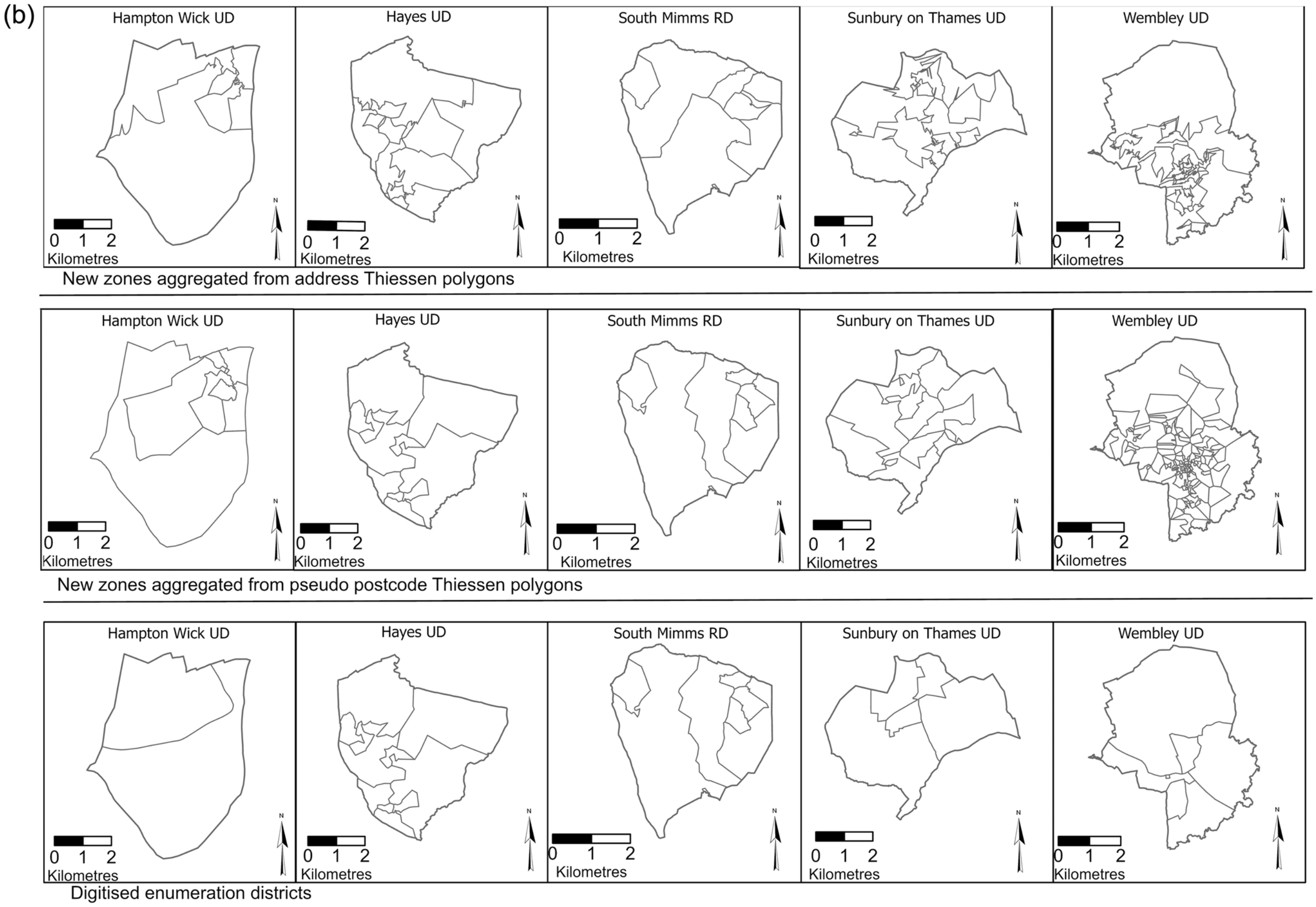


FIGURE 5 | (Continued)

related to the extent to which they allocated addresses and therefore individuals to the same new zone. Greater difference in the allocation accompanied by increased adherence to thresholds, homogeneity, and compactness could indicate that the new zones were relatively more demographically and socio-economically coherent. In contrast if the difference between the two methods for defining acceptable new zones (from a confidentiality perspective) is minimal then it could be concluded that continuation of their creation from postcode areas is reasonable. It should be stated that the results derive from the application of the two approaches to historical datasets and would not necessarily carry through to similar analysis with contemporary data.

Across the study areas the mean and standard deviation of the population counts were lower or the same for zones created from address-based Thiessen building blocks compared with the pseudo postcode-based ones. The same outcome occurred for household counts apart from Wembley. Considering differences between London and Middlesex, the latter included four areas where the mean population and household counts in the new zones were identical when expressed as integers. Higher homogeneity values (inter-area correlations) indicate greater homogeneity. Although some of the homogeneity statistics relating to the number of households per address (1, 2, 3, 4 and 5 or more) and the numbers of addresses in the different accommodation types (detached, semi-detached, terraced, flat or other) are relatively small, those for the sets of address-based building block zones were invariably higher. For example, in Deptford the number of households and accommodation type figures were 0.204 and 0.171 compared with 0.119 and 0.061 for the pseudo postcode-based zones. Lower mean compactness scores indicated zone sets tending towards circularity, whereas higher values indicate rather more elongated and irregular shapes (Martin et al. 2001). It is unsurprising that the new zones aggregated from address-based building blocks should have higher compactness mean scores as they start from stretched out angular building blocks (see Figure 4) whereas the pseudo postcode building blocks already related to a group of at least partially dispersed address points. The highest difference between mean compactness scores occurred in Deptford (15.59) and there were low scores in Hampton Wick (2.09), and South Mimms was the only area where the new zones from address-based building blocks were slightly more compact that the pseudo postcode set, which may be attributed to the comparatively dispersed location of most households in this largely rural locality.

Figures 5a, b (London and Middlesex counties respectively) continue to illustrate the output from the workflow sequence outlined in Figure 2 by showing the new zones aggregated from the address- and pseudo postcode-based Thiessen building blocks in the top and middle rows respectively. The third series in each part of Figure 5 show the boundaries used in the 1911

TABLE 2 | Statistical characteristics of new zones created from address-based and pseudo postcode-based buildings blocks in London and Middlesex study areas, 1911.

	New zones created from Thiessen polygons for		New zones created from Thiessen polygons for	
	Pseudo postcodes	**Addresses**	**Pseudo postcodes**	**Addresses**
	City of London		**Hayes UD**	
Building block count	632	2789	149	839
New zone count	44	46	11	11
Total population	427 (145.2)	411 (104.9)	385 (32.3)	385 (60.3)
Total households	87 (20.9)	80 (29.3)	80 (10.3)	80 (12.9)
Homogeneity				
Households per address	0.07	0.09	0.04	0.06
Accommodation type	0.07	0.11	0.03	0.04
Mean shape score	26.2	33.3	31.0	38.7
	Deptford MB		**South Mimms RD**	
Building block count	828	16095	124	615
New zone count	280	260	7	7
Total population	417 (127.1)	387 (30.4)	396 (35.0)	396 (18.4)
Total households	96 (29.6)	89 (10.1)	90 (9.2)	90 (7.7)
Homogeneity				
Households per address	0.12	0.20	0.06	0.07
Accommodation type	0.06	0.17	0.02	0.02
Mean shape score	24.7	40.3	28.7	27.3
	Greenwich UD		**Stoke Newington UD**	
Building block count	924	14883	584	7329
New zone count	239	235	127	58
Total population	404 (127.5)	386 (58.3)	383 (56.2)	383 (21.9)
Total households	86 (19.7)	86 (16.0)	90 (16.7)	90 (14.2)
Homogeneity				
Households per address	0.17	0.22	0.16	0.17
Accommodation type	0.07	0.09	0.08	0.12
Mean shape score	25.0	35.0	26.2	37.9
	Hammersmith MB		**Sunbury on Thames UD**	
Building block count	970	15335	181	972
New zone count	296	302	12	12
Total population	400 (150.9)	389 (82.3)	383 (15.2)	381 (19.2)
Total households	94 (31.9)	90 (14.5)	85 (7.2)	85 (6.1)
Homogeneity				
Households per address	0.18	0.18	0.04	0.06
Accommodation type	0.11	0.16	0.03	0.00
Mean shape score	24.6	36.9	25.7	35.1
	Hampton Wick UD		**Wembley MB/UD**	
Building block count	70	489	181	2180
New zone count	6	6	28	28
Total population	395 (32.0)	395 (22.4)	383 (29.2)	381 (17.7)
Total households	90 (9.6)	90 (8.0)	85 (9.6)	87 (7.7)

(Continues)

TABLE 2 | (Continued)

	New zones created from Thiessen polygons for		New zones created from Thiessen polygons for	
	Pseudo postcodes	Addresses	Pseudo postcodes	Addresses
Homogeneity				
Households per address	0.06	0.06	0.05	0.10
Accommodation type	0.00	0.04	0.02	0.05
Mean shape score	31.0	38.7	28.0	36.3

Abbreviations: MB, Metropolitan Borough; RD, Rural District; UD, Urban District.
Number of households at address was 1, 2, 3, 4 or 5 +); accommodation type categories were detached, semi-detached, terraced, flat and other.

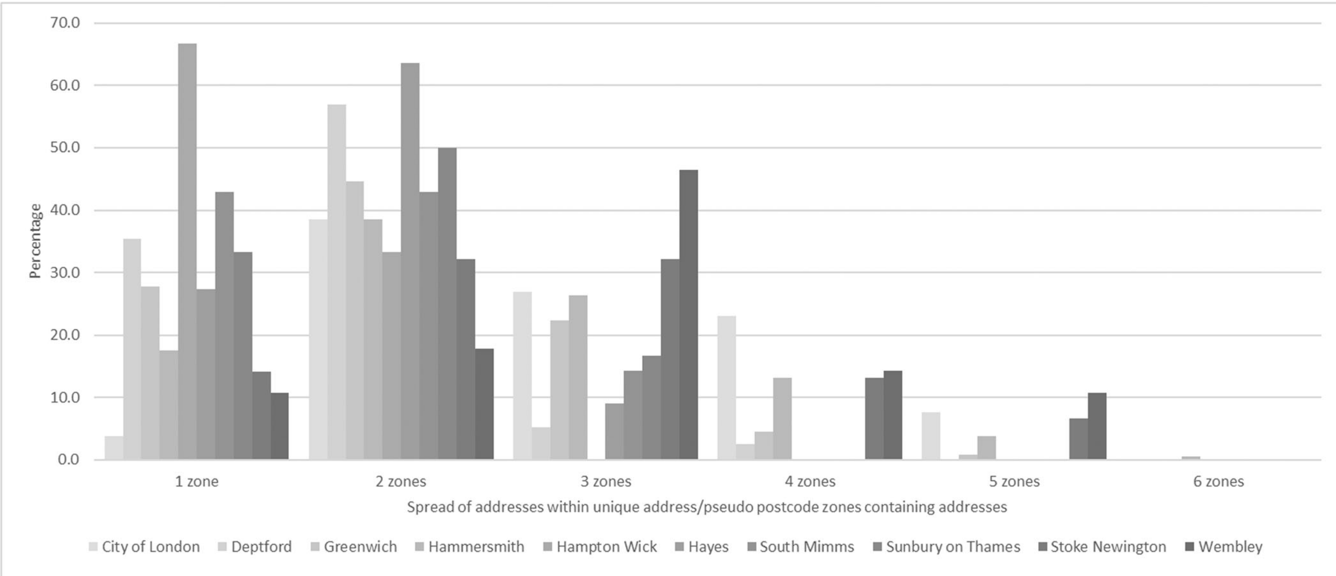


FIGURE 6 | Spreading of addresses across address-based new zones after the latter were intersected and split with new postcode-based zones.

census enumeration districts which were digitised using the detailed descriptions contained in the historical census records. For example, in Sunbury on Thames enumeration district 14 is defined as "On the north by the Civil Parish of Feltham. On the east by Vicarage Road. On the south by the railway. On the west by the Civil Parish of Ashford and the Ecclesiastical Parish of Upper Sunbury St Saviour." (Record identifier: GBC_1911_RG78_00335_0207-DESC). Visual inspection of the three maps for each study area reveals that, even in the simpler Middlesex area, the boundaries of the newly created zones and the enumeration districts coincide only occasionally.

Table 2 compared the results of applying the AZTool to both sets of building blocks in respect of satisfying population and household thresholds, homogeneity and shape compactness. An alternative approach was used in Figure 6, which examines the extent to which individual addresses were aggregated together or separately by the parallel series of analyses. Using the new zones created from the address-based Thiessen building blocks, Figure 6 examines the percentage of addresses that occurred in the same, or 2, 3, 4, 5 or 6 different new zones using the pseudo postcode-based new zones to split the address-based ones. For example, in Hampton Wick 66.7% of addresses were colocated in the same area and 33.3% of addresses were in two areas. This contrasts with Hammersmith where some addresses were spread across six split new address-based zones when they were intersected with the postcode-based ones. The City of London, Stoke Newington and Wembley had the lowest percentages of addresses (< 15.0%) colocated within a single new address-based zone, contrasted with Hampton Wick, South Mimms and Sunbury on Thames with over 30.0%. At least 30.0% of addresses were spread across two new address-based zones in all areas apart from Wembley, which has the most distributed into three, four and five zones.

The impact of addresses being spread across the split address-based zones is examined further by considering the effect on the population and household totals. Figure 7 provides an example of the process for a small area in the south of Stoke Newington. The upper and lower parts of Figure 7a show the grid reference point locations of 124 addresses and the mean grid points for seven pseudo postcodes together with the Thiessen polygons created around them. Similarly, the upper and lower parts of Figure 7b show the new zones created as a result of using the AZTool with the address- and postcode-based Thiessen polygons as building blocks (note: the address points shown have been limited to those occurring in five pseudo postcode-based zones.). Figure 7c shows the outcome of intersecting the two

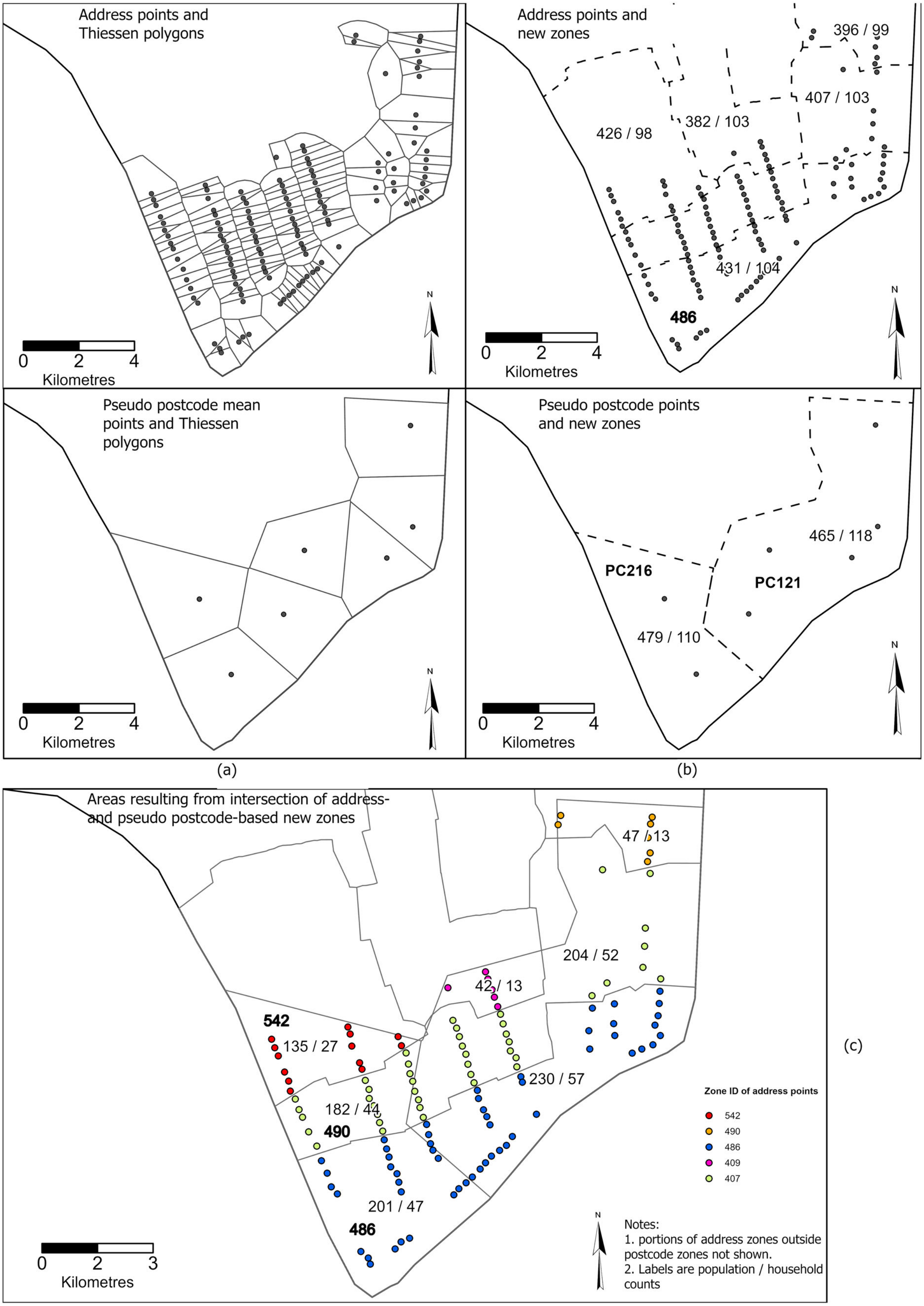


FIGURE 7 | The impact of addresses being spread across the split address-based zones. Note: the address points shown have been limited to those occurring in five pseudo postcode-based zones. (a) Address grid reference points, (b) New zones around grid points, (c) Distribution of population and households to split new zones. Source: Ordnance Survey, Edina. © Crown copyright and/or database right 2022 OS.

(a)

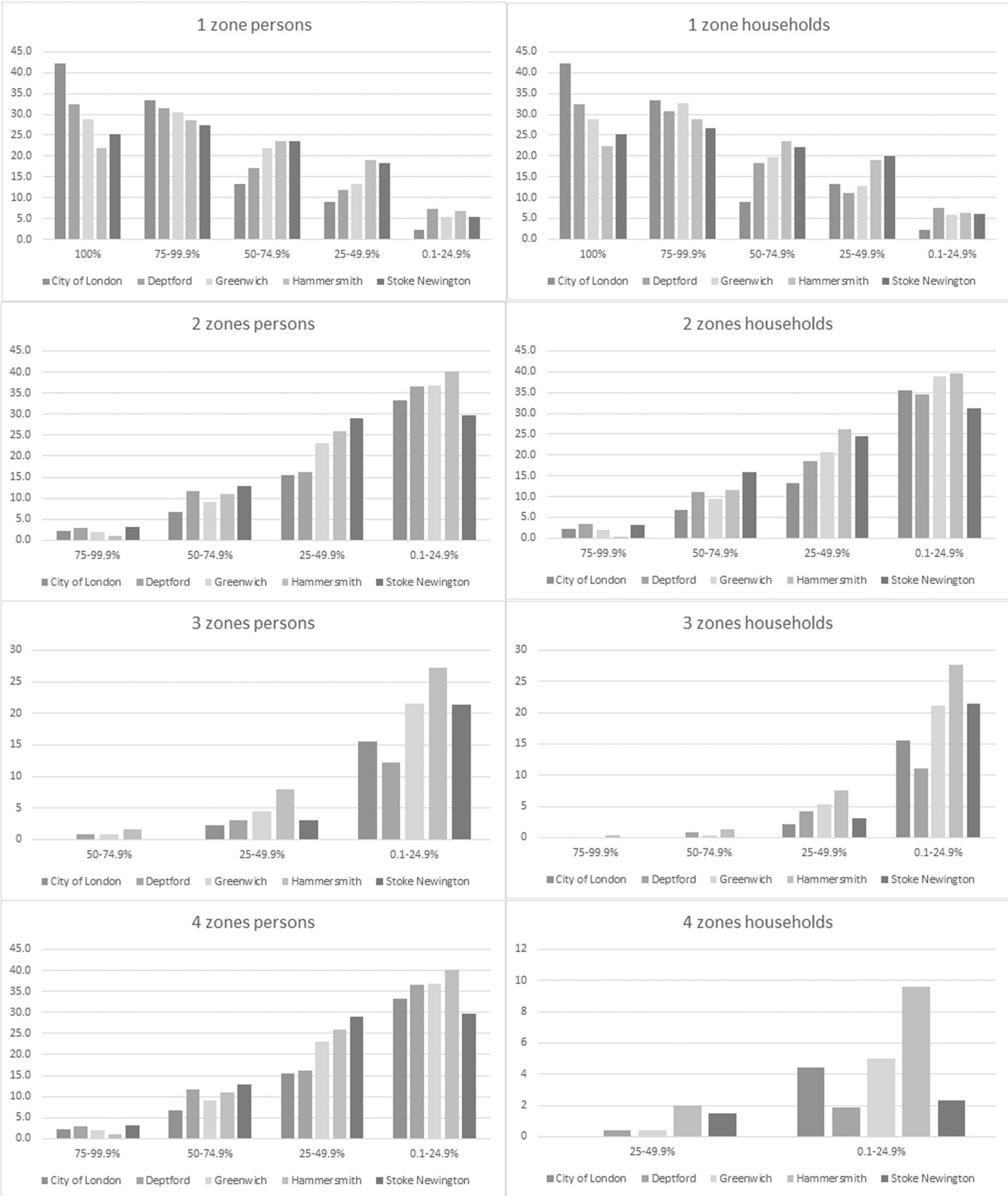


FIGURE 8 | Spreading of addresses across address-based new zones in each of the 10 study areas in the Counties of London and Middlesex, 1911. (a) London. (b) Middlesex.

sets of new zones and how some individual address points in one address-based zone were located together whereas others were assigned to a different zone. For example, the southern-most address-based zone (numbered 486 in Figure 7b) had 431 people and 54 addresses (104 households). The addresses (and households) in zone 486 were spread across two split portions when it was intersected with the pseudo postcode zone. These portions have been labelled as 201/47 and 230/57 in Figure 7c to indicate how the population and households were distributed. The pseudo postcode-based zone, labelled PC216 in Figure 7b,

(b)

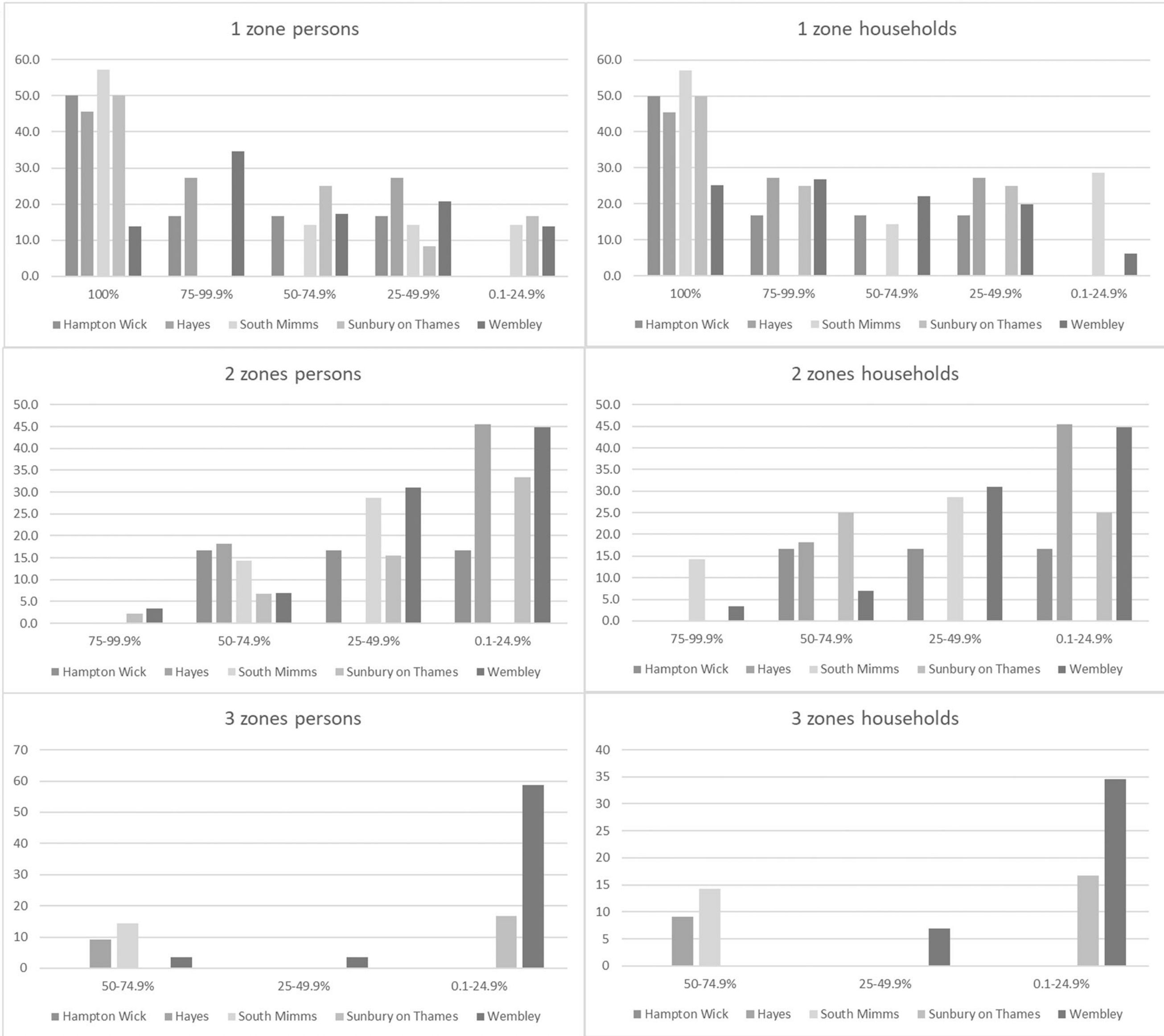


FIGURE 8 | (Continued)

contained people and households in addresses from three address-based zones (542, 490, and 486) with respectively 135/27, 182/44 and 201/47 people and households (total 479/110). The greater the 'mixing up' of people and households indicates the higher the level of inconsistency between the two new zone systems.

The outcome of this process is examined in Figure 8a (London areas) and 8b (Middlesex areas). The first group of columns in each of the top rows of the charts shows the percentage of address-based new zones where 100.0 per cent people and households were colocated irrespective of splitting the two zonal systems. The remaining groups of columns in the top row show the percentages of people and households in the first or reference address-based zone classified into approximately quartile classes. For example, just over 30.0% of the address-based zones in the City of London, whose population was spread across more than one area, were inhabited by between 75.0 and 99.9 % of their population. This implies that less than 25.0% were in one other zone. The second, third and fourth rows (fourth row not required in Figure 8b) shows the percentages of people and households, who were not colocated in one zone, classified in the same way spread across two, three and four new split address-based zones portions. The charts for households are broadly similar to those for population counts, although some differences such as a switch in percentages for Hammersmith and Stoke Newington in respect of where the spreading of people and households was over two zones inhabited by 25.0–49.9 of their respective totals. Considering the outcome of the zoning from the perspective of the percentages of people and households spread from their address-based new zone, the most dispersed (spread) study areas were Greenwich, Hammersmith, Stoke Newington and Wembley with over 70.0% of their population and households spread in this way, although the percentage of Stoke Newington's households was smaller. The lowest percentages were for Hampton Wick, South Mimms and Sunbury on Thames (less than 55.0% for both population and households.

TABLE 3 | Comparison of two new zoning methods using Spearman's Rank correlation between a group of census counts against persons per hectare.

	Address-based Zones	Postcode-based Zones	Absolute difference	Address-based Zones	Postcode-based Zones	Absolute difference
		City of London			**Hampton Wick UD**	
Persons aged 25–44	−0.111	0.067	0.178	−0.143	0.600	0.743
Detached housing	0.048	0.118	0.070	−0.257	−0.143	0.114
Terraced or end of terrace housing	0.014	0.270	0.256	0.371	−0.029	0.400
Living at > 1.5 persons per room	0.347*	0.550**	0.203	0.580	0.395	0.185
Households without children	−0.218	0.090	0.308	0.029	0.029	0.000
		Deptford MB			**Hayes UD**	
Persons aged 25–44	−0.347**	−0.030	0.317	0.345	−0.638*	0.983
Detached housing	−0.340**	−0.204**	0.136	−0.724*	−0.205	0.519
Terraced or end of terrace housing	0.047	0.194**	0.147	0.853**	0.218	0.635
Household living at > 1.5 persons per room	0.557**	0.489**	0.068	0.731*	−0.059	0.790
Households without children	−0.322**	−0.136*	0.186	0.274	0.174	0.100
		Greenwich UD			**South Mimms RD**	
Persons aged 25–44	−0.256**	−0.320**	0.064	0.071	−0.607	0.678
Detached housing	−0.537**	−0.564**	0.027	−0.750	−0.893*	0.143
Terraced or end of terrace housing	0.193**	0.030	0.163	0.357	0.714	0.357
Household living at > 1.5 persons per room	0.422*	−0.507**	0.929	0.306	0.164	0.142
Households without children	−0.379**	−0.353**	0.026	0.468	−0.500	0.968
		Hammersmith MB			**Sunbury on Thames UD**	
Persons aged 25–44	−0.180**	−0.175**	0.005	0.133	0.636*	0.503
Detached housing	−0.468**	−0.433**	0.035	−0.109	0.510	0.619
Terraced or end of terrace housing	0.297**	0.074	0.223	0.322	0.315	0.007
Household living at > 1.5 persons per room	0.534**	0.406**	0.128	−0.049	0.677*	0.726
Households without children	−0.002	−0.124*	0.122	−0.261	0.690*	0.951

(Continues)

TABLE 3 | (Continued)

	Address-based Zones	Postcode-based Zones	Absolute difference	Address-based Zones	Postcode-based Zones	Absolute difference
	Stoke Newington UD			Wembley MB/UD		
Persons aged 25–44	0.094	0.406**	0.312	0.425*	0.638**	0.213
Detached housing	−0.126	0.233**	0.359	−0.036	0.176	0.212
Terraced or end of terrace housing	0.070	0.300**	0.230	0.075	0.430*	0.355
Household living at > 1.5 persons per room	0.063	0.394**	0.331	0.129	0.449*	0.320
Households without children	0.116	0.376**	0.260	−0.145	0.331	0.476

Abbreviations: MB, Metropolitan Borough; RD, Rural District; UD, Urban District.

An indication of the difference between the two zoning methods as a means of allocating households and individuals to new areas for statistical aggregation purposes has been carried out using Spearman's Rank correlation. Differences in the number of new zones produced by each method (see Table 2) means they cannot be correlated with each other; however, a group of typical census counts was each correlated against persons per hectare. The detached and terraced or end of terrace type housing counts have been obtained from inspection of Ordnance Survey maps showing the georeferenced addresses and the other three counts, persons aged 25–44, households living at > 1.5 persons per room and households without children have been determined for the 1911 Population Census records themselves. The correlation analysis results are summarised in Table 3, which reveals the coefficients were similar for some of the local authority areas, but in other cases there was a difference in the strength and/or direction of the relationship. For example, the number of households in detached properties in the City of London and Hampton Wick were respectively 0.048 and 0.118, and −0.257 and −0.143 for the new zones produced by both methods. In contrast, there was a strong positive nonsignificant correlation (0.600) for persons aged 25–44 in the postcode-based new zones in Hampton Wick and a low negative correlation (−0.143) for the address-based new zones. However, there was a moderately strong and significant correlation (−0.322) for address-based new zones in Deptford compared with a weaker but still negative correlation for the postcode-based ones. Considering the absolute difference between pairs of Spearman's correlation coefficients across the local authorities the majority of larger values were in the more rural Middlesex areas with the notable exception of Greenwich in respect of overcrowded households with a moderate positive correlation (0.422) for the address-based zones and a slightly stronger negative one (−0.507) for the postcode-based ones.

5 | Discussion and Conclusions

This paper has returned to a perennial question facing those concerned with designing and creating sets of zones that can be used as statistical reporting units from national census enumerations, while at the same time preserving confidentiality thresholds to avoid the unintended and prevent the deliberate release of personal information about individuals and households. Specifically, it has used historical census records from the British Population Census in 1911 for a group of 10 study areas in the former counties of London and Middlesex. The areas range from the relatively sparsely inhabited South Mimms on the border with Hertfordshire in the north to the already well-populated metropolitan boroughs of Deptford, Greenwich and Hammersmith towards the centre and south-east of the overall area covered by the two counties. It has compared two approaches to creating zones, one from building blocks generated as Thiessen polygons around individual addresses and the other building blocks of Thiessen polygons created around the mean grid reference for groups of 1911 addresses sharing the same modern postcode The AZTool zoning tool was applied in each case. The results of the re-zoning were assessed in three main ways: first, by examining population and household means and standard deviations, inter area correlation and compactness measures; second, by investigating the extent to which addresses in one set of new were located together or separately by the two zoning approaches; and third by correlation of aggregate counts of census statistics.

The address-based new zones were somewhat less compact than those produced from building blocks that were already grouped around a collection of addresses, whose group's location was determined as the mean grid reference of the individual addresses. The new zones based on the aggregation of address building blocks generally resulted in higher levels of homogeneity in respect of the attributes assessed (apart from South Mimms), although the IAC values were lower than achieved in some studies (Mokhele et al. 2016). The mean and standard deviation of their population and household counts (apart from Wembley in the case of households) were slightly smaller or the same as those from the pseudo postcode building blocks. In summary the results indicate the feasibility of creating zones around addresses and improving socio-economic homogeneity while retaining the population and household thresholds to preserve confidentiality. Differences in the correlation coefficients for a selection of typical census counts in relation to population density (persons per hectare) were varied, but offered some evidence that the two methods produced more divergent outcomes in comparatively less population rural districts in the former Middlesex County.

Some 100 years after the UK's 1920 Census Act, which formed the primary legislation governing the British over this period, we moved into an era that Coleman (2013) termed "the twilight of the Census". The drivers for change approach from different directions, but leading the field are two issues: the rising cost of holding a traditional census enumeration, even at a time when online data collection from the large majority of the population has already become a reality; and the insatiable desire for more frequent, comprehensive information about the population than can be offered by a 'once every 10 years' national census. The final appearance and shape of the outcome of the transition from the traditional census to some form of 'big data' assembly of (hopefully) equivalent statistical information has yet to be determined. Unless there is a drive to dispense with the production of detailed area statistics created on a consistent and reliable national basis capable of being connected over time, there will remain a need to define the boundaries of relatively small areas without risking breaches of confidentially that have been a constant undercurrent in the work of the census authorities. Administrators, researchers, policy-makers and practitioners in public, commercial, and third sector organisations have become accustomed to using such information over the course of the last 50 years and this seems unlikely to diminish. The present article has indicated the feasibility of starting from Thiessen polygons around addresses to create zones that comply with established population and household thresholds.

Acknowledgements

The author gratefully acknowledges the use of the AZTool software, which is copyright David Martin, Samantha Cockings and the University of Southampton.

The work on which this article is based was carried out under a British Academy, Senior Research Fellowship (SF110183) entitled *The Geography of London's Population: Populating Places with Historical Census Statistics*.

OS MasterMap Address Layer 2 was used under Academic Research Licence. Ordnance Survey ©Crown Copyright. All rights reserved.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analysed in this study.

References

Royal Mail. nd. Royal Mail postcodes explained. https://www.postcodeaddressfile.co.uk/products/postcodes/postcodes_explained.htm.

Cockings, S., A. Harfoot, D. Martin, and D. Hornby. 2011. "Maintaining Existing Zoning Systems Using Automated Zone-Design Techniques: Methods for Creating the 2011 Census Output Geographies for England and Wales." *Environment and Planning A: Economy and Space* 43: 2399–2418. https://doi.org/10.1068/a43601.

Cockings, S., A. Harfoot, D. Martin, and D. Hornby. 2013. "Getting the Foundations Right: Spatial Building Blocks for Official Population Statistics." *Environment and Planning A: Economy and Space* 45: 1403–1420. https://doi.org/10.1068/a45276.

Cockings, S., and D. Martin. 2005. "Zone Design for Environment and Health Studies Using Pre-Aggregated Data." *Social Science & Medicine* 60: 2729–2742. https://doi.org/10.1016/j.socscimed.2004.11.005.

Coleman, D. 2013. "The Twilight of the Census." *Population and Development Review* 38: 334–351. https://www.jstor.org/stable/23655303.

Duke-Williams, O., and P. Rees. 1998. "Can Census Offices Publish Statistics for More Than One Small Area Geography? An Analysis of the Differencing Problem In Statistical Disclosure." *International Journal of Geographical Information Science* 12: 579–605. https://doi.org/10.1080/136588198241680.

Dumedah, G., N. Schuurman, and W. Yang. 2008. "Minimizing Effects of Scale Distortion for Spatially Grouped Census Data Using Rough Sets." *Journal of geographical systems* 10: 47–69. https://doi.org/10.1007/s10109-007-0056-y.

Duncan, J. B. 1978. Regional Highway Traffic Model: Non-home Based Trip End Model. Proceeding of the Summer Annual Meeting, University of Warwick, England, July, 1978. Transportation Research Board. Planning and Transport Research and Computation Co Ltd, London, pp. 158–167.

Duque, J. C., R. Ramos, and J. Suriñach. 2007. "Supervised Regionalization Methods: A Survey." *International Regional Science Review* 30: 195–220. https://doi.org/10.1177/0160017607301605.

Flowerdew, R. 2011. "How Serious Is the Modifiable Areal Unit Problem for Analysis of English Census Data?" *Population Trends* 145: 106–118. https://doi.org/10.1057/pt.2011.20.

Flowerdew, R., D. J. Manley, and C. E. Sabel. 2008. "Neighbourhood Effects on Health: Does It Matter Where You Draw the Boundaries?" *Social Science & Medicine* 66: 1241–1255. https://doi.org/10.1016/j.socscimed.2007.11.042.

Hayes, J., R. Dymond-Green, J. Stillwell, and V. Moody. 2018. "UK Census Aggregate Statistics: Characteristics and Access." In *the Routledge Handbook of Census Resources, Methods and Applications*, edited by J. Still. Abingdon: Routledge.

Haynes, R., K. Daras, R. Reading, and A. Jones. 2007. "Modifiable Neighbourhood Units, Zone Design and Residents' Perceptions." *Health & Place* 13: 812–825. https://doi.org/10.1016/j.healthplace.2007.01.002.

Higgs, E., R. Jones, K. Schürer, and A. Wilkinson (2013). Integrated Census Microdata (I-CEM) Guide. http://www.essex.ac.uk/history/research/ICeM/documents/icem_guide.pdf. Accessed 29 January 2016.

Martin, D. 1998a. "2001 Census Output Areas: From Concept to Prototype." *Population Trends* 94: 19–24.

Martin, D. 1998b. "Optimizing Census Geography: The Separation of Collection and Output Geographies." *International Journal of Geographical Information Science* 12: 673–685. https://doi.org/10.1080/136588198241590.

Martin, D. 2002. "Geography for the 2001 Census in England and Wales." *Population Trends* 108: 7–15.

Martin, D. 2003. "Extending the Automated Zoning Procedure to Reconcile Incompatible Zoning Systems." *International Journal of Geographical Information Science* 17: 181–196. https://doi.org/10.1080/713811750.

Martin, D. 2004. "Neighbourhoods and Area Statistics in the Post 2001 Census Era." *Area* 36: 136–145. https://doi.org/10.1111/j.0004-0894.2004.00210.x.

Martin, D. 2010. "Understanding the Social Geography of Census Undercount." *Environment and Planning A: Economy and Space* 42: 2753–2770. https://doi.org/10.1068/a43123.

Martin, D., A. Nolan, and M. Tranmer. 2001. "The Application of Zone-Design Methodology in the 2001 UK Census." *Environment and Planning A: Economy and Space* 33: 1949–1962. https://doi.org/10.1068/a3497.

Mokhele, T., O. Mutanga, and F. Ahmed. 2016. "Development of Census Output Areas With Aztool in South Africa." *South African Journal of Science* 112, no. 7/8: 7. https://doi.org/10.17159/sajs.2016/20150010.

Norman, P., J. Colbert, and D. J. Exeter. 2023. "Linking Individuals to Areas: Protecting Confidentiality While Preserving Research Utility." *Spatial Demography* 11: 10. https://doi.org/10.1007/s40980-023-00121-9.

Norman, P., P. Rees, and P. Boyle. 2003. "Achieving Data Compatibility over Space and Time: Creating Consistent Geographical Zones." *International Journal of Population Geography* 9: 365–386. https://doi.org/10.1002/ijpg.294.

Openshaw, S. 1977. "A Geographical Solution to Scale and Aggregation Problems in Region Building, Partitioning and Spatial Modelling." *Transactions of the Institute of British Geographers* 2: 459–472. https://doi.org/10.2307/622300.

Ralphs, M., and L. Ang 2009 Optimized Geographies for Data Reporting: Zone Design Tools for 4 Census Output Geographies. Statistics New Zealand Working Paper No. 09-01. Wellington: Statistics New Zealand.

Raper, J., D. Rhind, and J. Shepherd. 1992. *Postcodes: The New Geography*. Longman Scientific & Technical.

Sabel, C. E., W. Kihal, D. Bard, and C. Weber. 2013. "Creation of Synthetic Homogeneous Neighbourhoods Using Zone Design Algorithms to Explore Relationships Between Asthma and Deprivation in Strasbourg, France." *Social Science & Medicine* 91: 110–121. https://doi.org/10.1016/j.socscimed.2012.11.018.

Tait, A. (ed.) 2012. Changes to Output Areas and Super Output Areas in England and Wales, 2001 to 2011. ONS. Available at https://docslib.org/doc/1777591/changes-to-output-areas-and-super-output-areas-in-england-and-wales-2001-to-2011#:~:text=Of%20the%20175%2C434%20output%20areas%20created%20for%20the,super%20output%20areas%20%28MSOAs%29%20in%20England%20and%20Wales. Accessed 07 November 2022.

Tranmer, M., and D. G. Steel. 1998. "Using Census Data to Investigate the Causes of the Ecological Fallacy'." *Environment and Planning A: Economy and Space* 30: 817–831. https://doi.org/10.1068/a300817.

UK Meter Franking. 2017 Section 3: postcode publicity. http://meterfranking.co.uk/part2section_31.pdf.

Walford, N. S. 2019. "Bringing Historicalbritish Population Censusrecords Into the 21st Century: A Method for Geocoding Households and Individuals At Their Early-20th-Century Addresses." *Population, Space and Place* 25: e2227. https://doi.org/10.1002/psp.2227.

Walford, N. S., and K. N. Hayles. 2012. "Thirty Years of Geographical (In)Consistency in the British Population Census: Steps Towards the Harmonisation of Small-Area Census Geography." *Population, Space and Place* 18, no. 3: 295–313. https://doi.org/10.1002/psp.658.

Webber, R., and R. Burrows. 2018. *The Predictive Postcode: The Geodemographic Classification of British Society*. London: Sage Publishing.