Proceedings Paper:

Solomon, G. orcid.org/0000-0002-4394-1498 (2025) Estimating urban areas using historical census data and the H3 grid. In: 33rd Annual Geographical Information Science Research UK Conference (GISRUK2025). GISRUK 2025 (Geographical Information Science Research UK), 23-25 Apr 2025, Bristol, United Kingdom.

https://doi.org/10.5281/zenodo.15229219

Estimating urban areas using historical census data and the H3 grid

Guy Solomon*[1]

[1]Digital Humanities Institute, School of History, Philosophy and Digital Humanities, University of Sheffield

GISRUK 2025

Summary

Historical (1851-1911) urban areas in England and Wales have often been derived from population data, which typically utilise census administrative units. However, these frequently fail to conform to the spatial boundaries of urban settlements. Recent developments in the geocoding of individuals enumerated by the census to a street level now facilitate a re-evaluation of this approach. Utilising Uber's H3 grid, this paper outlines a method for estimating urban areas at each historical decennial census, using 1901 as a case study.

KEYWORDS: urban settlements, historical population, historic census, H3 grid

1 Introduction

Several methods exist for defining historical urban areas, typically based on population (Robson, 1973), built-up areas (Litvine et al., 2024), or function (Royle, 2001). Rapid nineteenth-century urbanisation had significant economic, social, and demographic impacts, making city development crucial to understanding this period (Williamson, 1990; Gunn, 2004; Day, 2023).

Population-based approaches have traditionally concentrated on number of people over spatial coverage, but recent advances in historical census data enable a new approach. This promises new avenues of investigation regarding the internal structure and composition of urban spaces (for example, residential clustering of particular occupations or demographics), the expansion and permeability of urban-rural boundaries over time, and the nature of urban growth (such as densification). This paper explores the 1901 census as a test case, as it offers comparative datasets—particularly Satchell et al. (2017)—for validation, before extending the method to years without such coverage.

2 Background

An urban classification exists for individual-level census data. However, the urban coding scheme from Smith and Bennett (2017) and Smith et al. (2018), as well as earlier works (Law, 1967; Robson,

*g.solomon@sheffield.ac.uk

1973), has been based on census parishes—tying it to administrative geographies. This is significant because, despite increasing efforts over the nineteenth century, parish boundaries did not always align with settlements. Moreover, parish geographies have only been reconstructed for the 1851 census—–later years require the use of registration sub-districts (larger units containing multiple parishes) for spatial analysis, limiting their ability to identify the spatial extent of urban areas (see Figure 1).

Figure 1: An example from Rotherham illustrating how census administrative boundaries do not align with settlement footprints and may include multiple settlements. OS 6-inch to the mile map reproduced with permission from the National Library of Scotland (CC-BY). Simplified administrative boundaries (registration sub-districts) from Reid et al. (2018).

Alternative approaches have used historical maps to identify built-up areas (Satchell et al., 2017), but until recent developments (see, for a different period: Litvine et al., 2024), this was a manual, time-consuming, and somewhat subjective process. Additionally, spatio-temporal coverage issues make consistent comparisons difficult. A population-based approach using the decadal census provides an alternative metric at regular intervals, easily linked to demographics and occupations.

Recent census research aims to move beyond administrative geographies for greater flexibility. It focuses on relating the street names collected by census enumerators to georeferenced datasets (Lan and Longley, 2019; Rhodes, 2024, 2025), enabling the creation of finer spatial units not found directly in the census.

3 Data and Methodology

This paper uses the geocoded census data from Rhodes (2024), which links individuals in the 1901 census to three geographic features: a) modern street vectors (assumed to reflect historical counterparts) from Ordnance Survey's Open Roads dataset (Ordnance Survey, 2022); b) historical street names from the GB1900 Gazetteer (Aucott and Southall, 2019), providing point data (for the map label) from OS 6-inch maps; c) a polygon combining two commonly-used geographic boundary sets from higher-level census units (Day, 2016; Satchell et al., 2018). All individuals in mainland England and Wales (with minor exceptions) are coded to c), and may also be coded to a), b), or both. For details on the process and matching rates, see Rhodes (2024). For this paper, each individual is assigned a single geographic feature, prioritizing a) over b), and b) over c).

Including different types of spatial features presents a methodological challenge for identification of urban areas, which typically includes some form of population density requirement. To address this problem, this paper aggregates Rhodes (2024) into Uber's H3 grid, a hexagonal global indexing system (Uber Technologies, Inc, 2025). At resolution 8, the H3 grid provides hexagons averaging 0.74 km^2, ensuring both sufficient granularity and a consistent unit of aggregation.

Individuals linked to a point are assigned to their enclosing hexagon. Those coded to a street geometry are distributed across hexagons based on the proportion of street length within each hexagon. For individuals only coded to a polygon (typically larger than hexagons), allocation is based on the proportion of polygon area within each hexagon. While this method may blur urban boundaries, especially for larger polygons, it significantly reduces area size compared to geolocating existing urban classifications.

Urban areas are defined as those containing a population greater than a specified number of individuals per hex. Law and Robson both used a value of one person per acre to offset large rural areas within 'urban' parishes, and directly acknowledge this as lower than expected for purely urban areas (Law, 1967, p.129; Robson, 1973, p.48). The appropriate value for this period is therefore examined in the next section. Hexagons are grouped into contiguous settlements based on shared edges, and these groups are linked to named urban units by intersecting with Satchell et al. (2017)'s manually digitised boundaries.

4 Results and analysis

The methodology from the previous section, applied at stepped population thresholds, produces the data in Figure 2. This is assessed by total population, area, and number of settlements. Including each of these metrics is useful given existing work. The 1901 Census Report provides contemporary totals (Dunbar et al., 1904). Smith et al. (2018) currently offers the best urban population estimates, but without lower-level spatial units, urban areas may be overstated when combined with Day (2016). Satchell et al. (2017) provides the most precise urban footprints, though their 1901 'urban' status is uncertain, and settlement limits remain open to interpretation. Settlements above 250,000, 50,000, and 10,000 people are included, reflecting classifications by nineteenth-century commentators and recent literature.

	Number of settlements				Area, km²				Population in settlements			
	>250,000 population	>50,000 population	>10,000 population	No specified minimum population	>250,000 population	>50,000 population	>10,000 population	No specified minimum population	>250,000	>50,000	>10,000	No specified minimum
100/hex	9	51	170	5,142	7,777	9,998	12,730	19,608	17,979,663	21,998,326	24,641,157	27,681,841
250/hex	10	49	193	2,532	3,277	4,563	6,256	9,365	14,676,508	18,868,786	22,261,555	25,294,824
350/hex	10	47	194	1,959	2,616	3,573	4,954	7,281	14,046,096	17,992,392	21,413,333	24,368,218
500/hex	10	52	197	1,553	1,775	2,789	3,859	5,619	12,333,278	17,158,931	20,443,170	23,323,482
750/hex	9	53	194	1,187	1,356	2,211	2,975	4,153	11,561,939	16,348,187	19,408,767	21,965,634
1000/hex	8	46	178	950	863	1,411	2,033	2,878	8,075,214	11,796,809	14,728,272	17,013,565
1901 Census Report[1]	9	75	436	1,122	786	2,314	7,723	15,576	7,973,406	14,506,863	21,959,998	25,058,355
Based on Smith-Law-Robson urban coding of parishes with Day's boundaries[2]	9	67	362	.	2,131	7,594	24,983	.	10,560,049	16,398,587	22,447,021	.
Satchell et al[3]	.	.	.	1,746/1,607[4]	.	.	.	2,230	.	.	.	.

Figure 2: Comparison of population thresholds to existing estimates.

Notes:

[1]*Beyond the 75 'Great' towns, the Census Report lists urban districts rather than towns. Space constraints limit discussion, but this likely overestimates the number of smaller towns.*

[2]*These figures result from combining Smith et al.'s urban coding of individual-level census returns with Day's registration sub-district boundaries, which significantly overstates urban areas. The PopulationsPast project also classifies Day's boundaries by 'type of place', but this is not urban-rural focused.*

[3]*Satchell et al. provide a separate database with population figures derived from multiple sources. For 1901, this is primarily based on Law-Robson.*

[4]*Satchell et al.'s documentation lists 1,783 settlements, but the GIS file contains 1,607 polygons. Area calculations are based on the latter figure.*

The results suggest that a population density of between 500/hex and 750/hex provide comparable coverage in terms of population, whilst providing a much finer spatial area. This choice is influenced by smaller settlements (greater than 10,000), which are prone to inflated populations due to mixed-area administrative units. Populations of larger settlements seem overestimated, likely due to boundary disambiguation where settlements meet. While area calculations lack the specificity of Satchell et al. (2017), this is expected given hexagon size relative to settlement footprints.

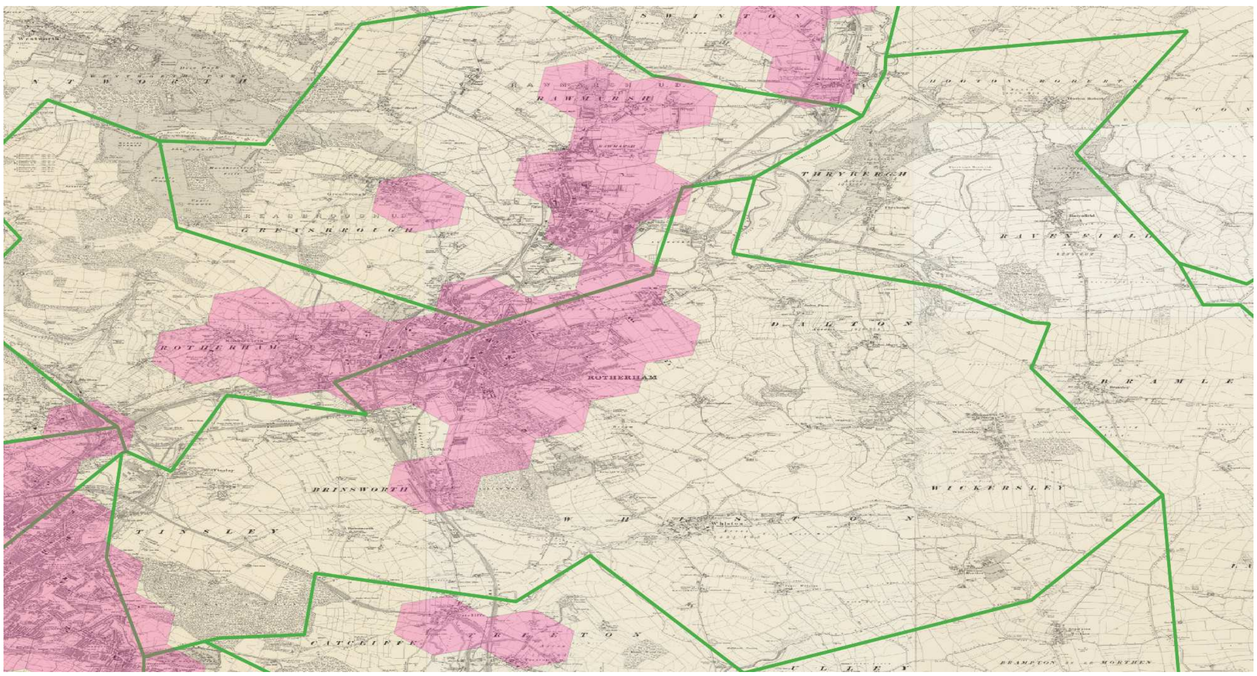

Figure 3: An example of revised settlement footprints, at 750/hex. OS 6-inch to the mile map reproduced with permission from the National Library of Scotland (CC-BY). Simplified administrative boundaries (registration sub-districts) from Reid et al. (2018).

5 Limitations and future work

The process of assigning parts of lines and polygons to the hexagonal grid is imperfect for several reasons. First, it assumes the population was distributed evenly over the former, which is unlikely to have been the case. Second, it may exclude urban areas which fall partially between hexagons. Both these problems can be addressed by relating individuals to a single geographic feature (i.e. road vectors), which will be addressed in future work.

Disambiguation of contiguous urban settlements (especially, conurbations) also poses a significant challenge, and requires further evaluation than is possible within the space constraints here. Further development of integrating different population density values provides a potential solution.

6 Acknowledgements

The author wishes to thank Joshua Rhodes for making an in-progress version of the census street-level data available prior to the publication of Rhodes (2024, 2025), Max Satchell for providing access to Satchell et al. (2017), as well as Joe Day and Alice Reid for supplying the registration sub-district boundary files (Day, 2016) which are in the process of being deposited with the UKDS. Some elements of this work were enabled by the Living with Machines project (AHRC grant AH/S01179X/1).

7 Biography

Guy Solomon is a lecturer in AI and Digital Humanities at the University of Sheffield. He is particularly interested in the dynamics of cities and urban systems in both historical and contemporary contexts.

References

Aucott, P. & Southall, H. (2019). Locating past places in britain: Creating and evaluating the GB1900 gazetteer. *International Journal of Humanities and Arts Computing*, *13*(1-2), 69–94, https://doi.org/10.3366/ijhac.2019.0232 https://doi.org/10.3366/ijhac.2019.0232.

Day, J. (2016). Registration sub-district boundaries for England and Wales 1851-1911 .

Day, J. (2023). Mapping the cultural divides of england and wales: Did the geographies of 'belonging' act as a brake on british urbanisation, 1851-1911? *PloS one*, *18*(5), e0286244–e0286244.

Dunbar, W. C., Humphreys, N. A., Tatham, J., Mundy, A. J., & Bellingham, A. (1904). *General Report with Appendices*. London: PPrinted For His Majesty's Stationery Office, By Darling Son, LTtd., 34-40, Bacon Street, E. Command Number: Cd. 2174; BPP Reference: 1904 CVIII, 1 https://www.visionofbritain.org.uk/census/EW1901GEN/4.

Gunn, S. (2004). *Urbanization*, chapter 14, (pp. 238–252). John Wiley Sons, Ltd https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470997147.ch15.

Lan, T. & Longley, P. (2019). Geo-referencing and mapping 1901 census addresses for england and wales. *ISPRS International Journal of Geo-Information*, *8*(8), https://doi.org/10.3390/ijgi8080320 https://www.mdpi.com/2220-9964/8/8/320.

Law, C. M. (1967). The growth of urban population in england and wales, 1801-1911. *Transactions of the Institute of British Geographers*, (41), 125–143 http://www.jstor.org/stable/621331.

Litvine, A. D., Starzec, A., Younis, R., Faula, Y., Coustaty, M., Shaw-Taylor, L., & Églin, V. (2024). Built-up areas of nineteenth-century britain. an integrated methodology for extracting high-resolution urban footprints from historical maps. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *57*(1), 1–19, https://doi.org/10.1080/01615440.2024.2324163 https://doi.org/10.1080/01615440.2024.2324163.

Ordnance Survey (2022). OS Open Roads https://www.ordnancesurvey.co.uk/products/os-open-roads.

Reid, A. M., Arulanantham, S. J., Day, J. D., Garrett, E. M., Jaadla, H., & Lucas-Smith, M. (2018). Populations past: Atlas of victorian and edwardian population https://www.populationspast.org/.

Rhodes, J. (2024). AddressGB: Geo-coded British Census Addresses, 1851-1911 https://doi.org/10.5281/zenodo.10473597.

Rhodes, J. (2025). Geo-coding addresses in historic british census data: An open methodology. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *0*(0), 1–23, https://doi.org/10.1080/01615440.2024.2431491 https://doi.org/10.1080/01615440.2024.2431491.

Robson, B. T. (1973). *Urban growth : an approach*. London: Methuen.

Royle, S. (2001). *The development of small towns in Britain*, (pp. 151–184). The Cambridge Urban History of Britain. Cambridge University Press.

Satchell, A., Kitson, P., Newton, G., Shaw-Taylor, L., & Wrigley, E. (2018). 1851 England and Wales census parishes, townships and places. [Data Collection].

Satchell, M., Potter, E., Shaw-Taylor, L., & Bogart, D. (2017). Candidate towns of england and wales, c.1563-1911 https://www.campop.geog.cam.ac.uk/research/occupations/datasets/catalogues/other/.

Smith, H. & Bennett, R. (2017). Urban-rural classification using census data, 1851-1911. https://doi.org/10.17863/CAM.15763 https://www.repository.cam.ac.uk/handle/1810/269522.

Smith, H., Bennett, R. J., & Radicic, D. (2018). Towns in victorian england and wales: a new classification. *Urban History*, *45*(4), 568–594, https://doi.org/10.1017/S0963926818000020.

Uber Technologies, Inc (2025). H3 https://h3geo.org/.

Williamson, J. G. (1990). *Coping with city growth during the British Industrial Revolution*. Cambridge: Cambridge University Press.