# Census statistics for Civil Parishes –

# When “best-fitting” just isn’t good enough

Bruce Mitchell[1]

Geography Research Unit

Geography Branch

Research Development and Infrastructure Directorate

Office for National Statistics

January 8, 2015

**Summary**

The Government Statistical Service (GSS) National Statistics Geography Policy published by the Office for National Statistics (ONS) defines the way that statistics for any geography larger than Census Output Area should be generated. For Civil Parishes in England this ‘best fit’ method is not satisfactory.

ONS is therefore examining the relative merits of other methods of aggregating Census data to parishes, especially with reference to grid cells. The outcome of the research will inform a decision on which method to mandate for the 2021 Census.

**KEYWORDS:** Census, ONS, Civil Parish, Eurostat, grid

[1] bruce.mitchell@ons.gov.uk

## 1. Introduction

The Office for National Statistics (ONS) publishes statistics related to the economy, population and society of England and Wales at national, regional and local levels. Data produced by ONS, including a decennial Census (most recently in 2011), inform policies, priorities and allocation of resources.

This paper details research into providing statistics for Civil Parishes in England using an INSPIRE-compliant[2] 1 km grid supplied by Eurostat. This would represent an exception to the Government Statistical Service (GSS) National Statistics Geography Policy (ONS, 2010, 2015) and a significant departure for the provision of Census data, which is supplied using the 'best fit' methodology based on Output Area population-weighted centroids (OAPWC).

Civil Parishes are one of the most commonly-requested geographies on the Neighbourhood Statistics website[3].

## 2. 'Exact fit' and 'best fit'

The smallest geographical area for which Census data are released is the Census Output Area (OA). This geography was developed for the 2001 Census of England and Wales, and was built from the Census data themselves. OAs were built from clusters of adjacent unit postcodes and designed to have similar population sizes.

The GSS Policy defines the way that statistics for any 'higher' geography larger than OA, such as Local Authority Districts (LADs) should be generated. 'Exact estimates', derived directly from the Census households located within them, are calculated for OAs. These estimates are applied to the OAPWC. Statistics for the individual instances of any higher geography are aggregated up from the OAPWCs they contain.

Where the higher geography comprises exact aggregations of OAs e.g. Lower and Middle Layer Output Areas (LSOA, MSOA) or LADs, this is called 'exact fit'. Elsewhere, the boundaries of Census OAs may not align with the higher geography: here, the procedure is known as 'best-fit'.

Best-fit works well for virtually every geography, but not for Civil Parishes in England[4]. These are a heterogeneous mix of ancient and modern, displaying enormous variations in size (0.04 $km^2$ to 256 $km^2$) and population (zero to 75,000) covering 91% of England but with less than 50% of the population. The rest of England is covered by 'unparished areas'. There is a very weak correlation between a parish's area and its population. OAs on the other hand were expressly designed with population thresholds: they are always small in densely-populated areas and large where population is thinly spread. In consequence, the mesh of OAPWCs completely misses 1,140 (10%) of parishes, and it is not possible to generate best-fit estimates for this group despite their combined Census population of nearly 120,000 residents. Whatever their size (the largest is 155 $km^2$), parishes without OAPWCs are all tenuously populated in relation to their surroundings. In some rural areas, clusters of such parishes without publishable data (for example 18% of the parishes in West Oxfordshire) present a

[2] INSPIRE stems from an EU directive and aims to establish an infrastructure for spatial information in Europe to support Community environmental policies.
[3] *https://neighbourhood.statistics.gov.uk/dissemination/*

[4] Wales is entirely covered by an equivalent geography – Communities – but the focus of this work is England.

challenge to the relevant Local Authority.

ONS is looking for an improved methodology for Civil Parishes. Because they are such a mix, it is difficult to develop a single methodology that can be applied consistently across the entire geography. We rejected a variant of the best-fit procedure whereby the statistics for the OAPWC which was closest to the orphan parish's geometric centroid was applied to any 'orphan' Parish without an OAPWC. Both the orphan parish and the OAPWC had to be within the same LAD. But the external OAPWC could represent an OA with a very different statistical character, resulting in wildly inaccurate estimates being generated for the orphan parish: also the statistics for the external OAPWC would serve once for the parish within which it actually fell, and then again for any qualifying orphan parishes nearby.

## 3. Investigating alternative options

Alternative approaches independent of OAPWCs include generating parish estimates from the Census microdata or from postcode-level estimates. A further option which is being actively pursued is to base parish estimates on a grid. Exact estimate statistics for the Census residents within each grid cell would be applied to the grid cell centroid (GCC) and the cell statistics aggregated up to parish.

The initial specification was that grid squares should extend not only across Parish boundaries, but that this should be consistent at every level – they should not be cut by any administrative boundaries: entire grid cells should cross LAD and national boundaries and the coastline. But this blanket consistency causes anomalies.

The mathematical simplicity of grid squares is at odds with the reality of contorted real-world perimeters, and at some level, this has to be recognised and addressed. Initial results for Census households applied to the geometric GCC of complete grid squares have revealed over 1.1 million Census residents allocated to the wrong side of unparished area or national boundaries, or in the sea. For any grid cell transected by the parished/unparished boundary, both sides are likely to be populated, but unparished areas more heavily, resulting in a net population gain[5] for parishes. Wythall CP in Bromsgrove (actual population 11,500) gains an additional 6,000 residents from Birmingham (Figure 2). Such cases would be avoided by clipping grid cells to the parished/unparished boundary. Clipping to the coastline would prevent a further loss of 66,000 residents.

Both of the coastline and parished/unparished boundary may be extremely indented. Clipping to either may produce partial grid squares where the area of interest[6] is split into two or more fragments by the perimeter of the grid cell. Each will be processed separately. The values for each polygon centroid are calculated and applied to its centroid.

## 4. Parish estimates from grid cells

Census Output Areas were constructed on the basis of upper and lower population thresholds, Their OAPWCs are the foundation of the best-fit method for higher geographies, which failed to produce publishable data for 1,140 parishes and 120,000 residents. Even though the mesh of GCCs is denser than the OAPWC set, 27 small parishes still contained no geometric GCC and 53 no population-weighted GCC[7].

If the same statistical disclosure thresholds were applied to parishes as were used for Output Areas[8], among those parishes that did contain a GCC, 1,099 come out below the threshold for households and

[5] 363,000 in parishes and 720,000 in unparished areas, so a net imbalance of 357,000
[6] E.g. the onshore/parished portion as opposed to the offshore/unparished portion.
[7] Compared with 1,140 parishes wiithout an OAPWC.
[8] A minimum of 100 usual residents or 40 households.

1,161 that for usual residents. For population-weighted GCCs, the figures are actually a little worse[9]: their poorer performance is due to their more closely reflecting the real settlement pattern.

**Table 1** Misplaced data with complete grid cells and geometric centroids

**Data not recorded in correct place if no account taken of any boundaries**

| | | Populated gridcells affected | | Parishes affected | | Usual residents affected |
|---|---|---|---|---|---|---|
| **English Parish data lost to Scotland** | | 14 | - | - | | 319 |
| **EW complete gridcells - data lost to the sea** | | 375 | - | - | | |
| As singleparts (e.g. Either side of Menai Strait) | | | | | | |
| In Wales | 123 | | 76 | | 8,918 | |
| in England | 314 | | 190 | | 45,974 | |
| | | 437 | | 266 | | 54,892 |
| **English Parish data lost to Welsh Communities** | | - | | - | | 3,875 |
| **Welsh Community data gained by English Parishes** | | - | | - | | 10,063 |
| **English Parish data lost to unparished Areas** | | 2,147 | | 1,066 | | 363,173 |
| **Unparished Area data gained by English Parishes** | | 1,814 | | - | | 720,299 |
| **Gross impact** | | 4,787 | | | | **1,152,621** |
| **Net effect on English parishes** | | | | | | **408,969** |

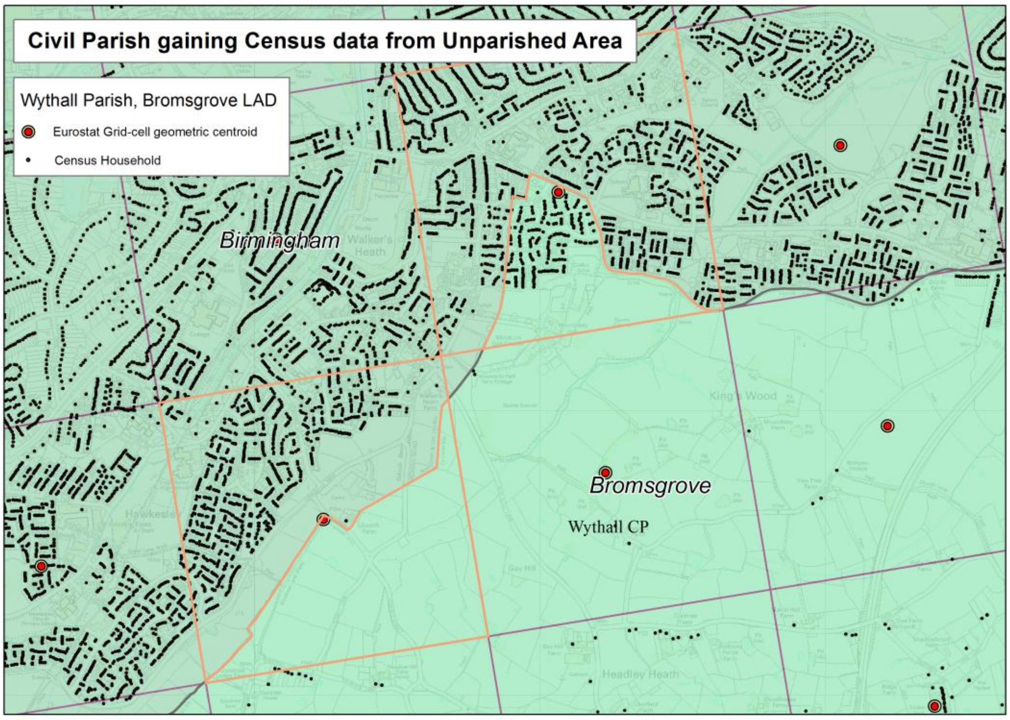


**Figure 1** Wythall CP, Bromsgrove

9 1,118 parishes contain fewer than 40 households and 1,255 contain fewer than 100 usual residents.

Assuming that similar statistical disclosure thresholds are retained for the 2021 Census as for 2011, these results suggest that Census data may still not be independently publishable for some sparsely-populated parishes, whatever method is used to generate estimates. For such parishes, some form of amalgamation will be required, and our objectives include the identification of a methodology that minimises this group.

A simple aggregation of a fictitious dataset to grid cells split by a border (Figure 2) produces values of 62 and 34 when aggregated simply to grid cell centroids.

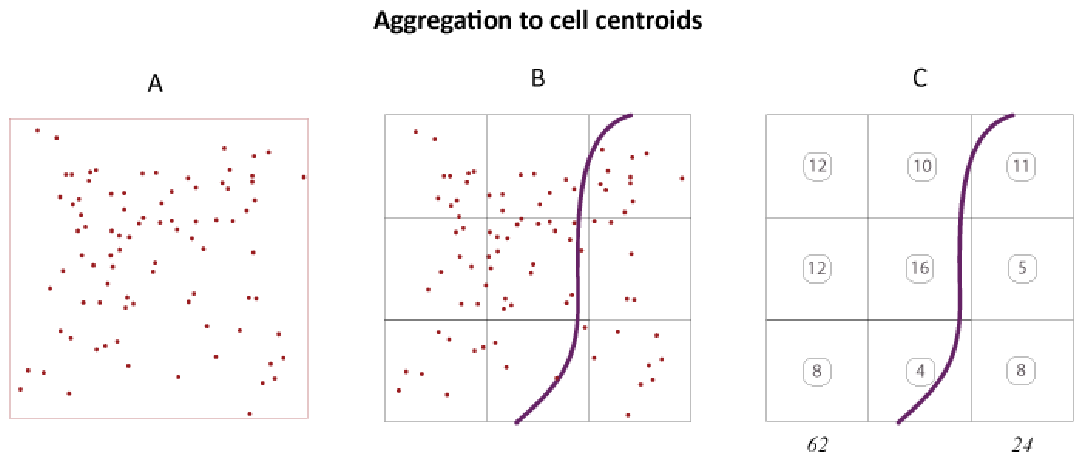


**Figure 2**: Simple aggregation to gridcell centroids

Taking the total value for each grid square and apportioning a quarter to each corner and then summing the combined values for each corner (Figure 3) produces 75% more centroids and produces a simplified ‘four-corner’ smoothing outcome.

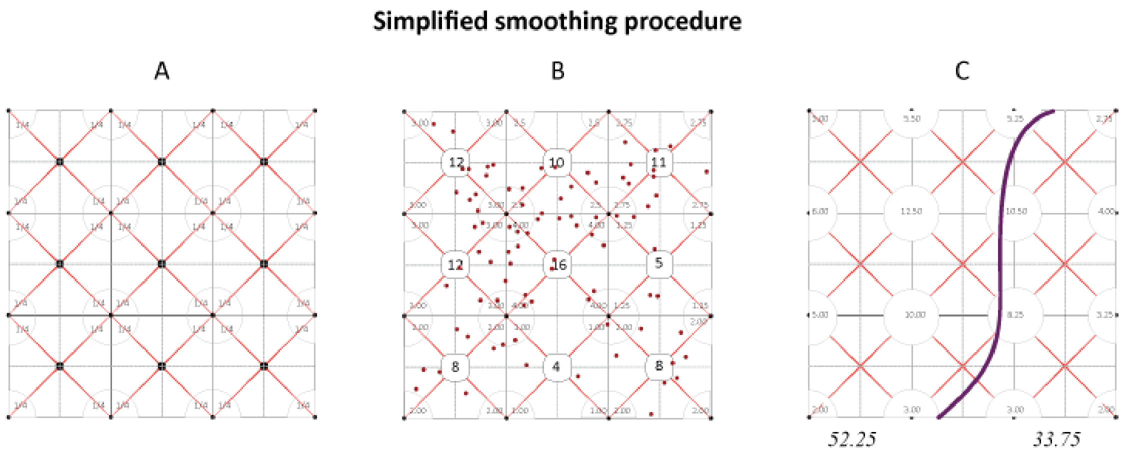


**Figure 3:** Simplified smoothing method

Alternatively, allocating each individual data point (household) to the nearest of five cardinal points in the cell – the centroid and the four corners (Figure 4) – effectively creates a new finer grid at a 45 degree angle to the first. It also produces a smoothed outcome, but one that is closer to the original.

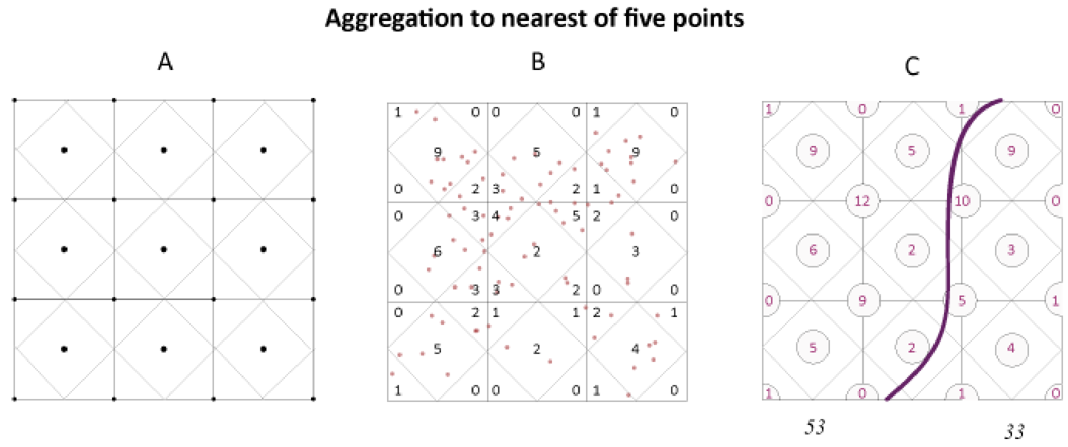


**Figure 4**: Five point method

Application of either will boost the centroid dataset and reduce the number of parishes without data. Both methodologies rely upon quadrilateral grid cells. Grid cells clipped to the coast or unparished areas may be made irregular, but this can be solved by generating a Minimum Bounding Rectangle (MBR) to produce four (or five) points for even the most contorted clipped polygon.

## Summary

ONS is examining the relative merits of three methods of aggregating household level Census data to grid cells and up to civil parishes in England. Each will be assessed against geometric or population-weighted centroids and against complete or cropped[10] grid polygons. Questions of accuracy, simplicity and cost, as well as the power to produce parish estimates and statistics on ‘misplaced’ data will inform the decision on which methodology to adopt for the 2021 Census.

Word count from ‘Introduction’ to here = 1,495.


## Acknowledgements

Andy Tait, Nick O’Rourke, Ian Coady and Dr Amy Fowler.


## Biography

Bruce Mitchell works in the Research Team of ONS’ Geography Branch. He has an MSc in Geographical Information Science from Birkbeck College, University of London. His research interests extend beyond GISc to geography, history, languages and forestry.

[10] To any or all of (in decreasing order of importance) unparished areas, the coast, the Welsh border and the Scottish border.

## References

Cockings, S., Harfoot, A., Martin, D. and Hornby, D. (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 43, (10), 2399-2418

Martin, D. (2002) *Geography for the 2001 Census in England and Wales* Population Trends 108, 7-15

ONS (2010) *GSS Geography Policy for National Statistics*, July 2010.

ONS (2014*):* Ian Coady, Samantha Cockings, David Martin, Andrew Harfoot, Bruce Mitchell *Publishing statistical grids in the United Kingdom*. Publication pending.

ONS (2015) *GSS Geography Policy for National Statistics,* forthcoming.

Open Geospatial Laboratory Southampton (2015) *Opengrid: an open gridded population dataset for England and Wales. Product Manual.* Unpublished report. Open Geospatial Laboratory, University of Southampton.