

# How serious is the Modifiable Areal Unit Problem for analysis of English census data?

Robin Flowerdew

*ESRC Centre for Population Change, School of Geography and Geosciences  
University of St Andrews*

---

## Abstract

Population data are often collected or presented for geographical areas which may have little or no connection to the processes generating the data. Such areal units are termed 'modifiable'. However analysis undertaken on such data is not independent of how these areal units are configured. Indeed, Openshaw (1984) and others have shown that the results of statistical analysis may differ wildly according to the scale and pattern of the areal units used. This phenomenon is called the modifiable areal unit problem (MAUP). It is clear that the MAUP exists, but far from clear about how often it occurs, how often it affects the conclusions from empirical data analysis, and in what contexts it makes most (or least) difference. British census data are well suited for investigating these issues, being available for different geographies which neatly nest within each other, and for a range of different variables of interest to central and local government and to many academic disciplines. This article is concerned with bivariate correlations (using Pearson's  $r$ ) between pairs of variables. The aim is to see if any variables seem particularly liable to display MAUP effects, and if so, why. The conclusion is that MAUP in many cases makes little or no difference to the results, but there are some variable pairs where the effect is substantial.

## Introduction

It is very common in human geography, demography and other social sciences to use census data to examine social trends, to test social theory and to evaluate policy recommendations. Following the recent taking of the 2011 Census in England (and the rest of the United Kingdom), there is likely to be a great deal of interest in research using the census. It may be appropriate at this time to raise issues in the interpretation of census data.

Census data are collected for individual households but are usually released only in aggregate, to avoid disclosure of personal information. Aggregation is often done on the basis of geographical location, and in many countries data are made available at two or more spatial scales. Thus in England, 2011 Census data are available for small areas called Output Areas (OAs) and larger units like wards, districts and Government Office Regions (now just referred to as 'regions'). The smaller units are nested within the larger ones, so that a district is made up of a set of wards, each ward is made up of a set of OAs and households can be uniquely assigned to a specific OA, ward and district. The same set of data can therefore be analysed at different scales, making it more likely that a meaningful comparison can be made between census and other variables.

The problem is that the results of statistical analysis are not independent of the scale at which analysis is done. For example the correlation between variable X and variable Y may be different, in some cases very different, at the OA scale, the ward scale and the district scale. This is an example of the 'scale problem' (Openshaw and Taylor 1979), one of the two forms of the modifiable areal unit problem (MAUP). The other form of the MAUP is that the results of statistical analysis are dependent on the configuration of the zonal systems used, even if scale is the same. This is termed the 'aggregation problem' by Openshaw and Taylor, although here the term 'zonation problem' is preferred on the grounds that the concept of aggregation applies to both problems.

Some studies indicate that correlations over virtually the whole range from -1 to +1 can be reached by judicious placement of zone boundaries (Openshaw 1984; Openshaw and Rao 1995). However, these correlations seem to have been achieved at the cost of highly convoluted (and hence empirically implausible) boundaries; see for example Openshaw and Rao's Figure 5 (2005). Are these dramatic results merely a matter of generating millions of correlations and choosing the extremes for publication?

The MAUP was first identified by Gehlke and Biehl (1934) but, despite the seeming importance of the problem, it is unusual to see published papers exploring the topic. Important exceptions include empirical studies such as Putman and Chung (1989) and Fotheringham and Wong (1991) and theoretical papers like Holt, Steel and Tranmer (1996) and Green and Flowerdew (1996). However, Marble (2000, p 33) has challenged advocates of the importance of the MAUP to find examples of studies where neglect of the MAUP has resulted in important and expensive errors. So there is strong evidence that the problem exists but not that it is always very important. This article is an attempt to evaluate how often MAUP effects occur and how strong those effects are, by taking a set of variables from the 2011 Census of England and examining their correlations with one another. This analysis is carried out at three scales - Output Areas, wards and districts - and the results compared. Out of a large number of pairs of variables, it will be possible to count how many exhibit the MAUP scale effect and how strongly they do so. It would be useful if it proves possible

to reach some conclusions about the extent to which the size of the effect can be predicted. Without such guidance, researchers may be tempted to go ahead with their analysis while ignoring the issue completely, or, perhaps just as bad, to use it as an excuse for avoiding statistical analysis completely.

Holt et al. (1996) undertook a related exercise using 1986 census data for Adelaide in South Australia, using aggregations to the 1,584 Census Divisions (CDs) and 97 Postcode areas (PCs), in addition to a 1 per cent public use sample at the individual level. They examined 82 socio-economic variables and hence 3,321 correlations, comparing values between the individual data and the two geographical scales. They found that CD-level correlations tended to have the same sign but to be much higher than at the individual level. Absolute differences varied from 0 to 0.85 but the median difference was only 0.16.

In the same paper, Holt et al. (1996) also developed measures intended to show the effect of the MAUP on each variable, based on the variances of the variables at different scales. The aggregation effect is defined as the ratio of the higher-level variance, weighted by population at risk, to the lower-level variance. Tranmer and Steel (2001, p. 116) described it as a 'simple way of summarising the scale effect for each variable separately'. Higher values of the aggregation effect for particular variables are associated with a greater likelihood of large MAUP effects. A related statistic is the intra-area correlation (IAC) (see Tranmer and Steel 2001; Manley 2005). It is a direct measure of within-area homogeneity, the correlation for a given variable between different people living in the same areal unit. Unlike the aggregation effects, the IAC measures are adjusted for population size, so they can be used to compare different scales.

## The data set

Eighteen variables were selected for investigation (**Table 1**), all of which are often used in economic, social and demographic research. They include data on ethnicity, housing tenure, marital status, car ownership, illness and employment. Because the raw census data are counts, the correlations may be inflated because of a size effect. Accordingly, the raw count data must be converted to proportions or percentages. This means that the population at risk must be downloaded and the number of people in any category must be divided by the population at risk to get proportions (then multiplied by 100 for percentages). Thus population and ethnicity counts must be divided by total population, employment counts by population aged 16 to 74, car ownership by the number of households, and so on.

For our purposes the study area can be taken as the whole of England. Data for all the lowest-level units in the study area can be downloaded using the CASWEB system at the University of Manchester. These are Output Areas (OAs), of which there are 165,665 with an average population of 297 - they can then be aggregated up in the statistical package SPSS to wards (7,969), with an average population of 8,168, to districts (354, including 46 unitary authorities), with an average population of 138,810, and eventually to regional and national units, not used in this project. At each scale, there is major variation in zone population. Districts range from Isles of Scilly with 2,153 people and the City of London with 7,185 to Birmingham with 977,087. Almost all districts contain hundreds of Output Areas; even Isles of Scilly and City of London having 9 and 38 OAs respectively, compared to 3,127 in Birmingham and 2,439 in Leeds.

Table 1 **Variables used in the analysis**

		Output Area		Ward		District		
		Mean	s.d.*	Mean	s.d.*	Mean	s.d.*	
1	pcmale	Percentage of residents who are male	48.6	3.5	48.8	1.7	48.8	0.7
2	pcfemal	Percentage of residents who are female	51.5	3.5	51.2	1.7	51.2	0.7
3	pcmarr	Percentage of residents over 16 who are married	42.9	12.9	45.6	8.2	44.8	5.3
4	pcdiv	Percentage of residents over 16 who are divorced	8.5	3.8	8.1	1.9	8.2	1.0
5	pcwhite	Percentage of residents of white British ethnicity	87.6	17.3	91.1	13.0	90.3	12.1
6	pcIndian	Percentage of residents of Indian ethnicity	1.9	5.6	1.3	4.0	1.4	3.1
7	pcPak	Percentage of residents of Pakistani ethnicity	1.2	5.2	0.8	3.3	0.8	2.0
8	pcCarib	Percentage of residents of Caribbean ethnicity	1.1	3.0	0.6	1.9	0.7	1.7
9	pcpt	Percentage of residents aged 16-74 in part-time employment	11.8	3.6	12.2	2.1	12.2	1.6
10	pcft	Percentage of residents aged 16-74 in full-time employment	40.6	10.2	40.3	7.1	40.9	4.6
11	pcunemp	Percentage of residents aged 16-74 who are unemployed	3.4	2.6	2.9	1.5	3.0	1.1
12	pc0cars	Percentage of households with no access to a car	26.6	18.5	22.0	13.2	23.8	9.8
13	pcowner	Percentage of households who are owner-occupiers	29.3	15.8	31.9	10.3	30.8	6.4
14	pcmort	Percentage of households who own with a mortgage	38.8	16.6	39.4	9.8	39.5	6.2
15	pcLA	Percentage of households renting from a local authority	13.1	19.9	10.6	11.1	11.4	7.8
16	pcHA	Percentage of households renting from a Housing Association	6.0	12.2	5.7	6.0	5.8	3.9
17	pcpriv	Percentage of households renting from a private landlord	8.8	10.2	8.3	6.0	8.3	3.9
18	pcill	Percentage of residents with limiting long-term illness	33.5	10.7	32.8	6.7	33.0	5.2
Number of cases			165,665		7,969		354	

\*Note: s.d. = standard deviation

The variables are defined in Table 1, which also gives the means and standard deviations at each of the three scales. Some of them represent conditions that are fairly common in the data set; at the OA level, the percentage of people of white British ethnicity is 87.6 per cent; other variables with high values include percentage of females (51.5 per cent), percentage of males (48.5 per cent), percentage of people over 16 who were married (42.9 per cent) and percentage of people aged 16 to 74 in full-time employment (40.6 per cent). Others represent more unusual conditions, such as percentage of people of Caribbean (1.1 per cent), Pakistani (1.2 per cent) and Indian (1.9 per cent) ethnicity, and percentage of people aged 16 to 74 who were unemployed (3.4 per cent).

A distinction can also be drawn (**Table 2**) between variables with a high mean and low standard deviation like percentage of males and percentage of females, indicating relatively even distributions, and those with low means and high standard deviations, like percentage of households renting from the local authority and percentages of Indian and Pakistani ethnicity, whose distributions are concentrated in certain parts of the country. There are of course other variables with high means and high standard deviations and others with low means and low standard deviations. For our purposes, we divide our data set variables into categories based on

the ratio of mean to standard deviation. These categories might be expected to show some relationship to the modifiable areal unit problem.

Descriptive statistics for the variables in the data set can be computed at all three scales. The mean might be expected to be approximately the same at each scale. That they are not the same is because the smaller units are not all exactly the same size. If the value at one scale is the weighted mean of the values at the lower scale, the means would be the same. The fact that it is the unweighted mean suggests that units with lower population size have slightly more influence.

Indeed, comparison of variable means shows that the three scales differ slightly. The percentage male for example ranges from 48.5 at the OA scale to 48.8 at the ward scale and 48.8 at the district scale. Larger differences occur for some variables, such as white British ethnicity (87.6 for OAs, 91.1 for wards, and 90.3 for districts), Caribbean ethnicity (1.1 for OAs, 0.6 for wards, and 0.7 for districts) and local authority renting (13.1 for OAs, 10.6 for wards, and 11.4 for districts). High values for a particular scale should occur where the variable tends to be more concentrated in smaller units.

**Table 2 Classification of census variables by mean and standard deviation**

<b>A High mean, low standard deviation</b>	<b>Mean Scale Difference</b>
pcmale	.117
pcfemal	.117
pcwhite	.181
<b>B Moderate mean and standard deviation</b>	
pcfull	.193
pcmarr	.187
pcpt	.220
pcill	.149
<b>C Moderate mean and high standard deviation</b>	
pcmort	.142
pcdiv	.219
pcowner	.198
pcunemp	.183
pc0cars	.150
<b>D Low mean, high standard deviation</b>	
pcLA	.184
pcHA	.167
pcpriv	.195
pcCarib	.180
pcIndian	.142
pcPak	.145

Definitions:

- A Mean at least 5 times higher than standard deviation
- B Mean between 3 and 5 times higher than standard deviation
- C Mean between 1 and 3 times higher than standard deviation
- D Mean lower than standard deviation

Source: Statistics derived from 2001 Census at the Output Area scale

Differences between scales in standard deviations are much larger, with the values for districts being the lowest and the values for the OAs being highest. Examples include owner-occupiers with a mortgage (6.2 for districts, 9.8 for wards, and 16.6 for OAs), Indian ethnicity (3.1 for districts, 4.0 for wards, 5.6 for OAs) and no-car households (9.8 for districts, 13.2 for wards, 18.5 for OAs).

## Correlations

The eighteen variables in the data set, taken pairwise, provide 153 correlations at each of the three scales. At the OA level, these range from -1.000 to +0.670 with a mean of -0.033 and a standard deviation of 0.332. The quartiles are -0.234, -0.003 and +0.199. At the ward scale, the range is from -1.000 to +0.841, with a mean of -0.022 and a standard deviation of 0.364; the quartiles are -0.244, -0.073 and +0.262. At the district scale, the range is from -1.000 to +0.856, with a mean of -0.034 and a standard deviation of 0.414; the quartiles are -0.321, -0.036 and +0.252. At all three scales, the strongest correlations involved the same two variables, pcmale and pcfemale having the strongest negative correlation and pcunemp and pc0cars the strongest positive.

Out of the 153 pairs of variables, 82 had negative correlations at the OA level compared to 79 at the ward level and 81 at the district level. 68 of these were negative at all three scales, while 60 were positive at all three. Of the remainder, there was no clear trend: 12 were negative at the OA level but positive at the district level, while 12 changed in the opposite direction and 10 changed in one direction and then back again. See **Table 3** for more details.

**Table 3 Positive and negative correlations at different scales**

Count				Ward scale		Total
District scale				Negative	Positive	
Negative	Output Area scale	Negative		68	2	70
	Output Area scale	Positive		6	5	11
	Total				74	7
Positive	Output Area scale	Negative		5	7	12
	Output Area scale	Positive		0	60	60
	Total				5	67

The correlations at the three scales can themselves be correlated. The results are strongly positive, but not as overwhelming as might be expected. The district-level correlations with the equivalent ward-level correlations have a correlation of 0.934. The correlation between OA-level and ward-level correlations is a little higher at 0.960, and between OA-level and district-level correlations it is lower at 0.881. The degree to which these correlations differ from 1.000 is an indicator of the extent to which analysis is complicated by the scale issue.

After calculating correlations for the same data obtained for the Output Area scale (and aggregated first to the ward scale and then the district scale), it is possible to compare the coefficients at each scale. A simple and intuitively appealing statistic can be obtained by observing the difference

between any two of the three correlations. We can then investigate if some of the variables seem to show a big difference between scales of analysis more often than others. It is also possible to see whether particular characteristics of variables are related to larger or smaller differences.

Another way of assessing the magnitude of the MAUP effect for this data set is to compare the value of the correlation coefficient at one scale with its values at the other scales. This can be most easily calculated by subtraction of one from the other:

$$\text{Diff}_{ij}(\text{OA, ward}) = r_{ij}(\text{OA}) - r_{ij}(\text{ward})$$

$$\text{Diff}_{ij}(\text{OA, district}) = r_{ij}(\text{OA}) - r_{ij}(\text{district})$$

$$\text{Diff}_{ij}(\text{ward, district}) = r_{ij}(\text{ward}) - r_{ij}(\text{district})$$

The  $\text{Diff}_{ij}$  values are direct measures of the MAUP effect. Note that they can vary between -2 and +2, although they are unlikely to be outside the -1 to +1 range. Values around 0 would be expected if there was no important MAUP effect. The three different values can also be compared to see if results at one scale are similar or different from those at other scales.

**Table 4 Pairs of variables showing the greatest scale effects**

	Variable 1	Variable 2	r(OA)	r(ward)	r(district)	Greatest Difference
1	pcmarr	pcunemp	-.524	.037	-.635	.598
2	pcmort	pcpriv	-.229	-.451	-.688	.459
3	pcdiv	pcowner	-.489	-.474	-.051	.438
4	pcpt	pcmort	.218	.286	.606	.388
5	pcwhite	pcpt	.451	.626	.825	.374
6	pcCarib	pcpt	-.279	-.431	-.644	.365
7	pcHA	pcill	.208	.106	-.137	.345
8	pcmarr	pcwhite	.247	.400	.586	.339
9	pcwhite	pcowner	.253	.376	.573	.320
10	pcdiv	pcHA	.276	.273	-.036	.312
11	pcdiv	pcft	-.061	.017	-.294	.311
12	pcdiv	pcpt	-.075	.094	.231	.306
13	pcCarib	pcpriv	.131	.231	.434	.303
14	pcft	pcpriv	.097	-.111	-.203	.300
15	pcft	pcLA	-.362	-.142	-.063	.299
16	pcmale	pcft	.216	.223	.496	.280
17	pcfemal	pcft	-.216	-.223	-.496	.280
18	pcmort	pcill	-.538	-.436	-.261	.277
19	pc0cars	pcpriv	.240	.343	.515	.275
20	pcCarib	pcowner	-.299	-.361	.529	.267

**Table 4** lists the twenty pairs of variables where the MAUP effect (as measured by the absolute differences in correlation coefficients between scales) is strongest. Fifteen of the eighteen variables are involved, including demographic, ethnic, employment, tenure and health variables; the exceptions being Indian ethnicity, unemployment and lack of a car. It is possible to work out the combined effect of the MAUP on the 17 relationships which each variable has with the others, by

summing the largest difference between scales and dividing by 17. This quantity is referred to as the mean scale difference (MSD), and is shown in **Table 5**.

It can be seen that there is considerable variation in mean scale difference. Certain variables seem to have larger MSD values than others, with *pcdiv* (divorce) and *pcpt* (part-time employment) having the highest values followed by *pcowner* (outright home ownership) and *pcpriv* (private renting). The lowest effects are found for *pcmale* and *pcfemal*, followed by *pcIndian* and *pcmort* (mortgaged). There appears to be little relationship between MSD and the classification in Table 2. It is tempting to regard MSD as a measure of the liability of each variable to the MAUP, but the measure is dependent on which other variables are involved. These results are of course specific to the particular data set used, and might look very different if the other variables being studied were different.

The most startling relationship between two variables is that of home ownership with a mortgage and renting from a private landlord, where the correlation is weakly negative (-.229) at the OA level and increasingly strong at the ward (-.451) and district (-.688) levels. This perhaps suggests that these two forms of tenure have complementary distributions at the local scale as well as at the city and metropolitan scales. The same process may be working at all three scales.

Table 5 **Mean scale difference between variables**

Variable	Mean Scale Difference
<i>pcmale</i>	.117
<i>pcfemal</i>	.117
<i>pcmarr</i>	.187
<i>pcdiv</i>	.219
<i>pcwhite</i>	.181
<i>pcIndian</i>	.142
<i>pcPak</i>	.145
<i>pcCarib</i>	.180
<i>pcpt</i>	.220
<i>pcft</i>	.193
<i>pcunemp</i>	.183
<i>pc0cars</i>	.150
<i>pcowner</i>	.198
<i>pcmort</i>	.142
<i>pcLA</i>	.184
<i>pcHA</i>	.167
<i>pcpriv</i>	.195
<i>pcill</i>	.149

## Significance testing using Fisher's Z

The results so far have shown that correlations at different scales do differ, but in many cases the differences are fairly small. This raises the question of whether these differences mean anything in the interpretation of results. It is possible to test the results for statistical significance using Fisher's Z. This can be done in two stages. The first stage is to use Fisher's Z to standardise the correlation

values at the scales that are being compared and the second is to test for the significance of the difference.

Note:

$$Z_{ijk} = 0.5 + \ln \left( \frac{1 - r_{ijk}}{1 + r_{ijk}} \right)$$

where  $\ln(x)$  is the natural logarithm of  $x$  and  $r_{ijk}$  is the correlation of variables  $i$  and  $j$  at scale  $k$ .  $Z$  scores can be calculated for each pair of variables at each scale.

The second stage is to compare the  $Z$  scores for two scales. Note also that the correlation of two variables at scale  $k$ ,  $r_{ijk}$ , is significantly different at the 0.05 level from the correlation of the same two variables at scale  $h$  if

$$(Z_{ijk} - Z_{ijh}) \geq 1.96 * (\text{SQRT} \left( \left( \frac{1}{m_k - 3} \right) + \left( \frac{1}{m_h - 3} \right) \right))$$

where there are  $m_k$  zones at scale  $k$  and  $m_h$  zones at scale  $h$ . If the left side of this inequality is smaller than the right side, the correlations are not sufficiently different to be statistically significant. The latter result, in other words, would suggest that the MAUP is not strong enough (for these two scales anyway) to affect our interpretation of the results.

Out of the 153 pairs of variables in this study, there is no significant difference between the OA and ward scales in 60 cases, between the OA and district scales in 68 cases, and between the ward and district scales in 61 cases. In other words, the MAUP does make a difference in about 60 per cent of cases. The specific pairs picked out in the  $Z$  score computation are very similar to those mentioned in Table 4 which was based on a rather crude arithmetic difference between correlation coefficients. The overall conclusion from Fisher's  $Z$  would seem to be a confirmation of the importance of MAUP; on the other hand, it is well known that very weak results may be statistically significant if the sample size is big enough (as it certainly is here).

## Factors affecting the size of the MAUP effect

It could be argued that the variables with relatively high values might be more likely to be stable in comparison to those representing rare occurrences. It might also be expected that variables with a high standard deviation might be prone to the MAUP, because it could be that the pattern picked out at one scale would be lost through aggregation. Perhaps the relationship between mean and standard deviation is what is important rather than either measure by itself. A further statistic which might be relevant is skewness - if a variable is highly skewed, those cases with the most unusual values might have different effects according to how they are affected by the aggregation procedure. There is also an issue concerning whether these effects should refer to one of the two variables concerned or to both in combination.

A further statistical relationship which might be relevant to MAUP is the relationship of correlation to population size. A useful statistic is the contribution of each areal unit to the overall covariation, referred to here as the cross-product. It is calculated as

$$C_k = (X_k - \bar{X}) * (Y_k - \bar{Y})$$

for areal unit  $k$ , with  $\bar{X}$  and  $\bar{Y}$  the means of variables  $X$  and  $Y$ . Again this is illustrated from the case of  $pcmort$  and  $pcpriv$ . At the OA scale, where the correlation was particularly strong, it is

possible to judge the contributions of each OA to the overall statistic. The top ten OAs by cross-product were in the wards listed in **Table 6**.

These are all places that combine small proportions of households buying with a mortgage and large proportions renting from private landlords. Note that several of the top OAs on this measure are in the same wards. This suggests that the aggregations from OA to ward should not have too dampening an effect on the correlation.

It is also possible to investigate whether size of areal unit has any impact on the MAUP effect. If cross-products are taken as local measures of contribution to the correlation, these measures can be correlated with population. For the ward scale, this results in an  $r$  value of  $-.023$ , suggesting very little connection (though it is still statistically significant). The  $r$  values at the OA and district scales were  $-.052$  and  $-.016$ , again indicating very little relationship between cross-product and population.

**Table 6**      **Wards containing those OAs making the biggest contribution to the cross-product of pcmort and pcpriv**

1	Grange, Newcastle upon Tyne
2	Goldsmid, Brighton
3	Headingley, Leeds
4	Headingley, Leeds
5	Grange, Newcastle upon Tyne
6	Rendlesham, Suffolk Coastal
7	Middleton-in-Teesdale, Teesdale
8	Headingley, Leeds
9	Rendlesham, Suffolk Coastal
10	Central, Manchester

### **Sensitivity of results to data modification**

Normally we might expect MAUP to be strongest if areal units showing a relationship between two variables are aggregated together with others not showing that relationship, or even showing the opposite relationship. If data are altered by modifying the values in a few places or by shifting areal unit boundaries, it may be possible to assess the robustness of statistical analysis to quirks of the data and the geography for which it is presented.

For example, our analysis picked out the variables pcmort and pcpriv as having a particularly strong MAUP effect, where a mild negative correlation at the OA level became stronger with data aggregation. Observations can be picked out from the data whose cross-product makes a particularly large contribution to the correlation and the relevant figures set to the variable means. In the case of the district level data, the biggest contributor (of the 354 districts) is Westminster, whose (low) proportion of households with mortgages and (high) proportion of private renters are

set equal to the national means. This single modification reduces the correlation of these two variables from  $-.688$  to  $-.676$ ; if the same treatment is extended to the places ranked from 2 to 5 on the cross-product criterion (Kensington and Chelsea, Isles of Scilly, Camden and City of London), the correlation goes down to  $-.617$ . If it is further extended to the top ten, the correlation is reduced to  $-.572$ . These reductions do not explain the MAUP effect, but they do show that individual observations can have fairly substantial impact on overall relationships.

With the larger number of cases at the ward level, this procedure is likely to show weaker overall impact. It is St Agnes in the Isles of Scilly that has the greatest cross-product for these two variables, emphasising the frequently observed tendency for very small zones to show extreme and sometimes unreliable behaviour in statistical analysis. It has a cross-product of  $-1529.49$ , followed by Headingley (Leeds,  $-1002.70$ ), Tower (London,  $-978.11$ ), Ford (Berwick-upon-Tweed,  $-974.88$ ) and Bishopsgate (London,  $-921.55$ ). Note that Tower and Bishopsgate are also very small in population. Changing the values to the means for the variables *pcmort* and *pcpriv* for these five wards results in a reduction in the correlation coefficient from  $-.451$  to  $-.447$ .

However, we can inspect the results of assigning wards with high cross-products to districts other than the one to which they are allocated - a direct approach to the MAUP zonation problem. The impact on the correlation at the aggregated scale will depend on whether the relationship in the zone to be allocated is in the same direction as that of the zone which is to be joined. In practice, this effect could be visualised as the effect of wards being swapped between districts. If this swap is geographically sensible (in terms of contiguity and social similarity), it seems unlikely that it would have much effect on MAUP results. Using Lancaster district as an example, it was found that wards eligible for swapping with neighbouring districts tended to be reasonably similar, and hence to make little difference to the correlation. The most promising swap on the basis of cross-products involved Lower Lune Valley ward in Lancaster and Ingleton and Clapham ward in Craven (North Yorkshire), but even this swap failed to change the overall coefficient at the third significant figure.

## Definition of universe of study (regional models)

In any analysis of geographical data, the definition of the study area may have some bearing on the results. If the study area is taken as a random sample of a larger population, one might expect the results of a correlation to be reasonably similar. However, if the study area is likely to contain a higher proportion of subareas which are major residuals, positive or negative, the correlation may be strengthened. Likewise if those residuals are excluded it may be weakened.

The main analysis discussed in this article has covered the whole of England, but many studies might choose to define the study area to be smaller, for example just one of the former Government Office Regions. How much difference would this make to the importance of the MAUP?

## Conclusions

The use of a large data set from the English 2001 Census enabled a number of aspects of the MAUP to be investigated. The main finding concerns the frequency of occurrence of the MAUP when applied to realistic data presented for sensible geographies. First, the majority of cases do

not show very big differences between correlations at different scales. There are no cases of variable pairs with large positive correlations at one scale and large negative ones at another, although there are several cases where the correlations are very different, although the sign is the same. These cases require recognition of the importance of the MAUP. It has also been shown by the significance testing performed using Fisher's Z that a change of scale in many cases may be statistically significant even though the estimated correlation coefficients are not very different.

It has been possible to pick out variables which seem to be more liable to suffer from the MAUP, but little progress has been made in determining why, or in being able to predict when it will be encountered. The relative sizes of means and standard deviations do not seem to be related, and nor does population. Excluding or editing the data to remove places which have a disproportionate influence on the overall result made marginal contributions to the understanding of the problem.

One type of information which may have a major impact on these scale differences is the extent and importance of spatial autocorrelation (Manley et al. 2006). A high degree of spatial patterning at the local level is more likely to be maintained at a higher level if the data are spatially autocorrelated, though zonation as well as scale effects are important.

An additional problem is the need to reconcile the spatial configuration of data collection and presentation units with the scale and shape of the processes we are interested in. This means that, even if the scale at which we want to study can be approximated by Output Areas, wards and districts, a great deal of flexibility is needed to ensure that data can be obtained in the way most useful for research purposes.

The intention of this article is to provide practical advice to people encountering the modifiable areal unit problem. It concludes therefore with a set of recommendations of points to consider:

1. If data are available at two or more scales, check whether results are compatible between two scales.
2. Could an apparent relationship between two variables actually be an artefact caused by a third variable controlling the first two?
3. Can you derive a different interpretation of your data by including or omitting particular cases?
4. Are your results affected by study area delimitation?
5. What data scale is most coincident with the geographical processes under study?

How serious is the MAUP? Most of the time, not very – but there are occasions when it can make a great deal of difference. Unfortunately it remains difficult to predict these occasions.

## Acknowledgements

Census data, which are Crown Copyright, are available for British censuses from 1971 to 2001. Access is available using the CASWEB system via the Census Dissemination Unit at the University of Manchester. Most of the computation was done by Samantha Barnwell, supported by a summer internship from the School of Geography and Geosciences at the University of St Andrews. This article has benefited from discussion with David Steel and David Manley, but they are not responsible for its weaknesses.

## References

- Fotheringham A S and Wong D W S (1991) The modifiable areal unit problem in multivariate statistical analysis *Environment and Planning A* 23, 1025-1044
- Gehlke C E and Biehl K (1934) Certain effects of grouping upon the size of the correlation coefficient in census tract material *Journal of the American Statistical Association, Supplement* 29, 169-170
- Green M and Flowerdew R (1996) New evidence on the modifiable areal unit problem, in Longley P and Batty M (eds) *Spatial analysis: modelling in a GIS environment*. Cambridge: Geoinformation International 41-54
- Holt D, Steel D G and Tranmer M (1996) Area homogeneity and the modifiable areal unit problem *Geographical Systems* 3, 181-200
- Manley D (2005) The Modifiable Areal Unit Phenomenon: An Investigation into the Scale Effect Using UK Census Data, unpublished Ph D thesis, School of Geography and Geosciences, University of St Andrews
- Manley D, Flowerdew R and Steel D (2006) Scales, levels and processes: Studying spatial patterns of British census variables *Computers, Environment and Urban Systems* 30, 143-160
- Marble D F (2000) Some thoughts on the integration of spatial analysis and Geographic Information Systems *Journal of Geographical Systems* 2, 31-35
- Openshaw S (1984). *The Modifiable Areal Unit Problem*. Norwich: GeoBooks.CATMOG 38
- Openshaw S and Rao L (1995) Algorithms for re-engineering 1991 census geography *Environment and Planning A* 27, 425-446
- Openshaw S and Taylor P (1979). A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In Wrigley N (ed.) *Statistical applications in the spatial sciences* (Pion, London) 127-144
- Putman S H and Chung S-H (1989) Effects of spatial systems design on spatial interaction models. I: the spatial definition problem *Environment and Planning A* 21, 27-46
- Tranmer M and Steel D (2001) Using local census data to investigate scale effects. In Tate N J and Atkinson P M (eds) *Modelling scale in geographical information science* (Wiley, Chichester) 105-122