Scale, Aggregation, and the Modifiable Areal Unit Problem

59

David Manley

Contents

Abstract

The modifiable areal unit problem (MAUP) is a serious analytical issue for analysts using spatial data. The MAUP manifests itself through the instability of a wide range of statistical results derived from analysis on spatially organized data. When spatial data are aggregated, the results are conditional on the spatial scale at which they are conducted, and the configuration of the areal units that are employed to represent the data. Such uncertainty means that the results of spatial data where the MAUP has not been considered explicitly should be treated with caution. Although solutions have been proposed, none have been applicable in more than a couple of specific cases. As such, it is likely that the MAUP will never be truly solved. This chapter charts the two related aspects of the MAUP, the scale and zonation effects, and details the role of spatial autocorrelation in understanding the processes in the data that lead to the

D. Manley
School of Geographical Sciences, University of Bristol, Bristol, UK
e-mail: d.manley@bristol.ac.uk

M.M. Fischer, P. Nijkamp (eds.), *Handbook of Regional Science*,
DOI 10.1007/978-3-642-23430-9_69, © Springer-Verlag Berlin Heidelberg 2014

statistical nonstationarity. The role of zone design as a tool to enhance analysis is explored and reference made to analyses that have adopted explicit spatial frameworks.

59.1 Introduction

A serious problem for analysts of spatial data is that while the phenomena they are investigating may be continuous, the data available frequently are not, and the areal units used to present the continuous data are arbitrary compromises designed to suit a wide range of uses rather than spatial equivalents of the day, month, or year. As a consequence, statistical analysis of individual data that has been aggregated into areal units is susceptible to nonstationarity across a wide range of measures. This problem is known as the modifiable areal unit problem (MAUP), and it has vexed users of aggregate data for many decades. Countless investigations have demonstrated that it is unlikely that an analytical solution to the MAUP will be identified, and those solutions that have been proposed frequently suffer from substantial flaws. Indeed, as yet, we have neither a full and detailed understanding of the problem nor the underlying causes. It is unlikely that an analytical solution to the MAUP will ever be realized due to the wide range of possibilities that arise when the partitioning of continuous space is implemented as well as the wide range of analytical tasks that aggregated data are required to perform (for comprehensive overviews of the MAUP, see Openshaw 1984; Wong 2009). Instead, the MAUP needs to be accounted for clearly in the research hypothesis that precedes analysis. In the twenty-first century, spatial data are an increasingly important factor in everyday life. Almost all nations in the developed world collect and publish data using administrative boundary systems – areal units. In the United Kingdom, the decennial population census is published using small, low-level areal units. Small area geographies, for a comprehensive range of area characteristics such as are available for the British Census, are valuable as the hidden aspects of the problem are less likely to occur, other things being equal, at fine levels of granularity than coarse ones. It is also worth noting that the small areal units of the British Census were designed explicitly drawing on the principles of the MAUP, promoting, amongst other things, internal homogeneity across a range of important indicators such as housing tenure. The problem of the MAUP is magnified by the temporary nature of the areal units and the frequent revisions that are made to the coverages to reflect changes in population data.

Despite the prevalence of the MAUP in spatial data, it is an issue that is all too frequently ignored or neglected in geographical analysis. A search in Google Scholar on the term “modifiable areal unit problem” reveals only 4,160 publications, a low number when you consider the number of papers that deal with aggregated data in their analysis (around 400,000). The lack of attention paid to the MAUP has, perhaps, two underlying causes. Firstly, the readily available nature of many areal unit systems means that the majority of research using aggregate data adopts areal boundaries that are generated a priori and an engagement with the creation of areal units is not required. Secondly, the results of many quantitative studies that employ aggregate data of one sort or another rely on the implicit assumption that the MAUP isn't a significant problem in order to present valid results. To acknowledge the MAUP, even informally, would be to question the validity of the analysis conducted and conclusions reached. Openshaw's conclusion from almost three decades ago remains as pertinent today as it was when he wrote it: "this is hardly a satisfactory basis for the application and further development of spatial analysis techniques in geography" (1984, p. 5).

This chapter explores the problem of the MAUP in the context of spatial data analysis, outlining the two major aspects of the problem, the scale effect and the zonation effect. Definitions are provided for both these aspects, and examples are drawn from the literature to illustrate the problems. Following these two sections, an overview of the evidence relating to the MAUP is provided.

59.2 MAUP Definitions

There are two aspects to the MAUP known as the scale effect and the zonation effect (also called the aggregation effect in some literature (for instance Openshaw 1977), but since the process of aggregation is involved in both scale and zonation decisions, an important distinction is made here, and the term zonation effect employed). This section outlines the two aspects with reference to relevant examples and provides the context for a discussion around empirical results in the following sections.

59.2.1 The Scale Effect

The scale effect arises because of the nested hierarchies within which human society is arranged and is expressed through the task of choosing the most appropriate scale for analysis (Arbia 1989) (Fig. 59.1). It is rarely that clear at which spatial scale an analysis should proceed, and frequently, there are multiple spatial scales at which an analysis could theoretically be conducted. Drawing on the United Kingdom Census as an example, output areas (OAs, typically 140 individuals) form the basic spatial units and can be aggregated into higher-level spatial units, such as wards (usually a couple of 1,000 individuals) and districts (many 100,000s of individuals).

The "classic" example of the scale effect was published by Gehlke and Biehl (1934) and used three different datasets including random coin tosses, census data, and experimental groups of rural counties drawn from the United States (see also Yule and Kendal 1949). They demonstrated that coefficients from correlation analyses between, for instance, census data reporting juvenile delinquency and monthly house rentals tended to increase as the number of areal units representing the data decreased. Table 59.1 reproduces the results of their correlation analysis. While the census data may be susceptible to structures within the data that cannot be observed, which in turn cause the instability of the statistical results, the coin toss data demonstrated that correlation coefficients changed even when the underlying data were generated randomly, and each data unit was independent of all others. From their analysis, Gehlke and Biehl concluded by questioning whether or not

Fig. 59.1 The scale problem: The three different scales could represent (**a**) output areas, (**b**) wards, and (**c**) districts

Table 59.1 Correlation coefficients under aggregation using juvenile delinquency and monthly rentals (from Gehlke and Biehl 1934, p. 169)

Number of areal units	Correlation coefficient (r)
252	−0.502
200	−0.569
175	−0.580
150	−0.606
125	−0.662
100	−0.667
50	−0.685
25	−0.763

"a geographical area is an entity possessing traits, or merely one characteristic of a trait itself" (p. 170). In essence, they urge caution in the treatment of data from areal units and "that variations in the size of the correlation coefficient seemed conditioned on the changes in the size of the unit used" (op.cit.).

Exploring the scale effect, Kirby and Taylor (1976) use data on referendum voting patterns to illustrate the potential pitfalls and identify pockets of the population who vote differently to the overall outcome for an area. The implication of this finding being that if analysis is conducted at difference scales it is possible to produce different area results from a single pattern of individuals voting. Kirby and Taylor also discuss the dilemma of choice of scale: at a scale that is too small, then it is not possible to compare data sources from different (modifiable) unit systems. However, with the scale too large, then much of the more local-level detail within an analysis is lost through the aggregation process. The scale effect has, therefore, a number of different elements, including the enhancing or smoothing of spatial processes, akin to the statistical smoothing of data to remove noise. The nontrivial nature of the scale effect was emphasized by Openshaw (1984), noting that even a relatively small set of zones can produce a sizable range of combinations: for instance, combining 1,000 zones into a new system of just 20 groups produces 10^{1260} unique combinations!

59.2.2 The Zonation Effect

Once the scale of the zonal system has been determined, then we can consider how the space is to be divided up – the zonation effect. The zonation effect occurs where there are

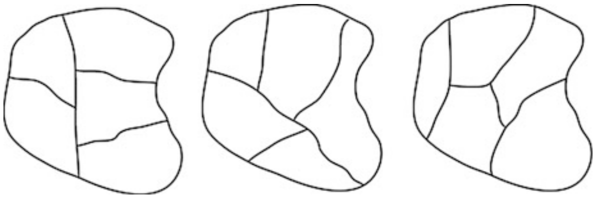

Fig. 59.2 The zonation problem. Each of these diagrams demonstrates a division of a sample space into five distinct areal units, yet each could potentially yield different results

Table 59.2 Correlation coefficients from Openshaw and Taylor (1979, p. 129) showing zonation effect (adapted)

Number of areal units	Correlation coefficient (r)
Six republican-proposed	0.482
Six democratic-proposed	0.627
Six congressional districts	0.265
Six urban/rural regional types	0.862
Six functional regions	0.713

"any variations in results due to alternative units of analysis where . . . the number of units, is constant" (Openshaw and Taylor 1979). There are potentially an infinite number of different ways in which a continuous space can be subdivided into discrete areal units. A diagrammatic interpretation of the zonation problem is presented in Fig. 59.2.

For Openshaw (1984) the zonation effect was by far the greater of the two aspects of the MAUP, as there is considerably more freedom choosing the delineation of boundaries than in choosing the number of zones required. The consequence of this is that "the process of zonation becomes susceptible to the whims of those involved in the overall aggregation process" (Openshaw and Taylor 1981, p. 61). While this position may be extreme, it makes the point that there are serious problems with the arbitrary nature of the many areal units.

Openshaw and Taylor (1979, 1981) conducted one of the largest investigations into the MAUP. Replicating the earlier work of Gehlke and Biehl, they used correlation analysis to assess the instability of statistical analysis as a consequence of the MAUP. In the first instance, they correlated the proportion of republican voters against the percentage of the population above 65, using the 1970 US Census. To assess the impact of the zonation effect, Openshaw and Taylor produced correlation coefficients for multiple arrangements of counties in the state of Iowa. They set the scale constant each time aggregating the base units into six counties. Table 59.2 reports the results of their analysis.

Openshaw and Taylor (1979) demonstrated that it was possible to obtain highly changeable correlation coefficients for a single set of data. They went further than this in the article by attempting to describe the universe of correlation coefficients that were possible to achieve using the different scales of zonation. For many of the scales, they claim that the theoretical range of coefficient was from –0.999 to 0.999. However, this was rarely the case for many of the zonation systems that they devised.

For instance, using 72 zones the minimum found was –0.579, and the maximum was 0.927. This demonstrates the impact of the zonation effect as differing boundary choices change the correlation coefficient values.

59.3 Approaches to Understanding

There is a vast body of research that has sought to gain greater understanding about the MAUP and how it can impact the results of statistical analysis. This section reviews work that has sought to unpick how the MAUP can lead to different results in statistical analysis. Starting with the simple examples of uni- and bivariate analysis, evidence is provided that shows the potential severity of the MAUP. This is built on by introducing research that has examined the impact of the MAUP in models that are more explicitly spatial in configuration. Attention is then paid to the role of spatial autocorrelation and spatial cross-correlation, two of the fundamental processes that lie behind incidence of the MAUP. Finally, attention is paid to the process of zonation, or zone design, and the research that has been undertaken to explore the MAUP from the perspective of the aggregation process.

59.3.1 From Univariate Statistics to Spatial Models

There are many examples of investigations into univariate and bivariate parameter instability as a consequence of the MAUP. In a recent article that clearly demonstrates the importance of preserving the availability of small area estimates for understanding societal processes, Flowerdew (2011) took 2001 Census data for England and presented an investigation on the severity of the MAUP. While there are many studies that demonstrate that it is possible to obtain different statistical results for different spatial scales and configurations, there are fewer studies that then provide statistical evidence that these differences are significant. Developing this theme and 18 common variables, Flowerdew demonstrates that even just using the three standard spatial scales that the data are released at leads to results with significant statistical differences. Flowerdew uses the Fisher transformation to standardize the correlation coefficients and concludes that after standardization the MAUP effect leads to different results in around 60 % of the cases. In general, under increasing scale aggregation, the increase in correlation coefficient is a consequence of the data smoothing properties associated with the aggregation process. As such, the variation between variables tends to decrease as aggregation increases – the heterogeneity *between* units will fall as greater number of the population are combined into single entities and the heterogeneity *within* units increases.

There are fewer examples of investigations into the MAUP in multivariate analysis. However, Fotheringham and Wong (1991) did tackle this problem using American Census data and demonstrated the problem with a regression model that related mean family income for multiple unit configurations at various spatial scales.

As with the work presented above, Fotheringham and Wong demonstrate that different spatial scales lead to systematic variability in the outcome of the regression analysis so that for some parameters (percentage elderly and percentage blue collar workers) the relationship to mean household income becomes more negative. Conversely, other parameters (percentage of home owners and the percentage of black residents) become systematically less negative as aggregation increases. Again, the importance of the MAUP is demonstrated by investigating whether the differences in the parameters obtained at the different scales and zonal configurations are statistically significant. Fotheringham and Wong use distribution of the parameter estimates and judge significance using a standard difference means test with 1.5 standard deviations (the 95% confidence level) as the key cutoff point. As they conclude, there are "many places[. . .] where the parameter estimates are significantly different" (1991, p. 1035).

Although the models above consider the MAUP, none of them explicitly incorporate the spatial structure of the data. Moving beyond these aspatial models requires a more complex modeling strategy and the explicit adoption of a spatial framework. An example of a spatial investigation into the MAUP has been conducted by Baumann and colleagues (1983). In their work, they investigate what they term as the scale hypothesis and the aggregation hypothesis (the scale effect and the zonation effect in other words) with respect to the supply of labor in multiregional labor markets. Adopting a standard MAUP approach, they suggest that the way in which the model of labor supply is measured through participation rates and commuting flows may be affected by the scale at which an analysis is conducted and the regions through which the multiple labor markets are realized. In their findings, Baumann and colleagues present a number of interesting outcomes: firstly, in terms of determining labor participation (the number of males and females in employment), the effects of scale are relatively small. Thus, there is little variation in the result as the spatial scale of the analysis is altered. However, in a model representing commuting patterns, the scale effects are much larger, a finding which intuitively makes sense as commuting is only realized in the framework when zone boundaries are crossed. Increasing the scale will, all other things being equal, reduce the number of boundaries and so the level of commuting. In surmising their findings, Baumann and colleagues highlight that the spatial framework that is adopted for an analysis is crucial, and it is "by no means admissible to ignore possible effects of the choice of a spatial framework in spatial model building" (p. 67). Finally, they suggest that when seeking out the most appropriate spatial framework, a range of criteria including model R^2, t-values, and a priori signs should be considered. This might lead the analyst to conclude, therefore, that the most appropriate spatial framework would be one that leads to the greatest level of explanation in the final model the best model performance overall. Within an econometric framework, this is an entirely reasonable assertion.

A major area of interest where the spatial organization of individual units within and between areas is segregation (see also Poulsen et al. 2011). It is a highly spatial phenomenon, and there are many examples within the literature where spatial statistics have been used to attempt to understand the role that the definition of the areal units and the scale of analysis can have on the resulting measures. Wong's investigation into segregation indices and the MAUP demonstrated that, in general, as the spatial resolution (scale) increases, the greater the degree of segregation identified (Wong 2003). As discussed above, the scale process is akin to data smoothing, so that sharp inconsistencies between smaller units are removed. Thus, as the areal units become smaller, the potential level of homogeneity within the areal unit will increase because there are fewer individual data points represented within each unit (up until the level of the single individual atomistic unit beyond which it is no longer realistic to decompose and represent a perfectly homogenous social unit). Using multiple scales of aggregation, Wong demonstrates that different scales produce different results for the dissimilarity index, D (see Duncan and Duncan 1955). To understand the impact of the MAUP scale effect, Wong proposes that the index can be decomposed into regional and local effects and that the local-level measure demonstrates the deviation of each unit from the global regional D value. The range of values achieved can give insight into how much each local unit influences the overall segregation pattern. High values record areas that deviate substantially from the global regional value, while lower values demonstrate congruence. Of course, one influence that Wong does not attempt to cover is the effect of zonation differences. It is clear however that with a small extension it would be possible to use Wong's methodology to effectively assess the impact of altering the boundaries on the resulting segregation outcomes. A second example using the diversity index, H, is used to highlight that with modification, it is possible extend the decomposition process to other segregation measures.

Two further examples of the MAUP impacting on the results of spatial statistical analysis are provided by the health literature, where research into the MAUP has been particularly active. The first study investigated the effects of the Dounreay Nuclear Power Plant in relation to instances of childhood leukemia as part of a public inquiry into an application to introduce reprocessing facilities (Heasman et al. 1984). In close proximity to the Dounreay plant were apparently high incidences of childhood leukemia. To investigate whether or not these represented significant clusters of leukemia in children, the Scottish Health Service analyzed data recording all incidences of cancer between 1968 and 1986. The initial results of the analysis reported that there was a significant excess of cases in the Dounreay area. However, at the subsequent public inquiry, a number of methodological weaknesses were identified, amongst which was the issue of boundary definition, the MAUP. Wilkie (1986) provided details of the methodological problems which included the potential gerrymandering (manipulation) of the time period studied and radial distances used to detect the cancer clusters. Creating tight boundaries around cancer points would have the effect of forcing the mortality rates upward, creating artificially high results because of the smaller population bases. Similarly, looking at a different time period, either by cutting the time series data into different lengths or curtailing the investigation at an earlier time point, would have the effect of altering the outcomes observed. Further problems arise from the presence of edge effects (cases appearing near the edge of the study space) and irregularly shaped areal units used for the aggregation. Finally, the use of areal units as a means to imprecisely locate individual incidence data introduced small errors which cumulatively could result in the erroneous generation of clusters where there were not any, or vice versa. In conclusion, the findings of the Dounreay analysis were difficult to evaluate robustly as the choice of radii and time periods for their study area "are arbitrary" (p. 266). Any clusters of cases in one area and time period could be eliminated simply through an alternative choice of radii or time periods. The second of the health examples is provided by Odoi and colleagues (2003). They were investigating the impact of the MAUP on the spatial distribution of human giardiasis (a parasitic infection causing diarrhea) in Canada. The study sets out to explicitly examine the impact of alternative spatial scales on the identification of infection clusters and whether the most appropriate statistical framework for assessing the clustering was using global or local statistics. Their analysis demonstrated that using a fine spatial scale with relatively small units enabled the detection of clusters that were hidden at the higher spatial scale. They also identified that local statistical measures provided more clustering detail than the global measures and as such were more appropriate for the exploratory analysis of patterns in spatial data.

59.3.2 The Importance of Spatial Autocorrelation

Tobler's First Law of Geography states that all things are related, but near objects are more related than distance objects (Tobler 1970). More formally, the degree of similarity is known as spatial autocorrelation, a concept developed by Michael Dacey in the 1950s at the University of Washington (see Getis 2010 for a comprehensive review). Cliff and Ord (1981) make the link between spatial autocorrelation and the MAUP more explicit, and note that the size of the cells in the areal unit system is important in determining the strength of the spatial autocorrelation. All other things being equal, larger areal units will have lower levels of autocorrelation than smaller ones. In other words, at different spatial scale, different patterns and degrees of spatial autocorrelation will be present and will impact on the structure of the data that are being analyzed.

Returning to the work of Fotheringham and Wong (1991) after assessing for the significance of the changes in parameter estimates, they investigated whether there was a link between these changes and spatial autocorrelation in the variables included in the analysis. Their conclusion was that there was little link between the severity of the MAUP and the degree of spatial autocorrelation in a (pair of) variable(s). They reinforced this conclusion by citing the examples of the percentage of black individuals and the percentage of home owners as displaying regression parameters that behaved very similarly under aggregation in terms of the significant change magnitude but that possessed very different spatial autocorrelation structures.

The work of Flowerdew and Green (1994) provides a way into understanding the properties of data with spatial autocorrelation. Using simulated data, they explore the outcomes of multiple realizations of areal units at a given scale. The use of simulated data was important as it enabled them to analyze data with known spatial autocorrelation properties in comparison with real data where spatial autocorrelations are not known and may be impacted by other (unmeasured) biases as well. Green and Flowerdew aggregated their basic grid of raw simulated data into new areal units in three ways: (a) randomly; (b) systematically, based on the value of one of the simulated variables; and (c) spatially, by combining spatially contiguous blocks. The new zones that were constructed aspatially with random aggregation show no change in the subsequent correlation or regression outcomes (although the standard error is increased as a consequence of having fewer data points); the systematic aggregation increases the correlation coefficient but has no effect on the regression parameter, while spatial aggregation alters both coefficients. In conclusion, they argue that the effects of spatial autocorrelation may "result from contiguous processes affecting the distribution of one or more of the variables being analysed, or the spatial distribution of other variables which have effects on these." This explicitly expresses the realization that the variables of areal units may display linked characteristics.

Developing their work on spatial autocorrelation further, Green and Flowerdew (1996) and Flowerdew and colleagues (2001) extend their analysis to consider the impact of spatial autocorrelation between variables as well as within variables, a phenomenon which they term "cross-correlation." They define cross-correlation as the relationship not only between variable X and variable Y at a specific point in space but also being between X and Y at neighboring points in space. In Green and Flowerdew (1996), they continue using the simulated data but this time aggregated into spatially contiguous zones. They then model the relationship between the simulated X and Y firstly using a standard regression model and then using a model that incorporates the simulated cross-correlation between X and Y. Green and Flowerdew call the cross-correlation a regional effect, and they introduce a regional term into the regression model so that there is a regression coefficient for the local effect and a regression term for the regional effect. Having used simulated data for an initial exploration, attention is then turned to repeating the analysis with real data derived from the UK population census. Setting up an investigating into unemployment and ethnicity, Green and Flowerdew find evidence that confirms their cross-correlation hypothesis and demonstrates the usefulness of the local and regional regression approaches. In Flowerdew et al. (2001) they illustrate the same concept using the example from Fotheringham and Wong (1991, see above). They theorize that cross-correlation can occur because the relationship between the "attractiveness of housing (and hence its value and the likely income of the residents) may depend not just on race and class in the immediate vicinity but also on such characteristics in neighboring areas" (Flowerdew et al. 2001, p.91). Within this work is the useful conclusion that while the presence of spatial autocorrelation is important in determining the incidence of the scale effect in correlation coefficients, it does not impact on the regression coefficients. The regression coefficients are altered when cross-correlation is present between the X and Y variable.

Arbia (1989) introduced the term “systematic spatial variation” to create a formal framework to understand the relationship between the MAUP and spatial autocorrelation using Cliff and Ord’s work (1981) as a starting point. Using data relating to the residential location of population organized on a 32 by 32 lattice, Arbia simulated the MAUP by aggregating the grid into combinations of 16 by 16, 8 by 8, 4 by 4, and 2 by 2. The results of the investigation demonstrate that with aggregation there is an increase in the level of variance and that as the level of aggregation increases, the estimates of the variance of the data become more unreliable as the number of observations diminishes with fewer degrees of freedom. Arbia concluded the effects of the MAUP under aggregation were the result of the relationships between near objects. Building on this finding, Manley et al. (2006) demonstrate that spatial autocorrelation structures rarely match the boundaries of the zones that have been used to represent the data and that these differences between the spatial extent of the autocorrelation is, in part, one of the causes of the MAUP.

Over time, more complex models were applied to the MAUP. For instance, Amrhein and Flowerdew (1989) investigated the effects of MAUP in relation to Poisson regression. The results of their analysis demonstrated that within the Poisson model there is little zonation effect to be found. However, this is not a cause for celebration by the spatial analyst because a methodology to overcome the MAUP has been identified: the lack of effect is the consequence of the analytical technique, not because the results are free from the MAUP. The finding of Amrhein and Flowerdew is important because they add a new dimension to the MAUP discussion. They demonstrate that the choice of model for an analysis is just as critical as the zonation and scale choice itself. This conclusion does not, however, mean that the world of the analyst dealing with spatial data is bleak as might initially be presumed. Amrhein (1995) uses the finding above to develop six heuristics for analysts and suggest that certain statistics and results (for instance, the standard deviation of coefficients, or the Pearson correlation coefficient) exhibit greater changes due to MAUP (scale) than other statistical methods (for instance, mean or the variance).

The work investigating spatial autocorrelation, and the related cross-correlation, has demonstrated that the MAUP is likely to be caused by the interrelated nature of the spatial variables being represented in the areal units. Thus, when aggregation is undertaken and the spatial structure of the data has a direct influence on the resulting zonations the MAUP occurs. Manley et al. (2006) further demonstrated the complexity of this problem by analyzing British Census data and showing that spatial autocorrelation rarely coincides with the boundary lines of areal units and when aggregation is undertaken it frequently incorporates small zones with differing degrees of spatial autocorrelation.

59.3.3 Exploring the MAUP Through Zone Design

A cursory overview of the statistical investigations into the MAUP would suggest that the vast majority of effort into explaining the MAUP has been concerned with the scale effect. In fact, the zonation issue has also been tackled extensively, and in some regards, with more success than the scale issue. The zonation issue research has largely focused on two aspects: how can zonations be created that are appropriate to the analytical task and what are the properties of zonation that lead to the MAUP occurring. The ability to provide multiple realizations of zonal systems within one analysis space enables the scale effect to be investigated further, as many different zonations can be derived as scale changes.

If zoning systems are problematic, then it is useful to consider why and how zoning systems may be (re)designed. The rationale behind is summed up by Openshaw and Rao (1995): "[t]he new opportunity provided by [the increasing availability of digital] boundaries is not to demonstrate the universality of MAUP effects, or to manipulate results by gerrymandering the spatial aggregation used, but it is to design new zoning systems that may help users recover from MAUP." Openshaw (1978) presented two extremes of zone design approaches to illustrate the problem. A conventional statistical approach within which spatially aggregated data can be viewed as fixed, or a model that assumes that the "undefined parameters [are] fixed, and the identification of an appropriate zoning system has to be made in some optimal manner." The first view is unacceptable due to the interdependence between the choice of zone and results achieved. From a statistical standpoint, the second solution is as poor as the first one was from a geographic perspective, as it could serve to remove the comparability between studies.

The process of zone design presents a compromise through the creation of the system that satisfies (or at least suffices) a set of criteria. One ideal outcome for a good zonal system would be a set of zones that was as simple as possible, homogenous (against a single or set of variables defined by the user), and compact. In contrast, Openshaw (1978) increased the complexity of the problem and suggested that shape (as distinct to compactness) and population size are also important elements to include. Depending on the task for which the zones are required, each of these criteria may be made more or less important. One of the first attempts at automated zone design was undertaken by Stan Openshaw (1978) with the Automatic Zoning Procedure, implemented in the Automatic Zoning Program (AZP). In more recent research, the process of zone design has become integrated with the mainstream literature around Geographical Information Systems (GIS) and enabled users to define their own zonal units. AZP was extended and became the Zone Design System (ZDES) and has been employed in a wide range of zonal scenarios. One prime example is explored in Openshaw and colleagues (1998) which commented on the first fully automated basic spatial unit (bsu) design process undertaken for the publication of the 2001 UK Census data. As Openshaw and colleagues point out, one of the major barriers to successful zone design is the realization that the problem is not one that can be tackled in the traditional software programming sense, where a global optimal solution is identifiable – if there was a global optimal solution, it is not clear how it would be identified, and in many cases there is no optimal solution. Rather, there is a range of suitable solutions which present sufficient solutions given the criteria that have been inputted.

Other systems have been developed specifically for zonal data analysis and redesign. An alternative to ZDES, is AZM (Automated Zone Matching). AZM "[i]mplements zone design on a set of zones described by polygon and arc attribute tables exported from Arc/Info or generated by users' own programs. [The program is designed to optimize] the match between two zonal systems, or the aggregation of a set of building block zones into output areas with a range of user-controlled design parameters" (Martin 2003). AZM uses the AZP procedure outlined by Openshaw (1978) and is conceptually similar. However, unlike ZDES, the AZM program was not designed specifically for the purpose of zone design. The primary function of the program is to provide a means to enable two incompatible zone coverages to be aggregated into a higher-level zone system that enables comparison (Martin 2003). However, through the input of two identical coverages, it can be used to perform an aggregation function (Martin (2003)). Nevertheless, the advantages of being able to control the aggregation process with regard to shape, key variable homogeneity, and population size mean that it is suited to the design of analytically appropriate zonal systems. In other words, zonal systems that better reflect the required uses of data, as opposed to purely "random" aggregations where there is little or no control over one or all of these factors, are not relevant in the context of research where desired scales of aggregation are required.

Finally, evidence of the potency of understanding zone design and exploiting it was presented by Boyle and Alvanides (2004). Using a case study involving the City of Leeds, and measures of deprivation, they demonstrate that it is possible to change the ranking of Leeds relative to other cities across the UK by using different boundary systems. This is of particular importance, as the European Commission was offering what are termed structural funds to aid the reduction of inequalities at a local level within member countries. Using the 1998 Index of Local Deprivation (ILD) based on the 1991 Census, as published, Leeds appeared 56th out of 57 cities. However, simply by redrawing the boundaries using alternative population thresholds to define the city area, the ranking could be changed to 11th. Applying another different criteria for the aggregation, whereby the scores were taken for wards, not local authority districts, enabled a further change in the ranking, making Leeds the 3rd most deprived city in England. The initial ranking of 56th would not have secured funding while the final ranking of 3rd would ensure a large flow of money into the city. Both of these examples highlight the potential difficulties, opportunities, and concerns that research using aggregated data should address.

59.4 Conclusion

This chapter has provided an overview of the modifiable areal unit problem (MAUP). With the growth of spatially coded data available, the potential for analysts to be confronted with areal units in analysis is increasing dramatically. Knowledge of the potential pitfalls of conducting analysis containing areal unit data is vital when dealing with areal unit data in analysis. This is true both when the areal units are the objects of the analysis as it is when the areal unit data are included to provide context to other sorts of information. In many cases, it is important to acknowledge the presence of the MAUP in analysis while accepting that the results may be conditional on the scale and zonation scheme employed.

Previous research has demonstrated that it is unlikely that a global solution to the MAUP will ever be found: indeed, to do so is to deny the inherent spatiality of the data that is under investigation, and the removal of the MAUP would be to remove the very object of interest! Previous research has also demonstrated that spatial autocorrelation and cross-correlation are likely to be very important in understanding the degree and severity of the MAUP. As such, these are key topics that the (spatial) analyst using aggregate data should be aware of and acknowledge in their analysis. Therefore, when dealing with spatially organized data, the analyst must adopt a geographically informed process of hypothesis formation. Analytical scale should become a primary factor that is explicitly considered rather than an issue that is implicitly dealt with and all too frequently assumed away in the name of pragmatism. In many cases, this will require the analyst to adopt an approach whereby multiple scales of measurement and analysis should be considered, or a highly rigorous spatial framework for an analysis constructed. This chapter is all too brief to provide a comprehensive view of all the work that has been conducted into the MAUP. Nevertheless, it hopefully sheds sufficient light on the subject and processes to provide the reader with the means to adopt a more critical and nuanced approach to their analysis.

References

Amrhein C (1995) Searching for the elusive aggregation effect: evidence from statistical simulations. Environ Plann A 27(1):105–119

Amrhein C, Flowerdew R (1989) The effect of data aggregation on a Poisson regression model of Canadian migration. In: Goodchild MF, Gopal S (eds) Accuracy of spatial databases. Taylor and Francis, London, pp 229–238

Arbia G (1989) Spatial data configuration in statistical analysis of regional economic and related problems. Kluwer, Dordrecht

Baumann J, Fischer MM, Schubert U (1983) A multiregional labour supply model for Austria: the effects of different regionalisations in labour market modelling. Pap Reg Sci Assoc 52(1):214–218

Boyle P, Alvanides S (2004) Assessing deprivation in English inner city areas: making the case for EC funding for Leeds city. In: Clarke G, Stillwell J (eds) Applied GIS and spatial analysis. Wiley, Chichester, pp 111–136

Cliff AD, Ord JK (1981) Spatial processes: models & applications. Pion, London

Duncan OD, Duncan B (1955) A methodological analysis of segregation indices. Am Sociol Rev 20:210–217

Flowerdew R (2011) How serious is the modifiable areal unit problem for analysis of English census data? Population Trends 145 (Autumn). Office for National Statistics, pp 1–13

Flowerdew R, Geddes A, Green M (2001) Behaviour of regression models under random aggregation. In: Tate NJ, Atkinson PM (eds) Modelling scale in geographical information science. Wiley, Chichester, pp 89–104

Flowerdew R, Green M (1994) Areal interpolation and types of data. In: Fotheringham S, Rogerson P (eds) Spatial analysis and GIS. Taylor and Francis, London, pp 121–145

Fotheringham AS, Wong DWS (1991) The modifiable areal unit problem in multivariate statistical analysis. Environ Plann A 23(7):1025–1044
Gehlke CE, Biehl K (1934) Certain effects of grouping upon the size of the correlation in census tract material. J Am Stat Assoc 29(Special Suppl):169–170
Getis A (2010) Spatial autocorrelation. In: Fischer MM, Getis A (eds) Handbook of applied spatial analysis. Software, methods and applications. Springer, Berlin/Heidelberg/New York, pp 255–278
Green M, Flowerdew R (1996) New evidence on the modifiable areal unit problem. In: Longley P, Batty M (eds) Spatial analysis: modelling in a GIS environment. GeoInformation International, Cambridge, pp 41–54
Heasman MA, Kemp W, Maclaren AM, Trotter P, Gillis CR, Hole DJ (1984) Incidence of leukaemia in young persons in west of Scotland. Lancet 323(8387):1188–1189
Kirby AM, Taylor PJ (1976) A geographical analysis of voting patterns in the EEC referendum. Regional Stud 10(2):183–191
Manley D, Flowerdew R, Steel D (2006) Scales, levels and processes: studying spatial patterns of British census variables computers. Environ Urban Syst 30(1):143–160
Martin D (2003) Developing the automated zoning procedure to reconcile incompatible zoning systems. Int J Geogr Inform Sci 17(1):181–196
Odoi A, Martin SW, Michel P, Holt J, Middleton D, Wilson J (2003) Geographical and temporal distribution of human giardiasis in Ontario, Canada. Int J Health Geogr 2(1):5
Openshaw S (1977) A geographical solution to the scale and aggregation problem in region-building, partitioning and spatial modeling. Trans Inst Br Geographr, New Ser 2(4):459–472
Openshaw S (1978) An empirical study of some zone-design criteria. Environ Plann A 10(7):781–794
Openshaw S (1984) The modifiable areal unit problem. CATMOG 38. GeoBooks, Norwich
Openshaw S, Rao L (1995) Algorithms for reengineering 1991 census geography. Environ Plann A 27(3):425–446
Openshaw S, Taylor PJ (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, pp 127–144
Openshaw S, Taylor PJ (1981) The modifiable areal unit problem. In: Bennet RJ, Wrigley N (eds) Quantitative geography. Routledge Kegan Paul, Henley-on-Thames, pp 60–69
Openshaw S, Alvanides S, Whalley S (1998) Some further experiments with designing output areas for the 2001 UK census. In: The paper presented at the 4th of the ESRC/JISC supported workshops Planning for the 2001 Census
Poulsen M, Johnston R, Forrest J (2011) Using local statistics and neighbourhood classifications to portray ethnic residential segregation: a London example. Environ Plann B 38(4):636–658
Tobler W (1970) A computer movie simulating urban growth in the Detroit region. Econ Geogr 46(2):234–240
Wilkie D (1986) Precognition on: review of the Scottish Health Service ISD report on Geographical distribution of leukaemia in young persons in Scotland 1968–1983 (document D/P/20) (Presented at EDRP public local inquiry, Thurso, September, 1986)
Wong D (2003) Spatial decomposition of segregation indices: a framework toward measuring segregation at multiple levels. Geograph Anal 35(3):179–194
Wong D (2009) Chapter 7. The modifiable areal unit problem (MAUP). In: Fotheringham AS (ed) The SAGE handbook of spatial analysis. Springer, Dordrecht, pp 95–112
Yule GU, Kendal MG (1950) An introduction to the theory of statistics. Charles Griffin and Company Limited, London