

**Modifiable Areal Unit Problem**

Dr. Tse-Chuan Yang

The Modifiable Areal Unit Problem (MAUP) refers to the “geographic manifestation of the ecological fallacy in which conclusions based on data aggregated to a particular set of districts may change if one aggregates the same underlying data to a different set of districts (Waller and Gotway, 2004).” The literature on MAUP can be traced back to 75 years ago when Gehlke and Biehl (1934) documented that the correlation coefficient of male juvenile delinquency differed by the scale of aggregation with the 252 census tracts in the Cleveland area. Until 1979, Openshaw and Taylor (1979) worked with the election data of the 99 counties in Iowa and first coined the term MAUP in geographical information sciences.

MAUP consists of two components: one is the scale problem or aggregation problem and the other is the grouping or zoning problem. The former concerns the different statistical inferences and estimates generated by the same data set that is aggregated into different spatial resolutions, especially aggregating small areas into a larger unit. The later refers to the variation in analytic results due to alternative grouping of the areal units at the same spatial scale (Openshaw and Taylor, 1979; Openshaw, 1984; Wong, 1996). These two components may introduce errors affecting the validity of results. Given that many policy decisions are made based on the statistical associations obtained from the analysis with aggregated data, researchers have been advised to be cautious when conducting spatial analyses on aggregated data (Unwin 1996; Bailey and Gatrell, 1995).

Currently, there is no agreement on the solution to MAUP yet, while several studies have attempted to solve it (Openshaw and Charlton 1987; Besag and Newell 1991; Gatrell et al. 1996). According to Wrigley (1995), to solve MAUP, two methodological tools must be created: an adequate statistical framework that can deal with the two problems above and a viable method that allows researchers to draw meaningful inferences about individuals on the basis of areal data analysis. Lately, it has been argued the only real resolution to the MAUP is to use individual-level data that are geocoded to a specific (usually residential) location (Weeks 2004). However, due to confidentiality and privacy issues, this approach has been rarely practiced. An alternative to this approach is to adopt a local instead of a global parameter modeling, such as geographically weighted regression (Fotheringham et al., 2002). Despite the lack of a solution to MAUP, recognizing the scale (aggregation) and grouping (zoning) problems is imperative.

Waller and Gotway (2004) provided several ways to ameliorate the effect of MAUP. First, researchers can refrain from making inferences at the individual level when analyzing aggregated data. Second, data collection can be based on the features about which researchers want to make inferences. For instance, if researchers want to make inference about individuals, they need to collect information on people. Third, researchers can use scale-independent statistics to make inferences when possible. For instance, including geography as a latent variable has been found to yield similar results in spatial linear regression models (Cressie, 1996). Statistically, choosing models whose parameters changes can be predictable should at least help researchers infer the correct magnitude and directionality of the effect of variables of interest (Tobler, 1989; Fotheringham, 1989).
References and Further Readings: