

Geo-referencing buildings to census grids: an optimization-based approach

Alexandre Kerff

University of Liège, HEC Liège Management School,
QuantOM, Research Centre for Quantitative Methods and Operations Management
e-mail: akerff@uliege.be

Jan Weymeirsch

Trier University,
Economic and Social Statistics Department
e-mail: weymeirsch@uni-trier.de

Morgane Dumont

University of Liège, HEC Liège Management School,
QuantOM, Research Centre for Quantitative Methods and Operations Management
e-mail: morgane.dumont@uliege.be

Élise Vandomme

University of Liège, HEC Liège Management School,
QuantOM, Research Centre for Quantitative Methods and Operations Management
e-mail: elise.vandomme@uliege.be

The geo-referencing of buildings and their inhabitants at detailed geographical levels can be required for multiple applications, for example, transportation planning, evaluation of housing policies, provision of health services, or flooding information. Indeed, the precise location of residence as well as place of employment of the individuals and households is crucial information in these fields. Spatial dynamic microsimulations especially project these individual units, over time and spatial dimensions, instead of aggregates to better understand behaviour at the individual level, when histories and heterogeneity are crucial. The German model MikroSim, in particular, represents the over 80 million German inhabitants and their households as a fully synthetic statistical twin and directly benefits from a detailed representation of their location.

Individuals and household data often comes from census datasets, which depict aggregate population information in a grid cell representation of various resolutions. On the other hand, micro-level buildings datasets, such as the Official Real Estate Cadastral Information System (ALKIS), represent individual buildings, their type, associated geospatial attributes and locations. The merging of these two sources of geographical information consists in a new task, the assignment of building objects to grid cells. The allocation of buildings to the census grid cells is neither fully transparent, nor conducted purely on geographic

localization alone, meaning that buildings can in principle be attributed to a neighbouring cell rather than the one they are geographically located in. It is suggested, that assignment of buildings to grid-cells occurs by address, rather than building location. However, the exact location of an address is also ambiguous. Indeed, this information is currently not published by the German National Statistical Institute (DESTATIS). Therefore, this allocation of buildings has to be conducted otherwise. This problem can be represented as a bipartite graph, where each building must then be assigned to close grid cells, ensuring that the total estimated population of assigned buildings does not exceed the grid cell's capacity. This problem is similar to the class of generalized assignment problems (GAP). In this study, we present different mathematical formulations of this problem that aim at minimising differences in estimated building inhabitants and grid cells' population level; and minimising distances between buildings and assigned grid cells. This problem presents some challenges, such as allocation of building complexes, heterogeneous building types or uncertainty in buildings capacity estimates as well as potential register errors.