



**Kandidat**

Johannes Uhl

**Masterthesis** (Jahr: 2011)

Limiting and Related Variables for Dasymetric Analysis of U.S. Census Demography

**Referent**

Prof. Dr. Heinz Saler; Prof. Dr. Barbara Buttenfield and Prof. Dr. Stefan Leyk,  
University of Colorado

**Keywords**

Dasymetric Modeling, GIS, Census Data, Python, ArcGIS Geoprocessing, Demography

**Zusammenfassung**

This master's thesis focuses on the creation and evaluation of derivation methods for limiting and related variables that can be used for the refinement of U.S. census data within a dasymetric approach. The thesis work is part of the NSF (National Science Foundation) granted research project "Putting People in Their Place: Creating a Geography for Census Microdata" and has been done at the Department of Geography, University of Colorado, Boulder (CO), USA.

Several types of U.S. Census data are freely available in the internet: Summarized for statistical entities (Census tracts, block groups and blocks) however containing less socio-economic attributes (Summary Files, SF), and richly attributed household-level records (PUMS, Public Use Microdata Sample), however published without precise spatial location (PUMA, Public Use Microdata Area). The NSF project attempts to develop a method for the creation of new summarized data on sub-tract level based on individual PUMS records, in order to enrich the summaries available as SF-1,2,3,4 data. Therefore, PUMS records are assigned to census tracts using a maximum likelihood method. Based on these assigned PUMS records, new tract summaries can be created that can be refined to sub-tract level using dasymetric modeling.

Dasymetric modeling requires ancillary data in order to exclude uninhabited regions of the census tract. Therefore, any kind of limiting ancillary data can be used. In a next step, the use of ancillary data taken from different sources can refine the demographic variable even further: If a qualitative or quantitative relationship between the demographic variable (target variable) and one or multiple related spatial variables can be established, dasymetric mapping can also be performed using the related spatial variables as ancillary data. The establishment of relationships between target and related ancillary variable is known at the most challenging part of dasymetric modeling and has a great influence on the quality of the results.

This thesis focuses on the creation of limiting and related ancillary data. Limiting ancillary data can be derived from any kind of spatial data that allows the extraction of residential areas at a scale finer than the scale to which the summarized input variable is referred (e.g. tract-level). In this thesis, several data sources (NLCD Land Use data, TIGER road network data and census block boundaries) are used in order to derive limiting ancillary data, the quality of the results is analyzed. Based on SF-3 tract summaries, limiting dasymetric mapping is realized.

In addition to that, related ancillary data is derived from different types of spatial data (DEM data and derived topographic and watershed-based variables, property taxes and hydrographic data). Related ancillary data allow the redistribution of the target variable within the extracted residential areas. Therefore, relationships between target variables and possible types of related ancillary variables are established and analyzed. The usability of certain kinds of related ancillary data for the area studied in the present work can be proved or disproved.

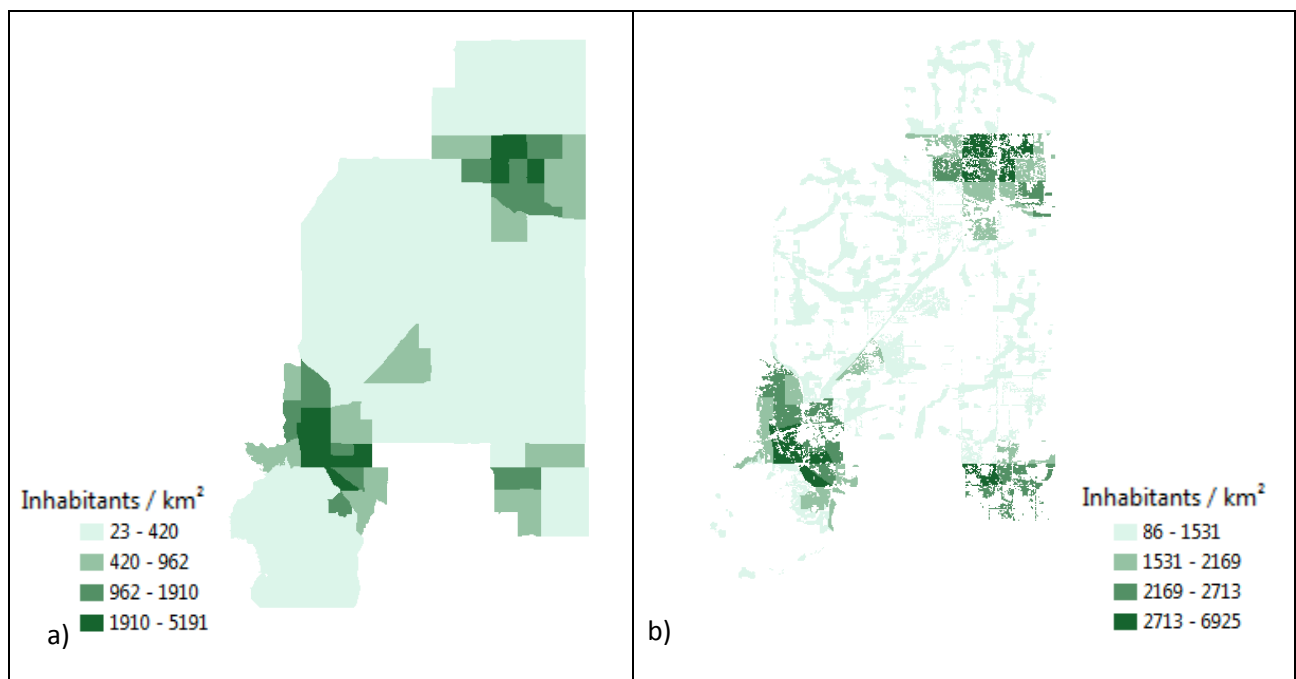
The idea of the present work is to provide ancillary variables for dasymetric modeling to the census tracts a) as limiting variables in order to exclude non-inhabited areas and b) as related

ancillary variables in order to refine the location of SF3-based target variables, such as population density and median income per household.

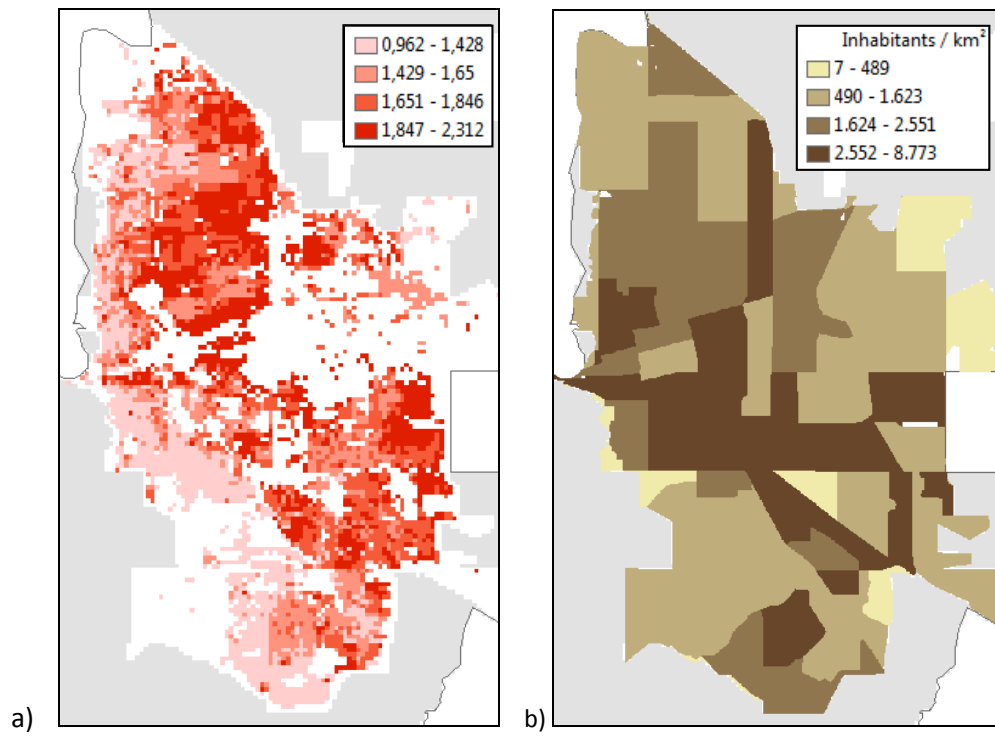
Multiple types of ancillary data are included and the results are compared. The lowest number of related variables that gives the geographically most logical result is adopted. In addition to this, these questions are attempted to be answered:

- What types of related ancillary data is available and relevant for the area of interest?
- What kind of relationships between related variables and the target variables can be established and how can this be done?
- A parsimonious solution to the problem should be found. Which and how many related variables are necessary to obtain a geographically logical result? Which variables do not contribute to a further refinement of the results?

For the realization of this work, ArcGIS geoprocessing tools have been used in combination with Python scripting. Evaluation of methods and validation of the results has been done using statistical methods in Python and SPSS, and visually making use of socio-economic knowledge about the study area.



**Figure 1: a) Initial and b) revised tract population density after applying the limiting variable for two PUMA regions in Boulder County, CO, USA.**



**Figure 2: Block group population density comparison in Boulder urban area: a) related variable prediction b) SF-3 block group summaries**