Is pycnophylactic interpolation a useful method for downscaling census attributes?

Hui Xu[1], Dan Brown[1], Shannon Brines[1]

[1] School of Natural Resources and Environment, University of Michigan, Ann Arbor

Abstract

Increasing spatial turn in the social sciences has led to increasing demand for spatially explicit demographic data with high resolution. However, publically available data are often spatially aggregated before release due to confidentiality concerns. Previous studies have developed various areal interpolation methods to downscale areal data to finer spatial units, but these studies are confined to constructing population density surfaces. It is not clear whether spatial interpolation methods, like pycnophylactic interpolation, can be a useful technique to downscale a broader set of demographic variables like age, race, and ethnicity. We downscaled census attributes of Michigan State to pixels using a modified pycnophylactic interpolation model. Comparing smoothed surface with actual data can help us to understand the spatial patterns of accuracy and variance inherited in interpolated data.

Keywords: Pycnophylatic interpolation, accuracy, downscale.

1. Introduction

Recent theoretical, methodological, and technological advances in spatial sciences provide opportunities for social scientists to include rich information about the geographic context of individuals in their studies (VanWey, et al. 2005), often based on information collected over block, tracts, counties, or other spatial units. The increasing integration of inquiry based on social and spatial location has led to rising demand for geographically rich micro-data. In responding to this demand, large national surveys typically collect survey data including both demographic attributes and geographic variables. However, publicly available data often have varying resolutions, and it is hard to access data with both high spatial and demographic resolution (Nagle, et al. 2014).

While researchers often desire demographic micro-data at fine spatial resolution (Nagle, et al. 2014), statistical agencies typically release these data at coarse spatial resolution due to concerns over confidentiality (Greenberg and Zayatz 1992). In order to disaggregate areal data to finer spatial units so that local variation can be investigated, many prior studies have explored the areal interpolation approach for population estimation (Kim and Yao 2010). Methods like area-weighted, kernel smoothing methods and geostatistical methods have been applied to downscale population data from source zones to target spatial units (Yoo, et al. 2010). However, previous studies have focused on constructing population density surface from areal population data, with little attention to the performance of areal interpolation methods on other demographic attributes, such as age, race and social-economic status. It is unclear whether spatial re-allocation methods like pycnophylactic interpolation can be useful to downscale a border set of demographic variables, other than population density, within certain geographic contexts.

This study provides an empirical investigation of the accuracy and uncertainty, and therefore utility, of areal interpolation methods when applied to downscaling demographic variables. We use Michigan as our study area, and 2010 US Census data as the source for demographic variables. Pycnophylactic interpolation (Tobler 1979) is a spatial interpolation method that has been frequently used in previous studies. This method smoothes transitions across the study area, while honoring mass in original discrete collection regions. A modified pycnophylactic model was developed using R and FORTRAN to handle a broader set of demographic variables. Using this model, we interpolated county and tract level data to finer spatial resolution pixels. These smoothed surfaces were then compared with actual data reported by the Census at lower geographic levels (i.e., tracts and blockgroups). Specifically, we calculated accuracy as the reciprocal of the mean squared error between the smoothed and actual data layers. In addition, we associated accuracy of smoothed layer with local variance of the predictor and size of census graphic units. Preliminary results show that magnitudes of interpolation errors tend to increase with local variance of the predictors but decrease as size of geographic units increases. Also, we compared accuracy of smoothed layers with those of source layers used for interpolation, assuming constant variable values across the coarser units. It seems that the largest differences between these two layers, either improvement or deterioration due to interpolation, tend to concentrate in urban areas. In this way, we provide empirical results showing the spatial pattern of accuracy and uncertainty of smoothed data layers generated using pycnophylactic interpolation.

References

Greenberg, BV, and Laura V Zayatz

1992 Strategies for measuring risk in public use microdata files. Statistica Neerlandica 46(1):33-48.

Kim, Hwahwan, and Xiaobai Yao

2010 Pycnophylactic interpolation revisited: integration with the dasymetric-mapping method. International Journal of Remote Sensing 31(21):5657-5671.

Nagle, Nicholas N, et al.

2014 Dasymetric modeling and uncertainty. Annals of the Association of American Geographers 104(1):80-95.

Tobler, Waldo R

1979 Smooth pycnophylactic interpolation for geographical regions. Journal of the American Statistical Association 74(367):519-530.

VanWey, Leah K, et al.

2005 Confidentiality and spatially explicit data: Concerns and challenges. Proceedings of the National Academy of Sciences of the United States of America 102(43):15337-15342.

Yoo, Eun-Hye, Phaedon C Kyriakidis, and Waldo Tobler

2010 Reconstructing Population Density Surfaces from Areal Data: A Comparison of Tobler's Pycnophylactic Interpolation Method and Area-to-Point Kriging. Geographical analysis 42(1):78-98.