Spatial Interpolation of Socioeconomic Conditions for Schools

Laura Nixon

U.S. Census Bureau

June 28, 2016

Problem - Need measures of socioeconomic status for neighborhoods for education programs

Proposal - Incorporate data from the U.S. Census Bureau

1) ACS Tabulations
2) Spatially Interpolated Demographic Estimates

What is socioeconomic status (SES)?

Common components are:

- Income/Poverty
- Parental Educational Attainment
- Parental Occupational Status and Type

What is a neighborhood?

- Census Tracts
- Block Groups
- Zip Codes
- Custom defined areas

Why American Community Survey (ACS)?

- Collected consistently
- Large survey
- Available for sub-state geographic areas
- Content includes primary items that make up SES
- Content also includes demographic information correlated with SES, like race and household type

Census Neighborhood Experiment for NAEP: American Community Survey Tabulation

The link between NAEP and Census tabulated data is Zip Code

Experiment Request:

Estimates of households with children

- By household type and racial composition
- For Zip Code Tabulation Areas (ZCTAs)
- From American Community Survey (ACS) 5 year data (2007 – 2011)

Obstacles with ZCTAs

- Only updated once every 10 years
- Created for mail service efficiencies - not representative of "neighborhood"
- Sizes vary
- Many ZCTAs do not have enough sample cases to meet disclosure concerns
- Poor quality for detailed household groups
- May need to combine ZCTAs for sufficient sample
- Combined ZCTAS could result in very large areas

What if a neighborhood is not predefined?

Can we use ACS data and spatial interpolations methods to create estimates based on surrounding sample cases?

Spatially Interpolated Demographic Estimates (SIDE)

Tobler's Law - *Everything is related to everything else, but near things are more related than distant things.*

Allow "neighborhoods" to be based on a particular number of inputs instead of predefined area

Predict estimate at any point location (anchor point)

Spatial Interpolation – Kriging

- Is a family of techniques for modeling and analyzing spatially continuous data.
- Assumes that the value of sample cases partly depends on their location and distance to other cases.
- Creates semivariogram based on a mathematical formula to make a surface of predicted values.
- Provides estimates for unsampled locations based on information from sampled locations. The closer the sampled location is to the unsampled location, the more weight the sampled value has on the final predicted value.
- Provides errors for the reliability of the predicted estimates.

Empirical Bayesian Kriging (EBK)

- Uses restricted maximum likelihood to estimate semivariogram.
- Uses repeated simulations to produce a large collection of regionalized semivariograms and uses properties of the semivariogram distribution to create prediction surface.
- Accounts for potential error inherent in other kriging methods that systematically underestimate the standard error of predicted values.
- Offers a more robust response to the problem of non-stationarity that may exist within a large study area.

SIDE test - inputs

- Ohio
- ACS 2009-2013
- Universe – Households with children ages 5 to 18 who are related to the householder and have poverty defined
- 9 household groups
  - Race/ethnicity and household type
- Continuous variable - Income to Poverty Ratio
  - Bottom-coded at 0 and top-coded at 999

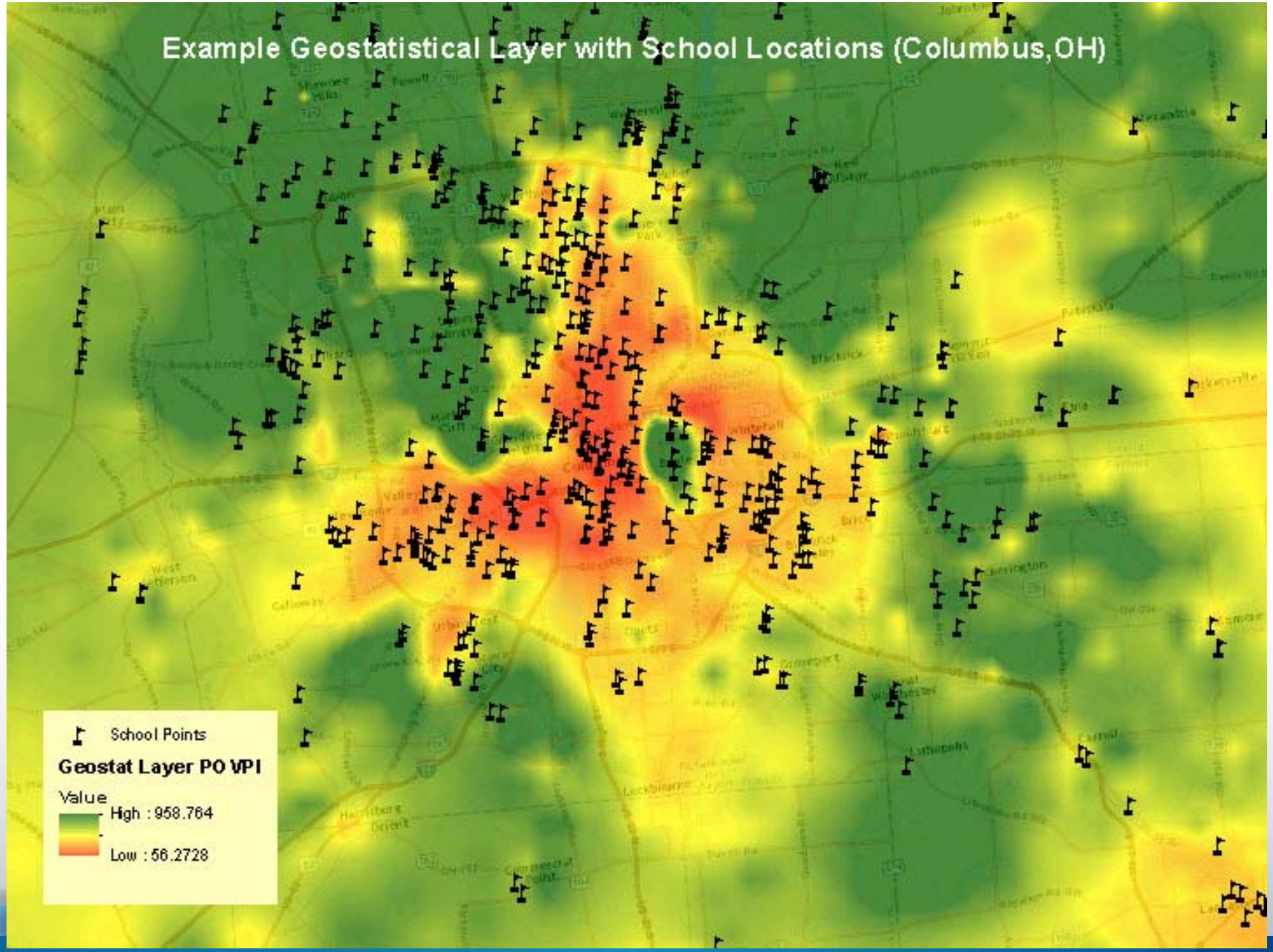
- Locations of schools serving 4th graders from the 2012/13 Common Core of Data were used as neighborhood anchor points

SIDE method

Use block centroids and ratio of income to poverty value for ACS sample cases to create a geostatistical layer for each household group using EBK.

Predict estimated values and spatial standard error of the ratio of income to poverty at 1,793 school locations (latitude and longitude) using prediction surfaces.

Example Geostatistical Layer with School Locations (Columbus,OH)



Examining Spatially Interpolated Demographic Estimates

- Do the estimates show plausible relationships with other demographic characteristics like race or household type?
- Are the neighborhoods surrounding the anchor point a reasonable size?
- How do the estimates compare with other SES measures?
- Do the estimates correlate with educational achievement?

SIDE "Neighborhoods"

Concept – Nearest 25 sample cases to the school point make up the neighborhood.

Size of neighborhood:

- Average size of "All households" neighborhood is less than 3 square miles. Median size is 0.59 square miles.
- "Other Race" household groups have smallest sample and largest neighborhood sizes, especially those neighborhoods anchored around rural schools.

ZCTA average size is 34 square miles (median is 23 square miles)

SIDE Limitations

- Continuous variables
- Poverty Index
- ACS limitations
- Standard errors
- Inaccurate school locations

SIDE Benefits

- All schools have estimates.
- All estimates are based on the same number of sample cases.
- Measures of error indicate spatial confidence.
- Interpolation methods prevent disclosure concerns.
- Estimates are not constrained by predefined boundaries.
- Household groups allow for differentiation.
- Potential to create for other social or economic conditions.

Questions?

Contact information

Laura Nixon

U.S. Census Bureau

laura.c.nixon@census.gov

Disclaimer: This presentation is to inform interested parties of research and to encourage discussion. The views expressed are those of the author and not necessarily those of the U.S. Census Bureau