Spatializing Global Urban Extent:
A Source Driven Approach

J. J. Moehl[1], A. N. Rose[1], E. A. Bright[1]

[1]Oak Ridge National Laboratory, Oak Ridge, TN 37831
Email: moehljj;rosean;brightea@ornl.gov

Abstract

"Urban" is something that intuitively feels very well defined; however, when it comes time to express this idea on a map, things get complicated. Definitions of "urban" vary globally, and as such there is not universal understanding of what makes a given place "urban". The common approach to spatially defining urban extents is through remotely sensed imagery. The alternative approach presented in this paper uses the percent of population in urban areas, which is a common macroeconomic (country-level) variable with the definition for urban generally defined by each country's statistical office, along with temporally-aligned high-resolution population data to spatially define urban extents for each country. Because the percent urban number is defined by the same producer as other urban/rural defined statistical data, such as household characteristics or birth rate, understanding the spatial aspect of urban from the same perspective is ideal for high resolution spatial modeling of these other phenomena.

1. Introduction

At Oak Ridge National Laboratory, we've been modeling population for nearly two decades. Whether using a top down approach such as the LandScan Global model (Dobson *et al.* 2000), or a hybrid top down/bottom up approach as is used to develop LandScan USA (Bhaduri *et al.* 2007), spatial data that aligns with tabular datasets is essential. Ensuring that national and subnational boundaries match the data from statistical organizations is extremely important for dasymetric mapping, as these boundaries play a large role in appropriately disaggregating zonal summaries. Many summary data values, such as population counts, birth rates, average household size, and age distributions are listed not only by administrative zone, but often times they are further listed by whether they occur in urban or rural areas. Spatial data for administrative boundaries are widely available, although of varying quality and level of detail. However, very rarely are spatial data available with urban area delineation from the perspective of the statistical organization.

Summary data designated as urban or rural for each administrative zone within the country is common. The idea of what is "urban" is defined and implemented by each country's statistical office. The UN's *World Urbanization Prospects: The 2014 Revision* provides an exhaustive comparison of these definitions which vary widely in their considerations (UN 2014). This may be an extremely detailed formal definition like the one provided by the US Census Bureau (US Census 2011) complete with spatial data freely available in multiple formats, or very generally defined with no publicly accessible spatial data. Administrative data and population density are the most common criteria used by countries to fit the urban definition, with 125 and 137 countries respectively using each criterion in part or whole (UN 2014). The HYDE datasets, which provide historical population and land use estimates, combine historical urban population estimates with population density in their models; however, these datasets are historical and at a coarser ~10 sq km resolution (Klein Goldewijk 2001, 2005).

2. Method

Our motivation for this study stems from a project where we are trying to model housing characteristics at a fine spatial resolution. These data are subnational and also include an urban/rural tag. This urban vs. rural designation often captures the variation in housing characteristics within the subnational zones. As such, having an urban extent is extremely important. We began exploring sources of spatial data for urban extent with global coverage. Potere and Schneider (2007) provide a useful comparison of these global urban area maps; the global urban area estimates ranged from 276,000 to 3.5 million sq. km. with variation coming from temporal, sensor, and methodological differences. While these datasets have been used to understand past and future urban expansion (Seto 2011), we felt these datasets were inadequate for our purpose. Our data summaries are provided by national statistical organizations whose decisions about which areas are urban generally have little to do with interpretations of remotely sensed data used to describe urban land cover. Thus, in many ways, it is irrelevant whether a particular spot in any given country could spectrally be described as urban.

Our goal is to accurately represent the urban area described by the summary data from the perspective of the producer of that statistical data. To model that extent, we combined other information from that producer, namely the percent of population living in urban areas, with high-resolution population data, LandScan Global. We are assuming that urban populations live in the most densely populated places and we otherwise know population density is often a factor in defining urban (UN 2014). With this scenario, it is possible to triangulate locally adaptive density thresholds for "urban" by summarizing the population, cell by cell, in each country from the most densely populated cell to the least. We use this population sum to determine the cumulative percentage of urban population at or above each density value. This percentage is then matched with the percent urban population provided by statistical organizations to determine the "break point" dividing urban and rural for each country. We use the break point to classify cells with greater or equal densities as urban for each country.

The first requirement for this process is a population dataset with high spatial resolution as well as global coverage. The LandScan Global dataset provides the granularity needed to calculate the most precise urban extent estimate possible. We developed an area grid of the same extent and resolution as the LandScan data in order to calculate population density. The next step in this process is to acquire urban population percentage data for the same time period as the LandScan data, 2014. There are several sources of this data. One option is to acquire this data from each national statistical organization. However, because this would largely repeat the processes employed by other organization whose mission is to acquire and assimilate such data, we feel it is sufficient to use data from these providers. Two producers, the CIA World Factbook (CIA 2016) and the United Nations Population Division (UN 2014), use similar methods; each acquires data from statistical organizations and adjusts them when necessary, for example when data are of poor quality. Figure 1 shows the CIA World Factbook versus the UN break point values; they are very similar and either is sufficient for our purposes. More importantly, Figure 1 shows the variation in the density break values. This illustrates the fact that "urban" means very different things statistically across the globe.

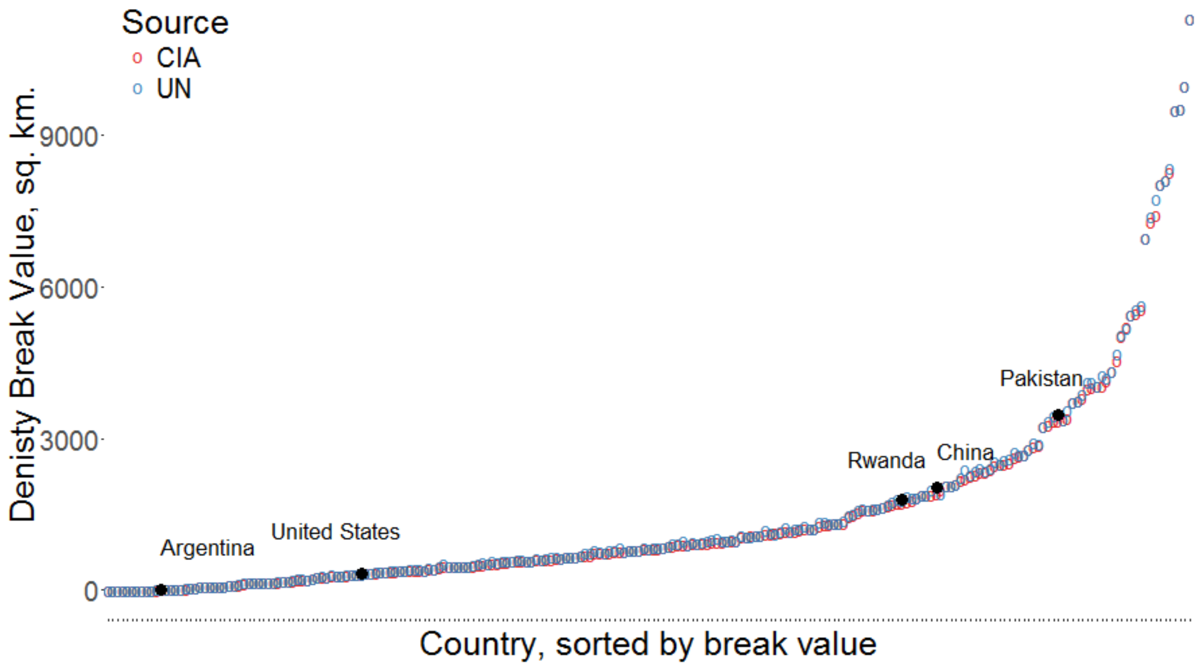

Figure 1. The population density value dividing urban and rural in our new method varies across the globe. Also, the CIA World Factbook and the UN Population Division data are extremely similar.

3. Results

Next we compared our urban extent with the GRUMP v1 urban extent layer (CIESIN 2011). GRUMP relies largely on nighttime lights and has been shown to estimate a larger urban area extent than other datasets (Potere and Schneider 2007). To calculate the total urban extent in GRUMP we used the same country boundaries and the same area grid and summarised the urban area in each country. Figure 2 shows a comparison of our new method and the GRUMP urban area extent. Each point in the chart represents the percentage of the total area, which is the sum of the GRUMP estimate and our method, accounted for by our new method. This allows for a relative examination of each country. The line at 0.5 means each country had 50% of the total, so they have equal area totals for the country. A break at 0.33 or 0.66 means one source has twice as much urban area in that country. The size of the points shows the urban area in sq. km. according to the new method. While most countries have much more urban area in GRUMP, there is no regional pattern to the variation, which shows the great difference in methodologies defining “urban” at the country level. For example, we estimate a much greater urban extent in Argentina than GRUMP due to the definition of urban in Argentina being “Localities with 2,000 inhabitants or more” (UN 2014). Figure 3 provides an example of the differences in spatial distributions between these models.

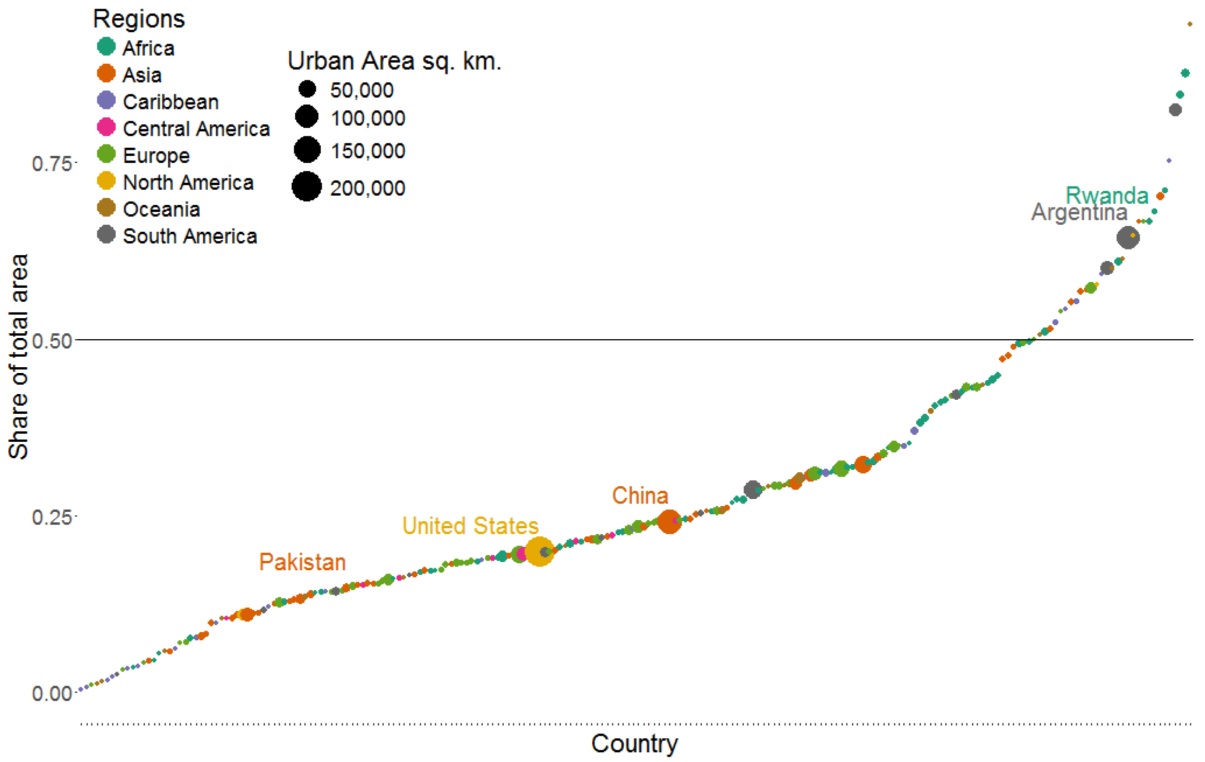

Figure 2: The new method urban area estimate shown as a percentage of the total area for the GRUMP estimate and the new estimate. Values above the line show where the new method estimates more urban area than GRUMP.

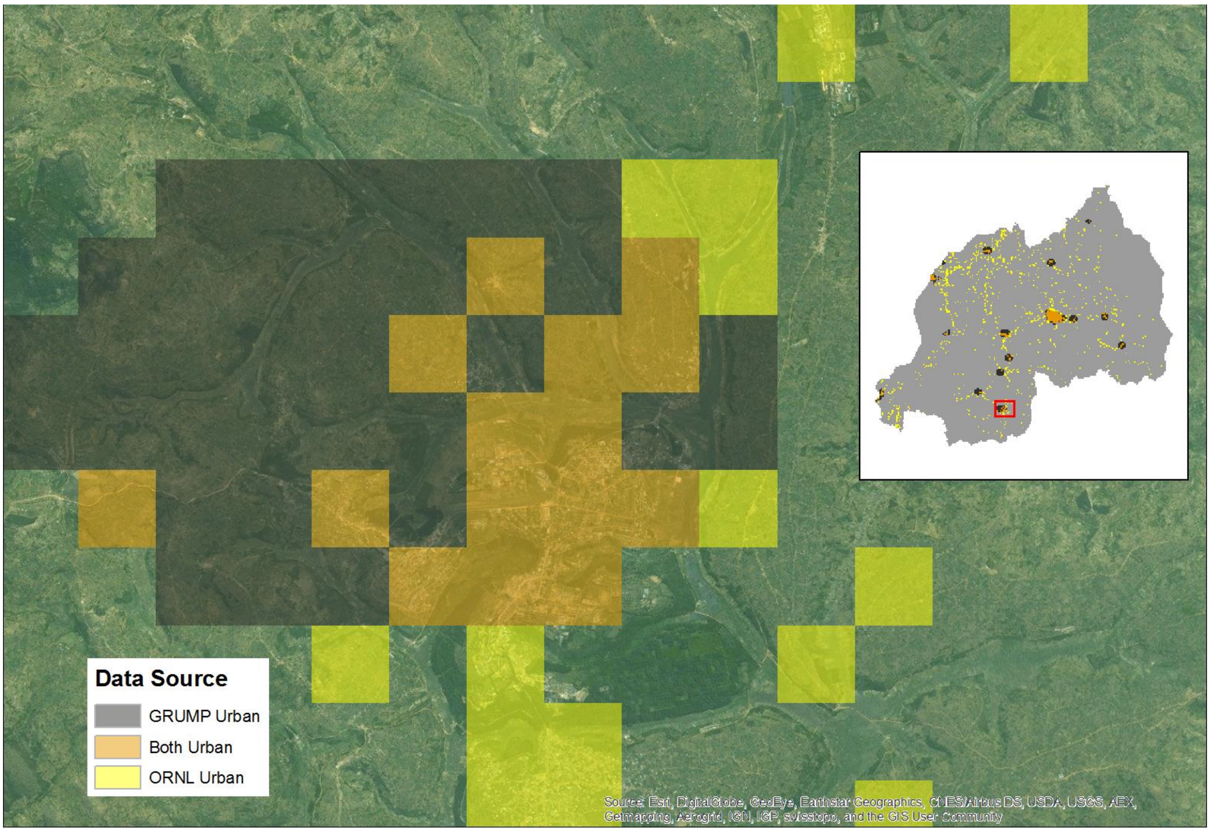

Figure 3: Close up and country level (inset) comparison of ORNL urban extent and GRUMP for Rwanda show how differences in methodology drive varied spatial distributions of urban and thus result in large differences in overall urban area between the models.

4. Conclusion

We conclude based on these results that it is appropriate to use data from the perspective of the statistical organization to calculate urban extent rather than to rely solely on remotely sensed imaging sources of "urban" data when trying to spatialize other phenomena from the same statistical organization. Further research is needed for a more in-depth comparison of urban extent as finer, city-level scales to fully test the veracity of our model. Additionally, explicitly including other data such as administrative boundary data may increase the accuracy of our model as these are often considered when defining "urban" at the local level.

Copyright

References

US Census Bureau, 2011. *Urban Area Criteria for the 2010 Census*. Department of Commerce, 76. *Federal Register* 164 (2011); 53030–53043.

Center for International Earth Science Information Network - CIESIN - Columbia University, International Food Policy Research Institute - IFPRI, The World Bank, and Centro Internacional de Agricultura Tropical - CIAT. 2011. Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Urban Extents Grid. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4GH9FVG. Accessed 28 April 2016.

Bhaduri, B., E. Bright, P. Coleman and M. Urban, 2007. LandScan USA: a high-resolution geospatial and temporal modeling approach for population distribution and dynamics. *GeoJournal* 69(1): 103-117.

Bright, E. A., Rose, A. N., & Urban, M. L. (2015). *LandScan 2014* [digital raster data]. Retrieved from: http://www.ornl.gov/landscan/

Central Intelligence Agency, 2016. The World Factbook 2015-2016. Washington, DC: Central Intelligence Agency, 2016.

Dobson, J., Bright, E., Coleman, P., Durfee, R., & Worley, B., 2000. LandScan: A global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing,* 66(7): 849-857.

Klein Goldewijk, K., 2001. Estimating global land use change over the past 300 years: the HYDE database. *Global Biogeochemical Cycles*. 15(2): 417-434.

Klein Goldewijk, K. 2005. Three centuries of global population growth: A spatial referenced population density database for 1700 – 2000. *Population and Environment* 26(5): 343-367.

Potere, D., and Schneider, A., 2007, A critical look at representations of urban areas in global maps. *GeoJournal* 69(1):55-80.

Seto, K., Fragkias, M., Güneralp, B., Reilly, MK., 2011. A Meta-Analysis of Global Urban Land Expansion. *PLoS ONE* 6(8): e23777.

United Nations, Department of Economic and Social Affairs, Population Division, 2014. World Urbanization Prospects: The 2014 Revision, CD-ROM Edition.