Abstract: World population increasingly lives in urban areas. Each country uses its own criteria to define urbanization level (urban percentage of population). To harmonize global urbanization analysis, the U.N. Statistical Commission endorsed a new methodology called the degree of urbanization and developed by the European Commission. It employs GHS-SMOD, a global satellite product categorizing each 1 km terrestrial cell to seven categories along the rural-urban continuum, based on their population and built-up presence. Studies show derived national urbanization levels using the new methodology may match well those by administrative definitions in some countries, but not so in others. To understand characteristics of GHS-SMOD to underlie globally harmonized urbanization analysis, this paper takes a further step to explore disparities in urbanization levels of subnational units from the two sources. This is important at the subnational level as many countries do not report urbanization records at this level while they matter for regional policy areas including health, housing, transportation, and climate change. Therefore, we compare subnational GHS-SMOD and census-based urbanization levels for a selection of countries. Our analysis reveals different levels of agreement between these two sources, mostly attributable to the urban status and criteria employed by national censuses.

Introduction

In this research, we investigate the agreement between subnational urbanization levels (urban percentage of population) estimated from GHS-SMOD (GHS Settlement Model layers) (Florczyk et al. 2019; Pesaresi et al. 2019) and those from official censuses in the continental United States, Mexico, and India. We select these countries because of their different stages of urbanization, as well as distinct criteria to define urban. This helps understand GHS-SMOD's characteristics in providing an alternative account of urbanization levels within these countries. GHS-SMOD has the unique capability of harmonizing global urbanization analysis, which is lacking from national censuses due to divergent criteria and shortage of records in data-poor countries. Understanding urbanization trends is important for public policy in terms of health, transportation, and climate change to name a few.

The U.S. Census employs a pure form of statistical urban classification based on contiguity, population density, land-use, and total population size (Balk et al. 2018), placing a mutually exclusive urban/rural class on blocks as its smallest set of administrative units. The National Institute of Statistics and Geography in Mexico also utilizes a statistical approach, classifying all localities with 2500 inhabitants or more as urban (United Nations, Department of Economic and Social Affairs 2019). India's official definition, on the other hand, has adopted a hybrid statistical/jurisdictional perspective with urban populations being those overseen by legally urban local governments (with some allowance for census towns and outgrowths, which are treated as urban by census authorities but are legally rural). This may lead to under-estimation of urban land and population in India stemming from the reluctance of some states to allow their large, urban-like settlements to be declared urban (Balk et al. 2019). In terms of inclusiveness, the U.S. definition ranks the highest followed by Mexico and India.

These countries are at different stages of urbanization. While the U.S. and Mexico are highly urbanized (82% and 79% urban in 2015), India is one of the least urbanized countries in the world (33% urban in 2015) although there is a strong upward urbanization trend in the country (United Nations, Department of Economic and Social Affairs 2019). Notably, all these countries share spatial disparity in urbanization levels across their subnational units. According to the census records of these countries, the urbanization level ranges between 39% and 95% across U.S. states in 2010 while the range is from 48% to 99% in Mexico in 2015 and from 10% to 98% in India in 2010. These heterogeneities facilitate interpretations on the comparison between GHS-SMOD based urbanization levels and the de jure records, which in turn could lead to more plausible understandings of de-facto subnational urbanization levels within these countries.

Datasets and Methods

GHS-SMOD delineates and classifies settlement typologies via a logic of cell clusters, population size, and population and built-up area densities in four epochs of 1975, 1990, 2000, and 2015. This multi-temporal global remote sensing product assigns a class to each 1 km cell along the rural to urban continuum. The classes are urban center, dense urban cluster, semi-dense urban cluster, suburban, rural, low density rural, very low density rural and water (Florczyk et al. 2019). We determine the first four classes as urban and extract the population of cells as such from another global remote-sensing product, called GHS-Pop (Freire et al. 2016). GHS-Pop is available at different spatial resolutions, including 250 m and 1 km, and its 1 km edition spatially aligns with GHS-SMOD. This facilitates population extraction for cells deemed urban. By summarizing total and urban population grids for each subnational administrative unit, we can reach both urban and total population aggregates of that unit and consequently its urbanization level. We then compare these figures with official census-based records. We select year 2000 for our comparisons in the United States and India, and 2015 in Mexico as these are the common years with both GHS-SMOD and census-based levels available for these countries.

Results and Discussion

Figures 1, 2 and 3 compare subnational GHS-SMOD and census-based urbanization estimates for the Unites States, Mexico and India, respectively. Tabe 1 also summarizes comparative analysis between both sources in these countries.

Figure 1 shows that in all U.S. states, regardless of their urbanization level, official census records are greater, pointing to the inclusiveness of the U.S. census criteria that encompass all factors that also underlie GHS-SMOD specifications, including populatiotion size, density, built-up presense, and spatial contiguity. Table 1 shows that percentage differences are always negative in the U.S., indicating larger census-based urbanization values. Notably, Vermont as the state with the lowest official urbanization level (38.2%) presents the greatest dissimilarity while in California, as the most urbanized state (94.4%), the agreement between the two sources is at its peak.

Figure 2 demonstrates that the two sources lead to similar urbanization levels in Mexico. This indicates that the condition of Mexican localities with population greater than 2500 inhabitants conforms to the criteria of GHS-SMOD. In contrast to the U.S., however, there are states whose official urbanization levels are smaller. According to Table 1, these positive differences do not exceed 20%. In general, more urbanized states present higher similarity in their urbanization levels from the two sources although some exceptions exist. The most positive dissimilarity at 20% occurs in Hidalgo, which is only 52.4% urban according to the official census records. On the other hand, the most negative difference at -11.3% pertains to Yucatan whose official urbanization level (84.5) is above the national level.

India, on the other hand, presents completley different outcomes. According to Figure 3, the GHS-SMOD urbanization levels are much larger than official records, and based on Table 1, up to 7.5 times higher. This indicates the combination of jurisdictinal discretions with statistical criteria result in subnational urbanization levels that differ significantly from the pure statsitcial approach implemended in GHS-SMOD. Particulalrly, this disparity pertains to the reluctance of some Indian states to declare their large, urban-like settlements as urban, while the population-based statsitcial approach of GHS-SMOD identifies them as urban, nonetheless. This, in turn, leads to much larger estimates of state-level urbanization by the latter source. Differences in the two sources negatively correlate with official urbanization levels, where the smallest difference at 7.3% is observed in Delhi, as the most urbanized state (93%), and the largest difference at 656% pertains to Bihar as the India’s second least urbanized state (10.5%).

Table 1. Summary of comparisons between census and GHS-SMOD based urbanization levels.

Country	Difference Range* (%)	# Positive Differences	# Negative Differences
USA	-38 to -3.5	0/48	48/48
Mexico	-11.3 to 20	14/32	18/32
India	7.3 to 656	35/35	0/35

*Difference = (GHS-SMOD – Census) / Census * 100

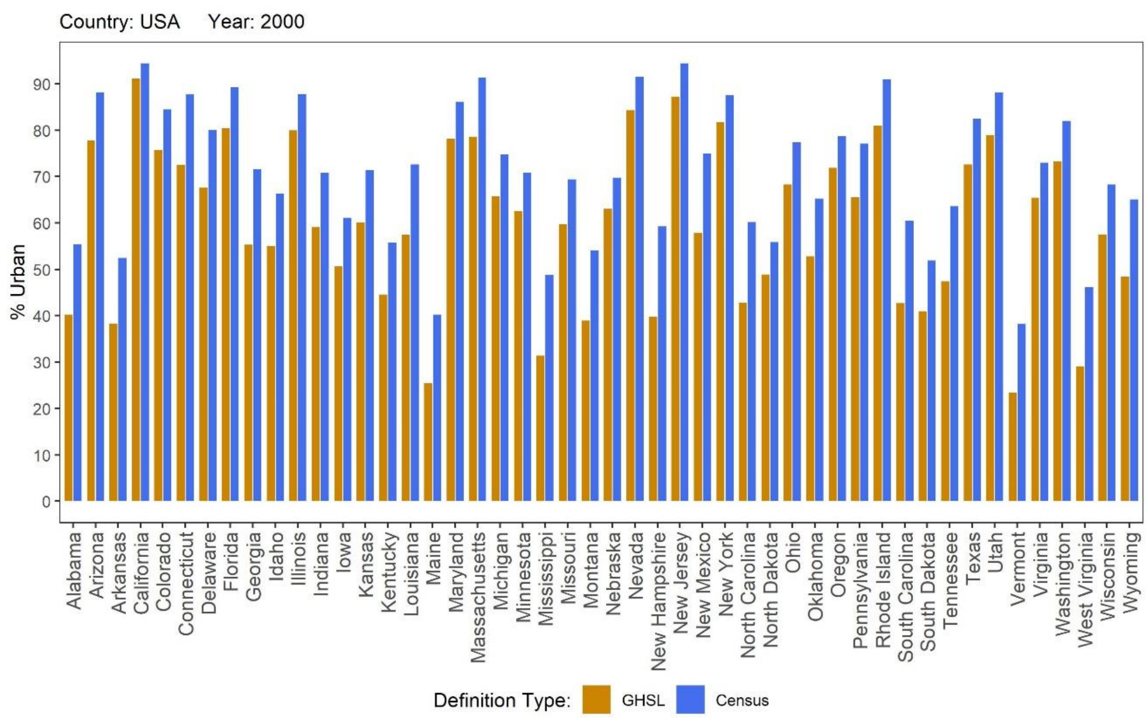

Figure 1. Comparison of Census and GHS-SMOD-based urbanization levels in the continental USA.

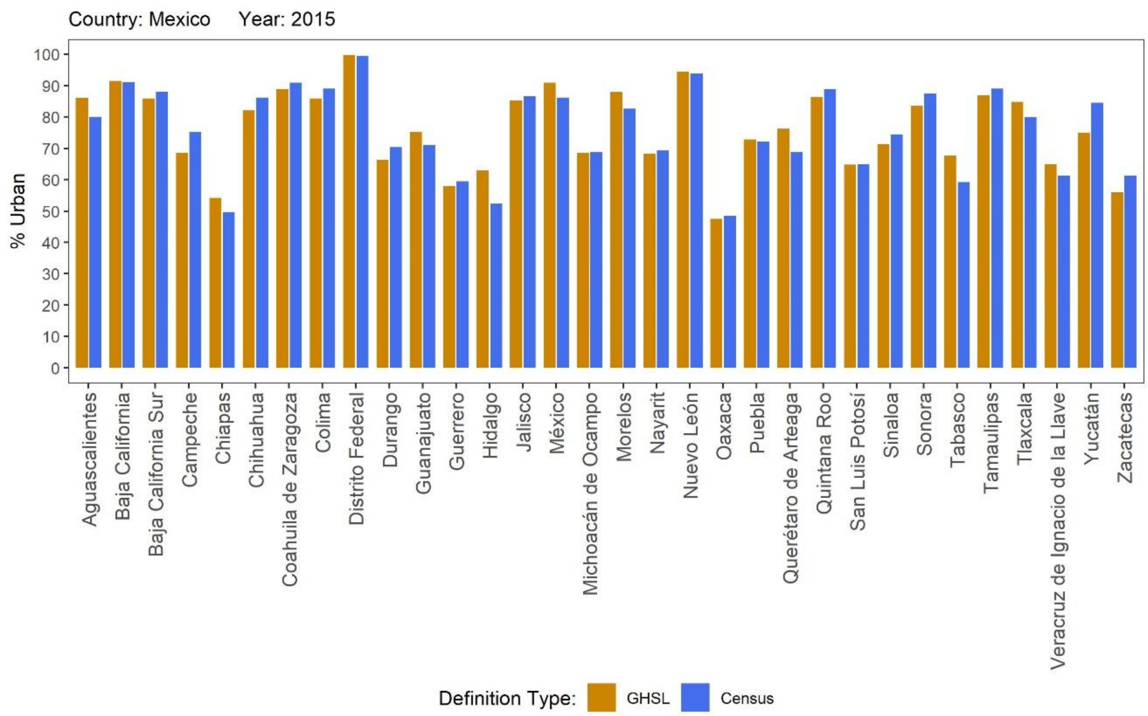

Figure 2. Comparison of Census and GHS-SMOD-based urbanization levels in Mexico.

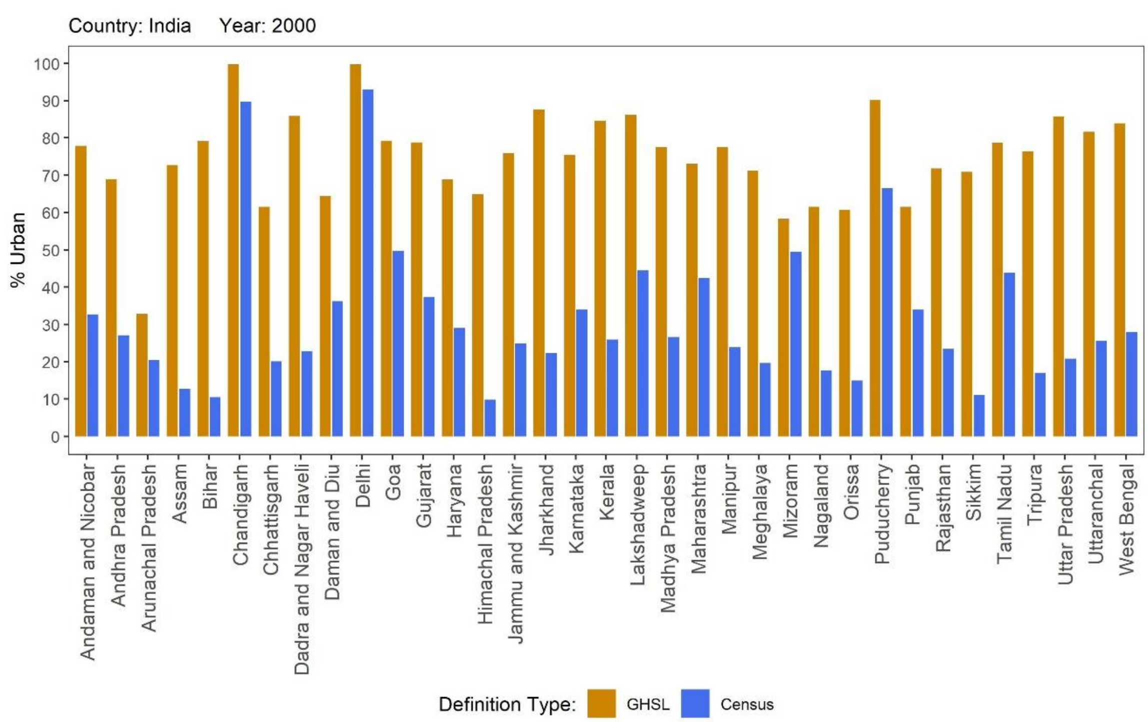

Figure 3. Comparison of Census and GHS-SMOD-based urbanization levels in India.

Conclusions

This research offers a comparative insight between census and GHS-SMOD subnational urbanization levels in three countries with different urbanization status, spatial heterogeneity, and distinct census criteria. Notably, this work does not aim at validating one source based on the other. Rather, it illuminates how GHS-SMOD based urbanization levels, stemming from an objective set of statistical criteria, should be perceived in different contexts. GHS-SMOD is a remote sensing product whose global coverage and harmonized set of defining criteria facilitate consistent comparative urbanization analysis across countries and at various geographic scales, even for regions of the world where official census records may be lacking. It also provides another perspective of global urbanization status merely determined by spatial population and built-up distributions. In contrast to official censuses, GHS-SMOD delineates the spatial distribution of different classes along the rural-urban continuum and is not limited to hard urban or rural classification. All of this is beneficial to human-climate dynamics, resource allocation, and health monitoring of population. It is important, nonetheless, to investigate and be cognizant of its differences with census records where possible, as this comparison allows uncertainty assessment and informed interpretations of GHS-SMOD outcomes. Therefore, we plan to enrich our comparative analysis and expand it to different parts of the world.

References

Balk, D., Leyk, S., Jones, B., Montgomery, M.R., and Clark, A. (2018). Understanding urbanization: A study of census and satellite-derived urban classes in the United States, 1990-2010. *PLoS ONE* 13(12):e0208487. doi:10.1371/journal.pone.0208487.

Balk, D., Montgomery, M.R., Engin, H., Lin, N., Major, E., and Jones, B. (2019). Urbanization in India: Population and urban classification grids for 2011. *Data* 4(1):35. doi:10.3390/data4010035.

Florczyk, A.J., Corbane, C., Ehrlich, D., Freire, S., Kemper, T., Maffenini, L., Melchiorri, M., Pesaresi, M., Politis, P., Schiavina, M., Sabo, F., and Zanchetta, L. (2019). *GHSL Data Package 2019*. Luxembourg: Publications Office of the European Union.

Freire, S., MacManus, K., Pesaresi, M., Doxsey-Whitfield, E., and Mills, J. (2016). Development of new open and free multi-temporal global population grids at 250 m resolution. *Association of Geographic Information Laboratories in Europe (AGILE)*.

Pesaresi, M., Florczyk, A., Schiavina, M., Melchiorri, M., and Maffenini, L. (2019). GHS Settlement Grid, Updated and Refined REGIO Model 2014 in Application to GHS-BUILT R2018A and GHS-POP R2019A, Multitemporal (1975-1990-2000-2015) R2019A. *European Commission, Joint Research Centre (JRC)*. doi:10.2905/42E8BE89-54FF-464E-BE7B-BF9E64DA5218.

United Nations, Department of Economic and Social Affairs, P.D. (2019). *World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420)*. New York: United Nations.