

Article

Global Harmonization of Urbanization Measures: Proceed with Care

Deborah Balk ^{1,2,*}, Stefan Leyk ^{3,4,†}, Mark R. Montgomery ^{5,6,†} and Hasim Engin ^{1,7,†}

¹ CUNY Institute for Demographic Research (CIDR), City University of New York, New York, NY 10010, USA

² Marx School of Public and International Affairs, Baruch College, City University of New York, New York, NY 10010, USA

³ Department of Geography, University of Colorado Boulder, Boulder, CO 80309, USA; stefan.leyk@colorado.edu

⁴ Institute of Behavioral Science, University of Colorado Boulder, Boulder, CO 80309, USA

⁵ Department of Economics, Stony Brook University, Stony Brook, NY 11794, USA; mark.montgomery@stonybrook.edu

⁶ Population Council, New York, NY 10017, USA

⁷ Center for International Earth Science Network (CIESIN), The Earth Institute, Columbia University, Palisades, NY 10964, USA; hengin@ciesin.columbia.edu

* Correspondence: deborah.balk@baruch.cuny.edu; Tel.: +1-646-660-6762

† All authors contributed equally to this research.

Abstract: By 2050, two-thirds of the world's population is expected to be living in cities and towns, a marked increase from today's level of 55 percent. If the general trend is unmistakable, efforts to measure it precisely have been beset with difficulties: the criteria defining urban areas, cities and towns differ from one country to the next and can also change over time for any given country. The past decade has seen great progress toward the long-awaited goal of scientifically comparable urbanization measures, thanks to the combined efforts of multiple disciplines. These efforts have been organized around what is termed the "statistical urbanization" concept, whereby urban areas are defined by population density, contiguity and total population size. Data derived from remote-sensing methods can now supply a variety of spatial proxies for urban areas defined in this way. However, it remains to be understood how such proxies complement, or depart from, meaningful country-specific alternatives. In this paper, we investigate finely resolved population census and satellite-derived data for the United States, Mexico and India, three countries with widely varying conceptions of urban places and long histories of debate and refinement of their national criteria. At the extremes of the urban–rural continuum, we find evidence of generally good agreement between the national and remote sensing-derived measures (albeit with variation by country), but identify significant disagreements in the middle ranges where today's urban policies are often focused.

Keywords: demographics; spatial; economic geography; urban economics; spatial demography; urbanness; built environment; rural–urban continuum; remote sensing

JEL Classification: J1; R0



Citation: Balk, D.; Leyk, S.; Montgomery, M.R.; Engin, H. Global Harmonization of Urbanization Measures: Proceed with Care. *Remote Sens.* **2021**, *13*, 4973. <https://doi.org/10.3390/rs13244973>

Academic Editors: Yuji Murayama and Yuyu Zhou

Received: 28 August 2021

Accepted: 23 November 2021

Published: 8 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The United Nations forecasts that, by 2050, two-thirds of the world's population will live in cities and towns, well above today's level of 55 percent, with further increases in store [1]. Rigorous, scientifically justifiable assessments of the changes in urban population and in the extents and forms of the built-up environment will be needed to face the challenges of the coming century, namely, urban and rural economic development [2–6], poverty alleviation and inequality [7–9], disease transmission [10,11], health and health disparities [12,13], carbon emissions [14,15] and climate change [16–20]. These issues demand rigorous investigation across countries and over time—but scientific studies of urban

change have often been thwarted by the heterogeneities of national urban definitions [1,21]. Even within the “statistical urbanization” camp, in which urban definitions are couched solely in terms of population density, contiguity and size, there exist divergent views on the proper measurement of urban population and land area [22] and, arguably, an over-reliance on population density as the definitive indicator of urban locations. Recognizing that density measures alone are likely to be inadequate for many purposes, researchers are now beginning to explore the prospects for merging satellite-derived data with on-the-ground socioeconomic measures drawn from spatially disaggregated population censuses and surveys [20,22–26]. A necessary first step in this emerging program of research is to assess areas of agreement and disagreement between global remotely-sensed and country-specific urban concepts and measures.

Although many researchers have lamented the non-comparability of country-specific definitions, there has perhaps been an insufficient appreciation of how these definitions have evolved to reflect national priorities and economic development strategies. Taking a comparative approach, this research study examines three countries that vary significantly in their urban criteria—the United States, Mexico and India. We aim to achieve a better understanding of how the definitions adopted by these countries compare with classifications derived from remotely sensed data [27–31]. Such differences have seldom been systematically quantified and evaluated from the shared perspectives of the scientific communities concerned with local governance, demography, land cover and remote sensing. In particular, given its endorsement in March 2020 by the United Nations Statistical Commission, multiple independent evaluations of the *Degree of Urbanization* method—to be discussed in detail below—are now very much in order [32–34]. This statistical urbanization model has increasing appeal in policy circles, where it is being combined with other types of data to describe the characteristics of population along the urban–rural continuum [24,35,36].

We find that internationally comparable “statistical” portraits of urban areas generally tend to agree with census-based classifications at the extremes of the urban–rural continuum—occupied by large, dense cities on the one end and sparsely-populated rural areas on the other—but with substantial variation by country. There is especially strong agreement in the case of the United States, which has adopted statistical criteria not unlike those now being applied internationally. In all three study countries, the significant challenges lie in the middle of the continuum, where national and internationally comparable systems often disagree. Our analysis underscores the need for new findings on urban land and population to be communicated cautiously, with recognition that the differences between national statistical estimates and those based on remotely sensed data should not be dismissed lightly.

2. Materials and Methods

This study compares and combines population-centric and land-structure-centric views of urbanization. At the core of the analysis is the finest-grained spatial unit of analysis for which (c. 2010) there exists complete national coverage, namely, Indian settlements (villages, legally constituted towns, census towns and outgrowths) and within-town wards, Mexican basic geographic areas and United States census blocks. Although they apply different definitions in doing so, all three countries assign urban designations to these fine-scaled units, which then become the building-blocks of larger urban entities. The census-based data are complemented by the Global Human Settlement Layer (GHSL-BUILT), a remote sensing-derived product that estimates built-up land area [37]. The land-centric perspective on settlement of GHS-BUILT is then blended with spatial population estimates to produce the *Degree of Urbanization* (DoU) model [32], a well-conceptualized system for urban–rural classification that is potentially applicable across the globe.

2.1. The Three Study Countries

Among the three countries we study, the U.S. has perhaps most closely adhered to pure “statistical” criteria in its official urban definitions, whereby a small census unit of

land (a census block) is declared urban on the basis of the block's population density, the identification of contiguous blocks that also meet a density criterion, and the total population of all such blocks. The algorithmic details are well explained by [38], who also recounts the intellectual history leading to these present-day definitions. As is well known—Atlanta, Georgia, is an often-cited case—peri-urban development in the U.S. has been dominated, in many regions, by the suburbanization of middle- and higher-income groups, who possess the means to secure good-quality public services and who connect to core cities and employment by way of automobile-dominant transport networks [39]. A complicating recent development is that poverty also seems to be decentralizing into suburban and urban outlying areas (for a review, see [40]).

Mexico presents a more complicated case. Especially over the 2000–2012 period, national policies and funding strongly encouraged the development of large-scale, low-density housing in urban peripheries, motivated by the desire to improve access to housing for poor and lower-middle-income groups [41]. These well-meaning policies did not cover the costs of land acquisition, an oversight that led to the development of rural plots of land that were only later converted to urban plots on a fragmentary, case-by-case basis. Consequently, many of Mexico's new low-income communities in the peripheries have lacked essential public services and adequate infrastructure, including transport to more centrally located sites of employment. For much of the time-span of our research study—before the initiation of post-2013 policy reforms—the contrasting situations of the U.S. and Mexico are well summarized by [41] (p. 31):

The resulting sprawl of Mexican cities is different from suburbanization in the United States during the 1960s and 1970s, where middle-class households moved to suburbs for more space with better amenities and schools. Instead, urban growth in Mexico has been connected to the fissure between new, peri-urban developments and more central neighborhoods in terms of the provision of infrastructure and services (including health and education), connectivity, access to sources of employment and urban amenities.

Judged in terms of built-up percentages and population density alone, the patterns of urban sprawl in Mexico and the U.S. might appear to be quantitatively similar, but any such similarities are superficial; the inhabitants of urban peripheries in these two countries differ in their socioeconomic standing, access to public services and connections to employment.

India occupies a distinctive position in the set of study countries. According to the country's official criteria, its urban percentage in 2011 was only 31.2 percent, well under half the percentages of Mexico and the U.S. at the time. However, in the Indian context, the meaning of “urban” is jurisdictional: urban populations are defined in terms of the people living within the boundaries of legally urban local governments. (A partial exception is made for census towns and outgrowths, which are legally rural despite having urban-like features in terms of density, size and the extent of non-agricultural employment. A legally rural settlement is eligible to be designated as a census town if it has a population expected to be 5000 or above in the upcoming census, an expected density of 400 persons per km² and if 75% of the male main workforce is likely to be engaged in non-agricultural activities. Outgrowths are areas of high-density, arguably urban settlement that are spatially adjacent to statutory cities and towns and which would thus seem to be poised on the threshold of becoming legally urban. In the meantime, however, they too continue to be governed by rural local governments.) Before approving the transition of a rural settlement to legally urban status, Indian state-level authorities must weigh the fiscal costs—loss of access to dedicated, relatively plentiful rural development funds—against the prospects of securing commensurate urban-dedicated funds, which have proven difficult for smaller cities and towns to obtain [42,43].

Perhaps as a consequence, a substantial proportion of India's rural population lives in large, dense, legally rural villages that elsewhere might be accorded urban status. According to [44–46], in 2011 there were 155,732 Indian villages with at least 1000 population and a density of 400 persons per km² or more, accounting for almost 80 percent of the

country's total rural population. The high densities of these villages raise concerns about the adequacy of sanitation and other services that would normally be addressed through urban infrastructure programs. High rural population densities also present a challenge to density-centered "statistical" definitions of urban.

Some researchers contend that the tight linkage between India's jurisdictional criteria and its fiscal system causes the country's urban percentages to be seriously understated [47]. A more sympathetic view is that urban entities in India are regarded as being fundamental units of decentralized governance, a position that was formalized in India's constitutional amendments of 1993 [43]. Since then, the country has struggled to find an operationally sustainable model for the local end of the urban governance tier.

As this brief account suggests, in the three study countries much recent policy attention has been directed to urban peripheries, smaller urban centers, and larger rural settlements—all of which are situated in the middle ranges of the urban–rural continuum. In these spaces, remotely-sensed methods may well excel in producing rigorous estimates of built-up land and population densities in fine spatial detail. However, for the foreseeable future, such estimates will need to be supplemented with measures of socioeconomic composition, service provision and governance if the densities are to be properly interpreted. In the near term, accepting that remotely-sensed methods will not soon provide persuasive evidence of socioeconomic variation within small spatial units, such measures can at least identify terrain that is rapidly developing over the years between censuses and prompt on-the-ground investigation by governments into the adequacy of services and transport. In this way, the new methods can help alert local authorities to the fast-paced changes underway in their own and nearby jurisdictions.

2.2. Official Data Sources

2.2.1. Indian Settlements and Within-Town Wards

Settlement-specific summaries of the population and socioeconomic characteristics of all officially defined Indian settlements covered in its 2011 Census—rural villages, statutory (i.e., legally urban) towns and their wards and the quasi-urban census towns and outgrowths—have been placed in the public domain by the Registrar General of India and are readily accessible via the national census website (<https://censusindia.gov.in/2011-Common/CensusData2011.html>, accessed on 1 April 2021). Unfortunately, the boundaries of 2011 Indian settlements (and within-town wards) are not yet publicly available. We have relied on a comprehensive collection of settlement boundaries in vector format, originally assembled by the private firm ML Infomap Ltd. (New Delhi, India), whose license allows the data to be displayed in research products but not redistributed (for discussion, see [25]).

2.2.2. Mexican Basic Geographic Areas

Mexico's official definition of urban begins with the identification of distinct settlements termed *localities*, which are classified as *urban localities* if the settlement contains 2500 or more inhabitants (see the account of [41], pp. 106–107). Each urban locality is spatially subdivided into basic geostatistical areas (AGEBs), which are groups of blocks delimited by streets, avenues, sidewalks, or other easily identifiable construction, in which land is used mainly for occupational, industrial, service provision, or commercial purposes. They are further constrained in size by the requirement that a single census enumerator should be able to canvass an AGEb. The boundaries of AGEbS are typically drawn so as to extend just beyond the current built-up area. Boundary adjustments are then made after each census to keep pace with local land development. All other land area outside urban localities is defined as rural and is spatially divided into rural AGEbS, within which *rural localities* (villages) are situated. To locate urban AGEbS spatially, we have relied on detailed boundary files that the National Institute of Statistics, Geography and Informatics (INEGI) has placed in the public domain (<http://en.www.inegi.org.mx/temas/mg/#Downloads>, accessed on 1 April 2021).

2.2.3. United States Census Blocks

Not unlike Mexican AGEs, U.S. census blocks are delineated by both man-made and physical characteristics of the landscape, such as roads and rivers as well as legal and administrative boundaries in some instances. They can differ greatly in area and total population, ranging from zero to several hundred people in cities. As [38] explains, in the 2010 U.S. census an urbanized area (UA) was defined as a contiguous set of blocks each having a population density >1000 people per mi^2 and, when taken together, a total population in excess of 50,000. Urban clusters (UCs) were defined as a core set of contiguous census blocks with densities greater than 1000 people per mi^2 but a total population across such blocks of 2500–49,999 persons. Any blocks in close proximity (within 2.5 miles) to UAs and UCs were defined to be urban if their population density exceeded 500 people per mi^2 (a complex algorithm defines proximity; it allows short “hops” and “skips” to connect otherwise dis-contiguous units). Also defined as urban are some categories of land in industrial and commercial use—non-residential blocks mainly covered by impervious surfaces (pavement, parking lots, and airports) within 0.25 miles of populated urban blocks of UAs and UCs. The boundaries, populations and urban–rural status of all 2010 U.S. census blocks are accessible in the public domain (<https://www.census.gov/programs-surveys/geography/technical-documentation/complete-technical-documentation/tiger-geo-line.html>, accessed on 1 April 2021).

2.3. Data Sources Derived from Remote Sensing

2.3.1. Global Human Settlement Layer

Produced by the Joint Research Centre of the European Commission, GHSL-BUILT estimates built-up land on the basis of Landsat imagery gathered in a series of global snap-shots centered on 1975, 1990, 2000 and 2014. For this research study, we used the 2014 version, whose imagery was assembled some 4 years after the population censuses of Mexico and the U.S. and 3 years after the Indian census. The GHSL team applies machine-learning methods to identify the presence or absence of structures at a resolution of 30 m^2 . Individual structures are not themselves identified, only the proportion of the grid cell occupied by one or more structures. The model has been trained not to mistake roads for structures—it does not currently identify roads as such—and does not (yet) distinguish residential from non-residential structural coverage [30,37,48–50]. Apart from a 30 m^2 water mask, we used the 250 m^2 aggregated version of GHSL-BUILT, which provides the proportion of each grid-cell that is built-up. Recent validation efforts have reported acceptable levels of accuracy of GHSL-BUILT except in low-density rural regions [51–54].

2.3.2. Gridded Population of the World and Global Human Settlement Population Layer

Two intermediate products serve to link GHSL-BUILT to the classification system of the *Degree of Urbanization* model. The Gridded Population of the World (GPW) ([55], Version 4.11) is based exclusively on sub-national administrative unit boundaries and total unit populations. To produce the Global Human Settlement Population Grid (GHS-POP), a dasymetric refinement method is applied to reallocate GPW population counts for administrative units to the grid cells within the unit boundaries, according to the finer-resolution, cell-specific proportions built-up as estimated by GHS-BUILT. As [56] note, “The benefit of GHS-POP is that it restricts population to built-up areas and makes its density directly proportional to the density of built-up areas (Freire et al., 2015). However, ..., population may be allocated to ‘non-residential’ areas such as commercial, industrial and recreational areas”. The assumption of direct proportionality between population and built-up densities is possibly too strong, although, admittedly, no compelling alternative is yet in hand. Alternative methodological solutions are explored in [57,58]; see [59] for a review of such approaches. Another concern warranting attention is that, due to the limited detection accuracy of GHS-BUILT in thinly settled rural areas, it is possible that GHS-POP over-concentrates administrative-unit population in the more built-up areas of the unit, which is likely to produce overestimates of urban populations relative to rural population when these data

are incorporated into the DoU algorithm. Despite this limitation at the rural end of the continuum, recent studies have shown that in urban and urbanized areas, GHS-POP can provide accurate estimates of pixel-level population [52,53,56,60].

2.3.3. The Degree of Urbanization Model

The *Degree of Urbanization* (DoU) model further refines GHS-POP by assigning settlement types on the basis of the spatially located population sizes and densities of GHS-POP. Along with other inputs, these data enable the construction of a seven-fold classification of urban and rural settlements [61]. (An eighth category identifies inland open water.) Table 1 presents a simplified version of these seven classes; see [61] for discussion of additional contiguity criteria not spelled out in the table. Note, in particular, the range of settlement types in the band of population density from 300 to 1500 persons per square kilometer; within this band, some difficulties can be anticipated in any effort to cleanly separate *semi-dense urban clusters* and *suburban or peri-urban* settlements from *rural clusters*. Taken together, the seven classes describe the full range of the urban–rural continuum insofar as population density, contiguity and size are concerned.

Table 1. *Degree of Urbanization* (DoU) population density and size criteria. For further detail on contiguity criteria, not described here, see [61].

		Minimum Population of the Cluster			No Minimum
		>50,000	5000–50,000	500–5000	
Population density (km ²)	>1500	Urban centre ①	Dense urban cluster ②	Rural cluster ⑤	
	300–1500		Semi-dense urban cluster ③		Suburban or peri-urban ④
	50–300				Low density rural ⑥
	<50				Very low density rural ⑦

Additional grid cells qualify for inclusion in the *urban centres* of the DoU model if they are at least 50% built-up. As [61] (p. 18) explain, “This assumption is useful for accommodating the presence in the city of large areas with low resident inhabitants but strongly functionally linked with the city, as for example large productive or commercial areas (typical case of cities in United States of America)”. A built-up density threshold of 3% was applied to add grid cells to the other three urban categories (*dense urban cluster*, *semi-dense urban cluster* and *suburban or peri-urban*), with the rationale being given by [61] (p. 19): “Grid cells are included in the urban cluster domain only if some minimal (evidence) of physical built-up structure was recorded by an independent source (with) respect to census data. The purpose of this assumption is to increase the robustness of the GHSL SMOD response by forcing consistency between census-derived sources (population grids) and land cover/land use sources (built-up areas) mitigating the effect of misalignment, thematic bias, scale gaps or other data gaps that may be present in the data”. In short, although the principal density criterion employed in the DoU is population density, a role for built-up density is integrated as well.

The implementation of the DoU depends crucially on the spatial resolution of the GPW administrative unit boundaries containing the unit populations. In the DoU global data set, fine-resolution spatial data are used for the United States and Mexico (census blocks and AGEs), but only moderate-resolution subdistricts for India (<https://sedac.ciesin.columbia.edu/downloads/docs/gpw-v4/gpw-v4-documentation-rev11.pdf>, accessed on 1 April 2021). Using restricted-distribution, settlement-level data for India in Supplementary Materials, we illustrate how the resolution of the population data can affect DoU classifications.

2.3.4. Data-Processing

We have conducted all data-processing steps using the Python geoprocessing capabilities of ArcGIS 10.6.1. The comparisons of vector boundary data (corresponding to settlement and within-town ward boundaries in India, the boundaries of AGEBS in Mexico and those of census blocks in the U.S.) to raster data were based on the overlap of vector units with raster cell centroids and employed zonal statistics geoprocessing functions. Vector and raster data for India were projected to *Asia South Albers Equal Area Conic*; for Mexico, the system was *North American Albers Equal Area Conic*; and for the U.S., *USA Contiguous Albers Equal Area Conic*.

2.4. Comparing Official Urban–Rural Classes with GHS-BUILT and the DoU

In the final DoU model, the fundamental role of built-up density (that is, the density of structures) is somewhat obscured by the intervening boundary and population layers used in constructing the model. In addition, the dependence of DoU classifications on country-specific administrative boundaries and population counts may limit the application of the DoU method to immediate post-censal periods and to countries that place their boundaries and population counts into the public domain. Hence, there is some value to be gained in a direct assessment of GHS-BUILT settlement proportions in relation to official urban–rural designations, since the remote-sensing and land classification programs operate independently of census data-collection and, in the future, can be expected to place new estimates in the public domain on a more frequent basis.

To explore an approach based on built-up densities, we have specified a threshold of $\tau = 50$ percent built-up to identify potentially urban land (recall that thresholds of 50% and 3% built-up are applied in classifying grid cells in the DoU model, in addition to population densities). Appendix A provides an extensive discussion of threshold choice in what can be viewed as a simple diagnostic test, whereby built-up densities above and below the threshold in a given grid cell serve as an (imperfect) signal of the official urban–rural status of the cell. The entries of Table 2 provide the terms we use in what follows to describe the diagnostic test outcomes in relation to the official urban–rural classification (selected results based on alternative τ thresholds are provided in Supplementary Materials).

Table 2. Terms used to summarize comparisons between official urban–rural classifications and GHS-BUILT built-up percentages, using a built-up threshold of $\tau = 50$ percent (see Appendix A).

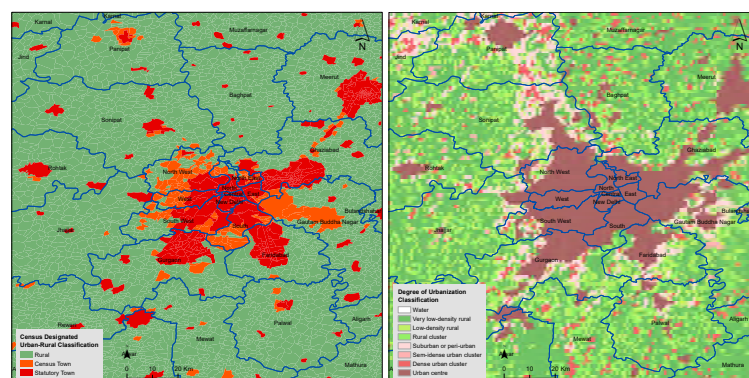
Official	GHS-BUILT Built-Up Percentage of 250 m ² Grid Cell	
	$\geq 50\%$	$< 50\%$
Census urban land	urban agreement	urban, not built-up
Census rural land	rural, but built-up	rural agreement

In extending this approach beyond land classification to address population densities and totals, we have adopted a conventional areal weighting approach [62,63] to distribute population uniformly within the boundaries of each individual census unit of land (settlements or within-city wards in India, AGEBS in Mexico and census blocks in the U.S.). The uniformity assumption may be acceptable within very small census units but has the potential to underestimate population in the highly built-up sections of any given unit and overestimate population in its less built-up sections. However, we do not expect significant errors in the average estimates of population densities, nor do we anticipate systematic biases overall.

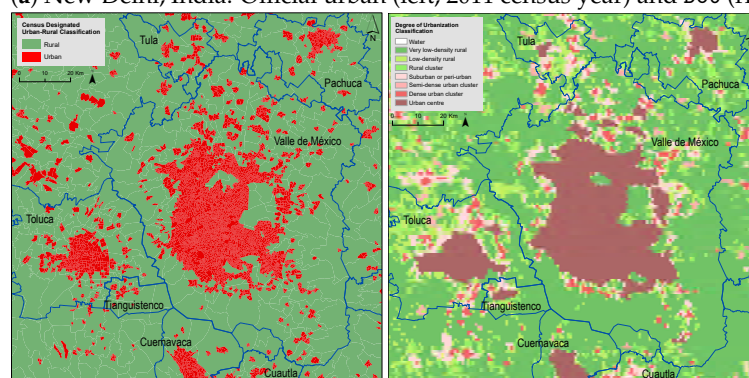
Official Urban–Rural Classifications and the DoU

Similar overlay methods were used to reallocate official population counts to grid cells and thus compare officially classified census units with the more refined seven-fold classification of the DoU. The maps of each of these input layers are shown in Figure 1 for one major city and its surrounding areas in each of the study countries. A visual scan of these examples suggests generally good agreement between the official urban–rural designations and the *Degree of Urbanization* (DoU) classification. However, on closer inspection, areas of

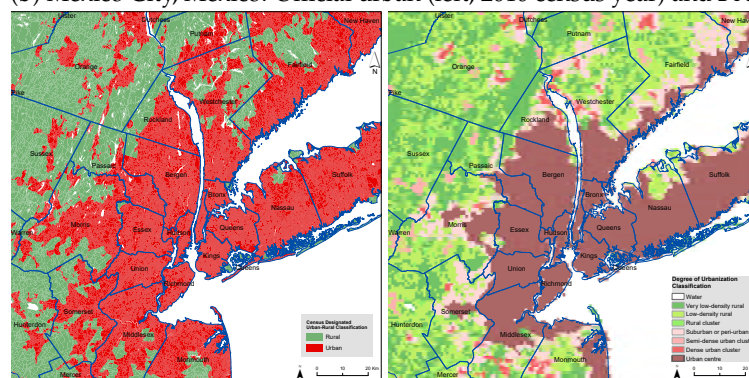
nuanced disagreement and ambiguity come into view. For example, in the northern part of Nassau County (to the east of New York City proper), a relatively small patch of land is officially rural, but, in the DoU classification, the rural extents are considerably larger and are intermixed with *suburban or peri-urban* areas. Significant discrepancies are also seen to the south of New York City in Monmouth county (New Jersey), most of which is officially urban yet DoU-classified as rural. The DoU classes (and their descriptive labels) could be regarded as helpful refinements of binary urban–rural official designations, or, when they are not in obvious agreement with such designations, could be taken as an invitation to engage in more detailed critical analysis.



(a) New Delhi, India: Official urban (left, 2011 census year) and DoU (right, 2015 estimate).



(b) Mexico City, Mexico: Official urban (left, 2010 census year) and DoU (right, 2015 estimate).



(c) New York City, United States: Official urban (left, 2010 census year) and DoU (right, 2015 estimate).

Figure 1. Official urban and *Degree of Urbanization* (DoU) gridded layers. Examples are from (a) New Delhi, India; (b) Mexico City, Mexico; and (c) New York City, United States. All images show a portion of the surrounding area. Administrative boundaries are indicated in blue. These are districts, in the Indian case; municipios, for Mexico; and counties, in the United States.

3. Results

3.1. Classifications of Land

Table 3 summarizes the findings on land area by urban–rural class. Note first that, by official urban criteria, India and the U.S. devote approximately equal percentages of habitable land to urban areas (3.4% and 3.6%, respectively); at 1.2%, Mexico’s allotment is much lower (for international comparisons, see [64]). Of all urban land in India (whether in statutory towns, or in the outgrowths and census towns that, while legally rural, are treated as urban in India’s official tabulations), about 73 percent is occupied by statutory towns, some 3 percent by outgrowths and the remaining 24 percent by census towns (figures not shown). The rural governments that have census towns in their jurisdictions do not have direct access to the urban development funds that could otherwise be used to improve infrastructure and sanitation, nor, in general, can they impose urban rates of property taxation to raise needed revenues.

Table 3. Total and percentage of habitable land area (in square kilometers), by official urban designation, GHSL–official comparisons and DoU category. Official figures from 2011 census year for India and from 2010 census years for Mexico and the United States. GHSL estimate for 2014. DoU estimate for 2015.

Status	India		Mexico		United States	
	Area	Percent	Area	Percent	Area	Percent
<i>Official Designation</i>						
Urban	109,578	3.39	22,942	1.17	274,962	3.58
Rural	3,124,420	96.61	1,942,272	98.83	7,398,680	96.42
<i>GHSL threshold versus Official Designation</i>						
Urban agreement	12,580	0.39	13,237	0.67	113,853	1.46
Urban, not built-up	97,203	3.01	9705	0.49	165,413	2.12
Rural, but built-up	5014	0.16	4872	0.25	13,618	0.17
Rural agreement	3,113,781	96.44	1,937,399	98.58	7,517,111	96.25
<i>Degree of Urbanization Category</i>						
Urban centre	75,094	2.31	12,223	0.63	91,009	1.18
Dense urban	51,522	1.58	4739	0.24	19,496	0.25
Semi-dense urban	4856	0.15	2538	0.13	16,373	0.21
Suburban or peri-urban	41,477	1.27	9010	0.46	64,035	0.83
Rural	234,587	7.21	19,264	0.99	31,426	0.41
Low-density rural	304,857	9.37	52,865	2.72	404,267	5.23
Very low-density rural	2,541,742	78.11	1,844,728	94.83	7,103,698	91.89

The *rural agreement* category—using the terminology of Table 2, estimated from a cross-comparison of official census land and GHSL built-up levels—is nearly identical to the figures for officially rural land. This close correspondence means that the sum of the *rural, but built-up*, *urban, not built-up* and *urban agreement* categories is also quite close to the officially urban land class. Of all officially urban land area, 58% in Mexico and 41% in the U.S. is found in the *urban agreement* category. However, in India, only 11% of officially urban land is classified as *urban agreement*; the remainder falls into the *urban, not built up* category. In other words, the great majority of officially urban land in India is estimated to be less than half built-up. If this seems surprising and scarcely credible, it may be that our perceptions of high densities in Indian cities are formed mainly by impressions of high *population* densities, rather than by the settlement proportions being described here.

For India, the DoU classes are distributed differently from what can be seen in Mexico and the United States. The most rural category of the DoU, *very low-density rural*, occupies a notably smaller share of Indian habitable land (78.1%) than the 94.8% of land in Mexico and 91.9% in the United States. (The DoU category labels refer mainly to *population* density, not [or at least not directly] to the settlement proportion. In addition, recall that GHSL–BUILT estimates of built-up land in sparsely-settled rural areas—on which the DoU system is based—may well be downwardly biased.) Hence, the remaining rural DoU classes in India

account for higher percentages of land than in the other two countries. Another difference worth noting is that, among the three most urban of the DoU land classes (*urban centres*, *dense urban* and *semi-dense urban*), the *urban centre* class takes a significantly larger shares of land in Mexico (63 percent of these three categories of urban land) and especially the U.S. (72 percent) than in India (only 57 percent).

3.1.1. Built-Up Density by Urban–Rural Category

Figure 2 confirms that built-up densities of officially urban land in India fall well below the densities of urban land in Mexico and the United States (panel (a)). As panel (b) shows, the difference is mainly attributable to the low densities of the *urban, not built-up* areas. There is close agreement among the three countries in the mean built-up percentages of the *urban agreement* and *rural agreement* groups and in the *rural, but built-up* categories; only in the *urban, not built-up* category does India appreciably diverge from Mexico and the United States in having lower built-up levels on average. Because we define “built-up” in terms of a $\tau = 50$ percent threshold, the averages for *urban agreement* and *rural, but built-up* must exceed 50 percent and the averages for *urban, not built-up* and *rural agreement* must fall short of that threshold. What is striking is the extent to which these categories depart from the threshold value.

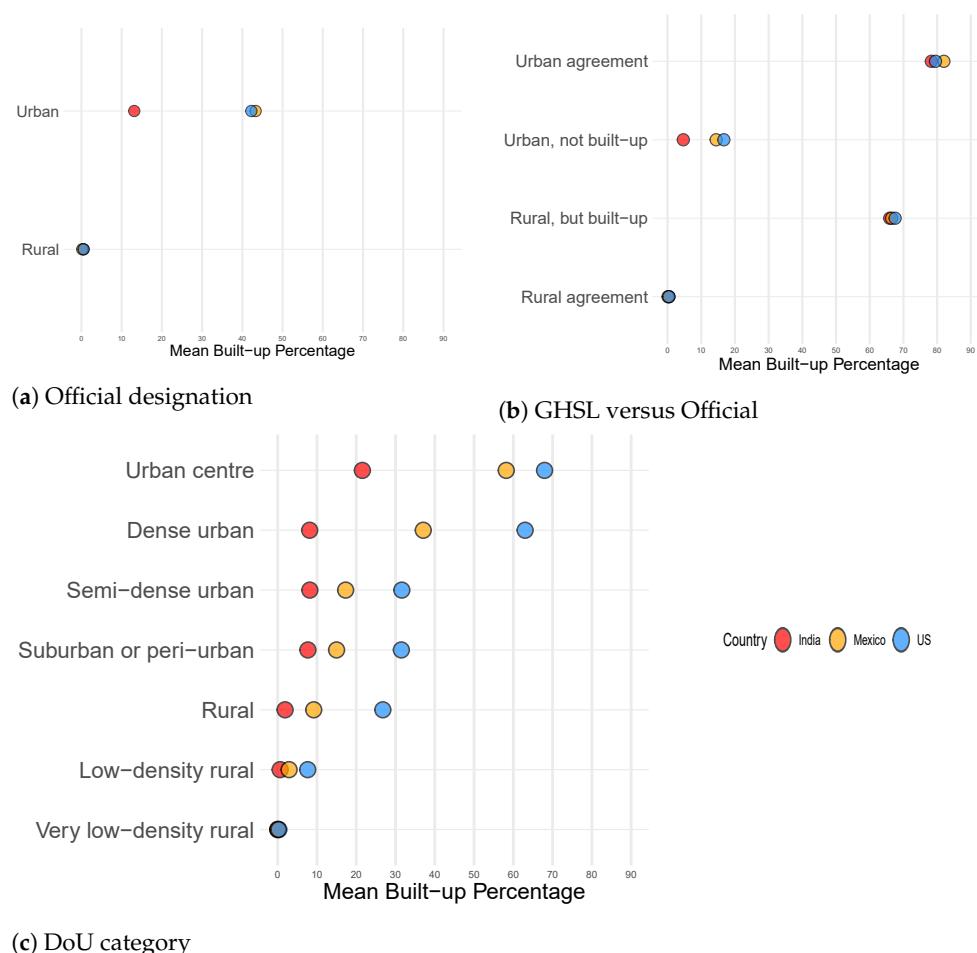


Figure 2. Mean built-up land density by urban–rural category and country. (a) Mean built-up densities by official urban–rural designation (2011 census year for India, 2010 for Mexico and the U.S.; 2014 GHSL estimate). (b) Comparison of means in the four-cell categorization of GHSL and the official designation. (c) Means by DoU category (2014 GHSL and 2015 DoU estimates).

Perhaps the most surprising results are those for the built-up densities by DoU class. Figure 2c shows a consistent ranking of the three countries across all DoU classes. The land

density percentages in the U.S. are the highest in each class, followed by those of Mexico and India, with a clear pattern of decreasing densities in Mexico and the U.S. as one moves down the urban–rural continuum. However, in India, *suburban or peri-urban*, *semi-dense urban* and *dense urban* areas exhibit essentially the same settlement proportions, which is noticeably lower than that found in the *urban centres* of the country and of course higher than the three rural DoU classes.

3.1.2. Official Urban Designations by DoU Class

Figure 3 displays one measure of agreement/disagreement between the official urban designations and those derived from the DoU model: the percentages of all grid cells in a given DoU class that are officially urban. At the most rural end of the DoU spectrum, the two classification schemes are in general agreement. However, even in the *rural clusters* class, differences begin to emerge—in the United States, over 55% of grid cells in this DoU class were officially urban—and, moving up the urban–rural continuum in India and Mexico from *suburban or peri-urban* to *urban centres*, a lack of consistent agreement with the DoU becomes evident.

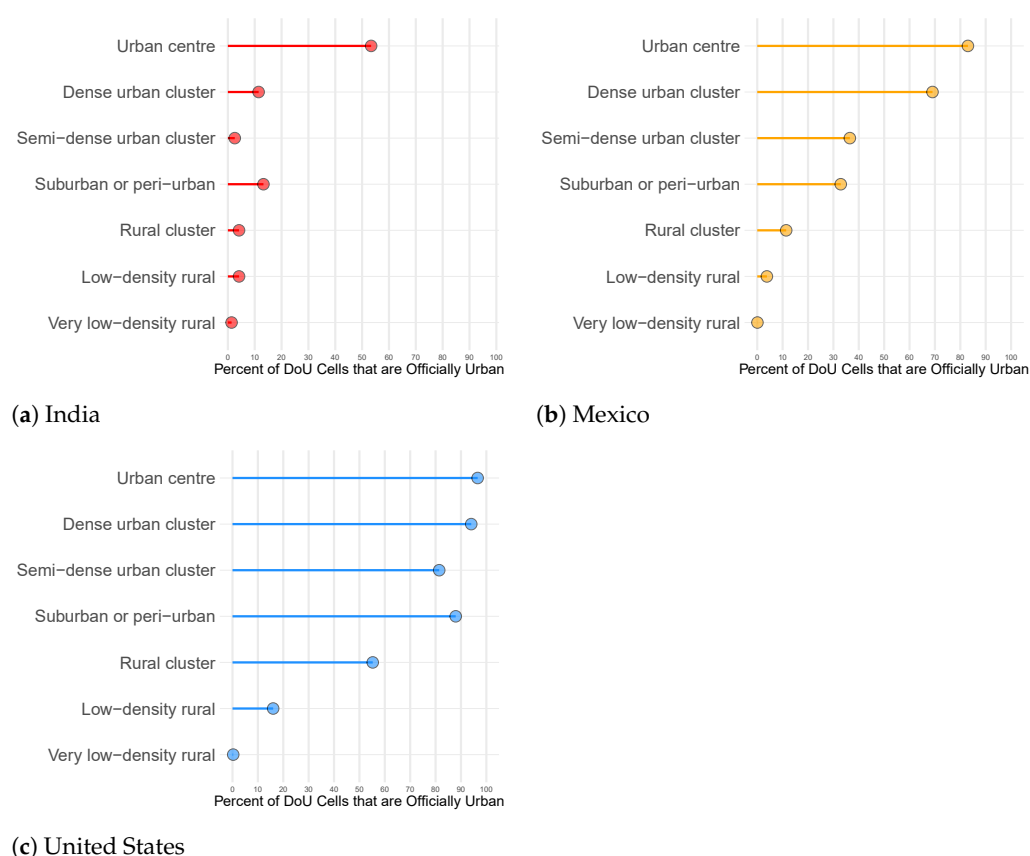
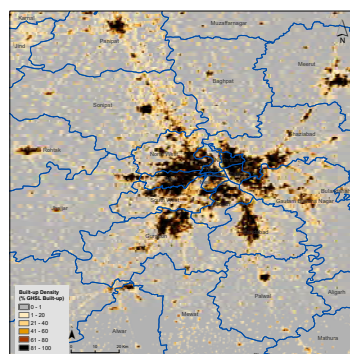


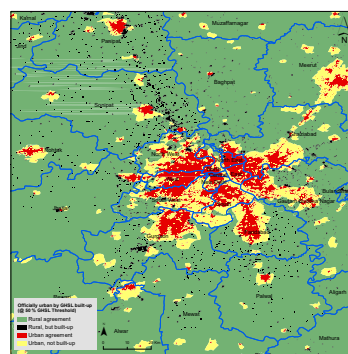
Figure 3. Official urban percentages by DoU class, calculated as the percentage of officially urban cells among all grid cells in each DoU class: (a) India, (b) Mexico and (c) United States. Official figures from 2011 census year for India, 2010 census years for Mexico and the United States. DoU estimate for 2015.

India presents an array of difficulties for the DoU conception of the urban–rural continuum, not only in defining urban settlements in jurisdictional terms as we have discussed, but also in the percentage of urban-designated land that is less built-up. Figure 4 illustrates the case of New Delhi and its surrounding areas. The areas depicted in yellow and red, when taken together, make up the land officially designated as urban. Evidently, a significant percentage of urban land—even in India’s capital—is less than 50 percent built-up. The composition of urban land near Mexico City—see Figure 5—also includes areas of less built-up land, especially in the smaller urban areas to the west of Mexico City; but as can

be seen, Mexico City itself is dominated by built-up land. As with Mexico City, Figure 6 shows that New York City proper (i.e., the five counties that jurisdictionally comprise the city) is nearly all built-up, but, as in New Delhi and to a lesser degree Mexico City, the surrounding areas are officially urban but not majority built-up. In these images (20 km is their common scale), the amount of officially urban land is much larger in the New York metro area—although at a range of built-up land densities—than in the other two regions.

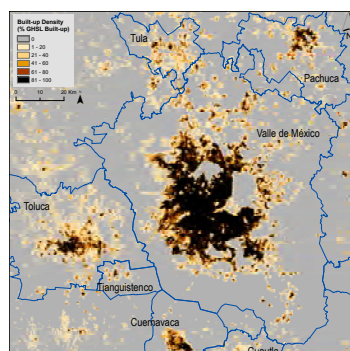


(a) Built-up land densities

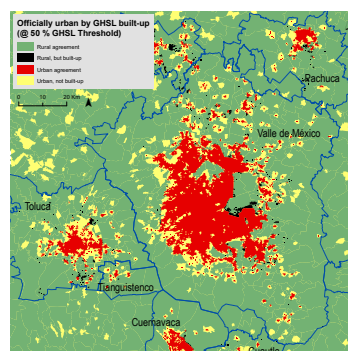


(b) Cross classification of land densities by official urban–rural status

Figure 4. Built-up land densities near New Delhi, India. Official classification for 2011 census year; 2014 GHSL estimate.

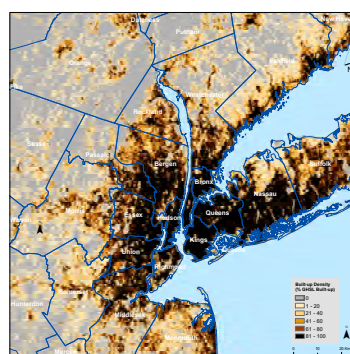


(a) Built-up land densities

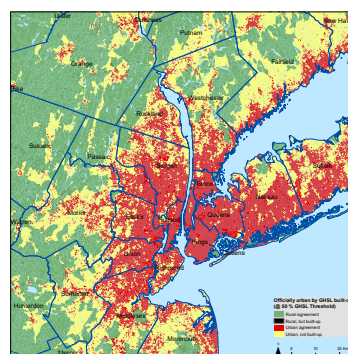


(b) Cross classification of land densities by official urban–rural status

Figure 5. Built-up land densities near Mexico City. Official classification for 2010 census year; 2014 GHSL estimate.



(a) Built-up land densities



(b) Cross classification of land densities by official urban–rural status

Figure 6. Built-up land densities near New York City. Official classification for 2010 census year; 2014 GHSL estimate.

3.2. Population Shares and Densities

We now turn attention to population and population density, illustrated in Figure 7. This figure shows the spatial distribution of population in New Delhi, Mexico City and New York City and their surrounding areas. For context, the United Nations estimated the 2020 population of the greater urban agglomerations of these cities to be 30.3, 21.8 and 18.8 million, respectively ([1], File 22). The relatively high density of population in India in areas outside the capital, many of which are officially rural (see Figure 1), is clearly evident.

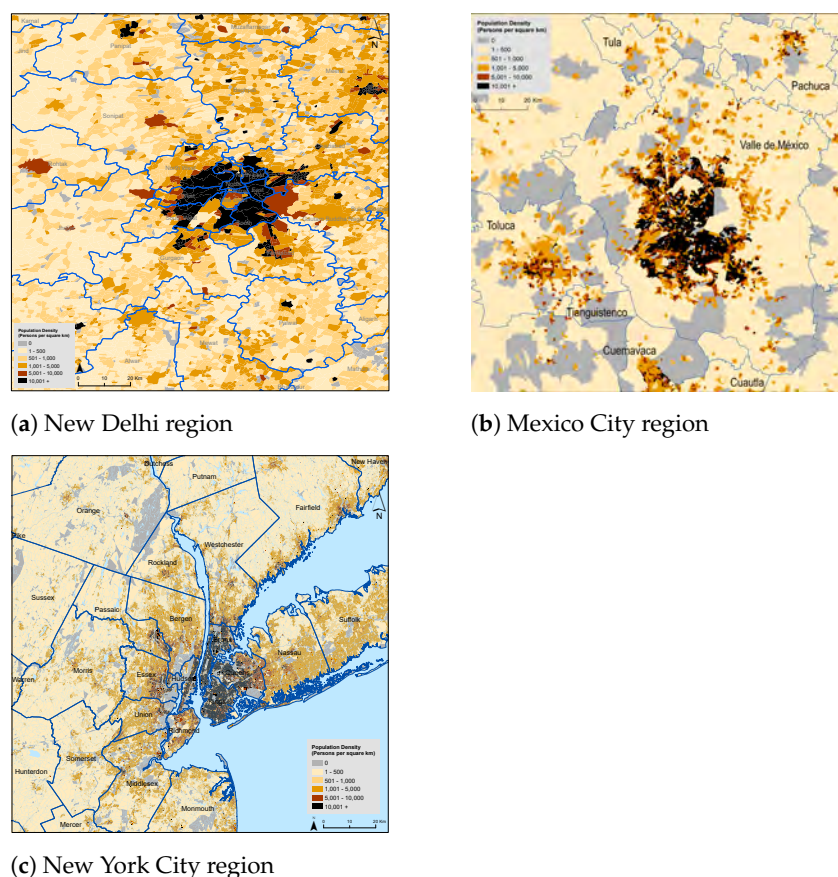


Figure 7. Population densities in selected regions of the study countries. Settlement boundaries and population from the 2011 census year for India (panel (a)) and the 2010 census year for both Mexico (panel (b)) and the United States (panel (c)).

Table 4 displays the total population size and percentage share for each officially designated urban–rural category, the official–GHSL cross-classification and the seven-fold DoU classes. The census-based classification indicate that 77.8% and 80.8% of the population is officially urban in Mexico and the U.S., respectively. As is often remarked-upon, India’s official urban share at the time of its 2011 census, based on legal jurisdictional criteria with additional consideration of census towns and outgrowths, was only 31.3 percent. The official–GHSL cross-classification (built-up threshold 50%), shows that 57–58% of the populations of Mexico and the U.S. lived in places of *urban agreement* in 2010—areas that were defined as urban by the census and in which built-up density was 50% or greater—whereas in India, only 10.7% of the total population lived in such areas. In all three countries, just over one-fifth of the population (20–24%) lives in areas that were officially urban yet less than 50% built-up. It should be added that, in India, these *urban, not built-up* areas are home to some 248 million people (20.5% of the population), a total that exceeds the number of urban residents of the United States. In the U.S., where 22.8% of people live in the *urban, not built-up* areas, this percentage has been the result of urban expansion (or “sprawl”) into areas occupied by relatively few people but which account for large shares of land [28].

Table 4. Total and percentage of population by official urban designation, GHSL–official comparisons and DoU category. Total population is expressed in thousands. Official figures from 2011 census year for India and 2010 census years for Mexico and the United States. GHSL estimate for 2014; DoU estimate for 2015.

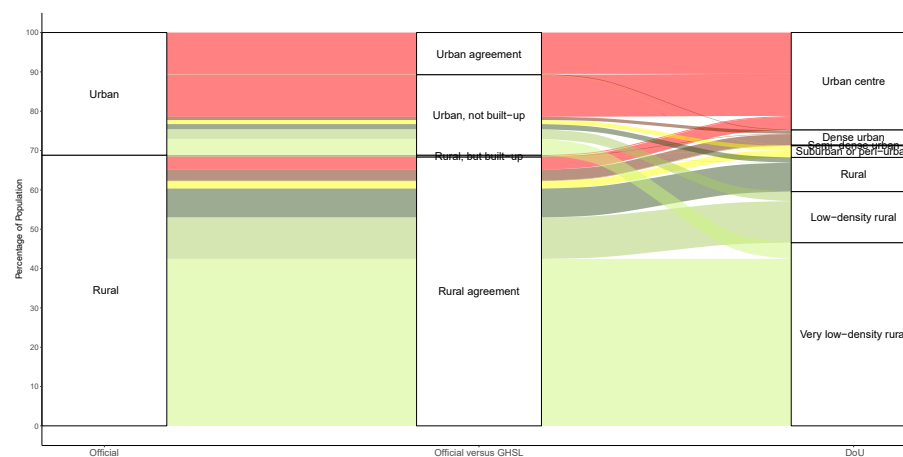
Status	India		Mexico		United States	
	Population	Percent	Population	Percent	Population	Percent
<i>Official Designation</i>						
Urban	377,108	31.28	85,568	77.82	247,516	80.77
Rural	833,746	68.72	20,685	22.18	59,141	19.23
<i>GHSL threshold versus Official Designation</i>						
Urban agreement	129,362	10.68	60,550	56.99	177,569	57.90
Urban, not built-up	247,747	20.46	25,018	23.55	69,948	22.81
Rural, but built-up	5521	0.46	96	0.09	2287	0.75
Rural agreement	828,225	68.40	20,589	19.38	56,854	18.54
<i>Degree of Urbanization Category</i>						
Urban centre	296,221	24.52	60,861	57.45	144,458	47.37
Dense urban	45,516	3.77	10,235	9.66	21,038	6.90
Semi-dense urban	1988	0.16	1679	1.59	9545	3.13
Suburban or peri-urban	38,501	3.19	5818	5.49	38,801	12.72
Rural	128,965	10.68	3785	3.57	12,905	4.23
Low-density rural	131,087	10.85	4529	4.28	40,953	13.43
Very low-density rural	565,639	46.83	19,026	17.96	37,252	12.22

The more detailed DoU classification place a greater share of India's population (24.5%) and fewer U.S. residents (47.4%) in *urban centres* than are found in areas of *urban agreement* (a detailed analysis of these differences will be presented in Figure 8 with accompanying discussion). The distribution of the population living in the middle-range urban categories of *dense urban*, *semi-dense urban* and *suburban or peri-urban* varies considerably across the three countries in ways that do not seem obviously attributable to differences in national urban percentages overall. With only 3.2% of India's population residing in areas classified as *suburban or peri-urban*, it seems likely that the country's large, dense villages must be distributed among this class and the *rural* and, possibly, even the *low-density rural* classes.

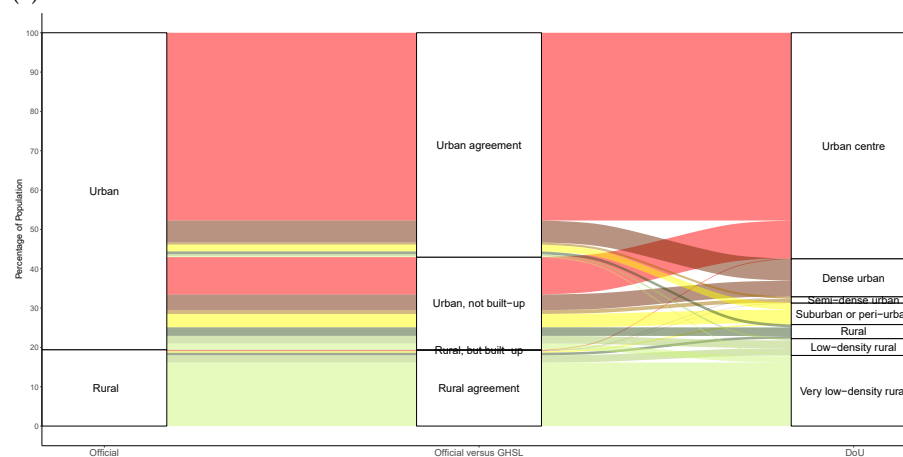
Table 5 presents the differences in population density by urban classification and country setting. We observe sizable differences across countries in the official urban census-based estimates; India and Mexico have very high urban population densities (>3300 people per km²) with much lower densities evident in the officially urban areas of the U.S. (only 886 people per km²). As for rural areas, India's are, respectively, 24- and 33-times denser than the rural areas of Mexico and the United States. In the census–GHSL cross-classification ($\tau = 50\%$ built up), we observe, in areas of *urban agreement*, very high levels of population density in India at 10,283 persons per km² and high levels in Mexico (5336 km²), but substantially lower levels in the U.S. (less than 1600 persons per km²). Interestingly, population densities in *urban agreement* areas in the U.S. are less than or similar to those found in areas classified as *urban, not built-up* in India (2549) and Mexico (1538), respectively. However, the ratios of population densities of these two classes (*urban agreement:urban, not built-up*) are similar in all three countries, ranging between 3.5 and 4.

Highlighting differences between countries, population densities of the *rural, but built-up* areas of India, Mexico and the U.S. are 1101, 168 and 57 persons per km², respectively. This suggests two (not mutually exclusive) possibilities: Even small, officially rural settlements may be very densely populated in India; or there may exist many legally rural settlements that might qualify by other criteria to be considered urban. Although the population share of the *rural, but built-up* areas is quite small, there is good reason to think that these built-up rural locations are likely to transition to census-urban over time; see Conclusions for discussion. Areas classified as *rural agreement* (that is, officially rural and below the $\tau = 50\%$ percent threshold) are much less dense in population but still exhibit notable variation across countries—266, 11 and 8 persons per km² in India, Mexico and

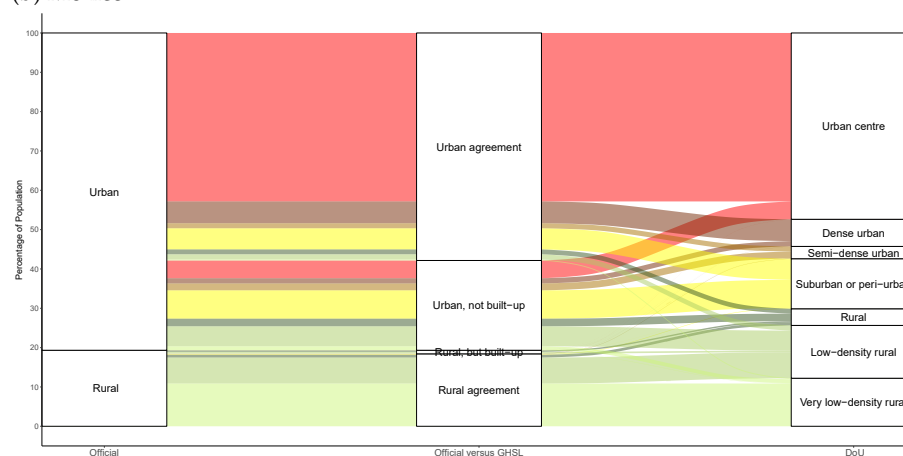
the U.S., respectively. These results indicate that, while there is general consistency among countries with regard to *ratios* of densities by urban class, the density *levels* of the study countries are quite different.



(a) India



(b) Mexico



(c) United States

Figure 8. Three-way cross-classifications of official urban–rural categories, GHSL–official cross-classification and DoU classes. Official figures from 2011 census year for India and 2010 census years for Mexico and the United States. GHSL estimate for 2014; DoU estimate for 2015, with classes shown in the order of Table 4.

Table 5. Mean population density (persons per square kilometer), by official urban designation, GHSL–official comparisons and DoU category. Official figures from 2011 census year for India and 2010 census years for Mexico and the United States. GHSL estimate for 2014; DoU estimate for 2015.

Status	India	Mexico	United States
<i>Official Designation</i>			
Urban	3388	3730	886
Rural	267	11	8
<i>GHSL threshold versus Official Designation</i>			
Urban agreement	10,283	5336	1560
Urban, not built-up	2549	1538	423
Rural, but built-up	1101	57	168
Rural agreement	266	11	8
<i>Degree of Urbanization Category</i>			
Urban centre	3945	4979	1587
Dense urban	883	2160	1079
Semi-dense urban	409	662	583
Suburban or peri-urban	928	646	606
Rural	549	196	411
Low-density rural	430	86	101
Very low-density rural	223	10	5

For the more refined DoU classification (Table 5), we see some curious variations along the urban–rural continuum. Surprisingly, India falls below Mexico in the population densities of *urban centres* and has lower densities than both Mexico and the U.S. in the *dense urban* and *semi-dense urban* categories. This is certainly not expected and may signal some difficulties in applying the DoU approach to urban India. (Recall that the DoU model takes no account of differences in official urban definitions across countries. However, the administrative units used for India are coarser than those employed in the DoU models of the U.S. and Mexico; see the Supplementary Materials.) Further investigation of DoU classification performance is clearly warranted in this case, not least because India accounts for about one-seventh of the world’s population.

3.3. A Combined Perspective

In this section, we assess how the DoU classes are composed, by tracing their connections to both the official–GHSL cross-classifications and to the binary urban–rural official designations. The “alluvial plots” of Figure 8 illustrate these three-way linkages, which are expressed in terms of population percentages. The colors of the figure correspond to the seven DoU classes that are shown on the far right of each plot; by following (from right to left) the color flow and the lines (and labels) that demarcate classes, one can trace the composition of these DoU classes (the Supplementary Materials provide the detailed percentages). The colors of the DoU classes are as follows: *urban centres* are shown in red; *dense urban* in brown; *semi-dense urban* in a lighter brown; *suburban or peri-urban* in yellow, marking the divide between urban and rural classes in the DoU scheme; *rural* population is indicated in dark green; *low-density rural* in a lighter green; and *very low-density rural* in the lightest of the green shades (the color scheme suggested in the DoU documentation).

As the figure indicates, *urban centres* and *very low-density rural* areas—the extremes of the urban–rural continuum in the DoU model—tend to agree with the official census-designated layers and census–GHSL cross-classifications. The percentage of the population in the *low-density rural* and *rural* DoU classes in India aligns fairly well with the *rural agreement* class, although a small portion of India’s DoU-rural, *low-density rural* and even *very low-density rural* classes originate in census-designated urban populations of the *urban, not built-up* type. The origins of the DoU *suburban or peri-urban* class in India are one part urban and the other part rural. Curiously, the *semi-dense urban* and *dense urban* DoU classes derive mainly from census-rural populations. However, the *urban centre* class largely originates in census-urban populations, although a significant percentage of this DoU class is traceable

back to census-rural populations. In short, there are a number of cross-over cases in India (whereby urban DoU classes stem from census-rural populations, or vice-versa) that call for further analysis. On the whole, such cross-overs are less evident in Mexico, although there certainly exist linkages worth investigating between DoU-rural and *low-density rural* populations having origins in census-urban populations. Much the same can be said for the United States, where both DoU-rural and *low-density rural* populations can also be traced back to census-urban origins.

Looking across the three countries, it is evident that the *urban, not built-up* class of officially urban populations living on less than 50% built-up land is highly mixed in terms of its representation in multiple urban and rural DoU classes. In setting priorities for further investigation, we would place these *urban, not built-up* areas at the top of the priority list. It is certainly possible that the DoU algorithm successfully refines and correctly labels this important segment of national populations—which, let us recall, accounts for over one-fifth of the total population in each of the study countries—but it is also possible that further tuning of the algorithm will be needed to prevent misclassification.

Likewise, the *urban agreement* cross-classified GHSL-official category is somewhat mixed, contributing to several urban DoU classes and making small contributions to two or more DoU rural classes. Note that, in India, the population of DoU *urban centres* is roughly twice that of the *urban agreement* class; in other words, a substantial share of the *urban centre* population lives on land that is less than half built-up. This can occur in the DoU algorithm for land areas that meet the population size and population density criteria; the 50% built-up test is only applied to grid cells that do not meet these population criteria. A similar if less pronounced pattern is also evident in Mexico.

We should again emphasize that apparent inconsistencies in classification do not in themselves cast doubt on the DoU method. Instead, they may reveal that urban forms are simply too varied to be forced into national, binary, urban–rural containers. Nor is it obvious that apparent mis-classifications in India necessarily contain any lessons for Mexico and the United States—what is considered urban in one country may not be regarded as urban in another. Another complicating factor needs to be mentioned: Rural-like areas can be found amid surrounding urban terrain, such as when natural amenities akin to parks lie within otherwise fully urban areas, a situation that may be expressed in what appears to be misclassification. Finally, as shown in the Supplementary Materials, all these results are dependent on the chosen thresholds of built-up percentages used in both the DoU and the census–GHSL cross-classification and are also dependent on the spatial resolution of the input population data.

4. Discussion

In exploring several ways to represent rural–urban classes from population-based, land-based and combined perspectives, we have conducted what is, in effect, one independent evaluation of the DoU classification system, which can be considered the first global data product that represents the full range of the urban–rural continuum. The DoU system is well conceived and, given its purpose, is appropriately free of country-specific urban–rural designations. From one point of view, it can be seen as a diagnostic tool that advances our understanding of official urban differences among countries, with the potential to produce insights that could eventually prompt revision of the official statistics and improve planning and economic development programs [65,66]. Before such benefits can materialize, however, users must first understand precisely how the DoU model departs from country-specific urban conceptualizations. Only then can such a global schema be taken up by countries with differing urban definitions or, indeed, with no urban–rural designations at all.

Our results reveal some similarities among country-specific and harmonized models at the extremes of the urban–rural continuum. In the middle ranges, however, there exist inconsistencies that need to be resolved. Areas that we term *urban, not built-up* have an especially high priority for further investigation. If country differences are most pronounced

in this heterogeneous span of the continuum—because, say, growth in population and land-convergence is most likely to occur here, or because these spaces present challenges to transportation or climate adaptation policies—then no global classification model is likely to capture all important local variation. Modern technology, data and methods enable the creation of valuable, scientifically comparable new global datasets—but such datasets must inevitably sacrifice some aspects of local knowledge. Researchers and policy-makers need to be aware of the DoU limitations and constraints to fully understand its fitness for use. The point applies more generally: For all such urban-proxy research programs, both data-producers and the growing community of users stand to benefit from sustained engagement and critical feed-back.

Our research study and findings are subject to several limitations that should be borne in mind. First, it goes almost without saying that the more accurate is the detection of built-up land, the more reliable is the integration of the land perspective in urban definitions. The GHSL-BUILT method is known to face detection and classification challenges in low-density rural areas; improving its performance is the focus of much current research [51,58]. Second, there are concerns about variation in the resolution and units of the population data. As an illustration for India in the Supplementary Materials shows, quite different patterns of DoU classification emerge when the population census data are used at a settlement-level resolution rather than in the coarser subdistrict resolution of the global DoU dataset. Third, as discussed in Appendix A, the appropriate selection of built-up thresholds—which affect not only the GHSL-official cross-classification, but also the composition of the urban classes of the DoU model—needs to be established more rigorously than it has been in the literature. A fourth difficulty has to do with the technique by which population data are allocated spatially; there is potential to further improve the accuracy of the population statistics relative to the dasymetric methods used here. On this point, it has been shown that at least in the global South, the dasymetric refinement of populations using multivariate approaches typically outperforms the simpler GHSL-POP method by integrating additional explanatory variables and implementing different allocation techniques [57,58]. Although the development of advanced methods for population allocation is an active field of research [59,67,68], this on-going research has yet to engage in spatially-explicit studies of place-based characteristics in areas marked by agreement and disagreement between alternative urban classifications. Therefore, it is not yet known whether disagreement is more likely to be found in the peripheries of large or small cities, or in more or less economically developed subnational regions, etc. Focused analyses would help to improve future urban modeling.

5. Conclusions

The production of harmonized, globally-comparable statistical measures of urbanization is a remarkable scientific achievement. Even in their first-generation versions, GHSL-BUILT and the *Degree of Urbanization* model exhibit the potential to revolutionize understanding of country-specific urban definitions; night-time lights and similar remotely-sensed measures might add even further value [32,69–75]. In particular, radar-based methods for measuring urban vertical expansion and building volume are likely to complement and extend our understanding of the multiple dimensions of urban growth [76,77]. When they are combined with population and socioeconomic data, such models can produce much-needed, finely-detailed portraits of urbanization along the full urban–rural span. There are excellent near-term prospects for even deeper integration of social and demographic data than the *Degree of Urbanization* model has achieved. In many countries, data to be released in the 2020 round of national population censuses will be more spatially detailed, with more easily accessible boundaries, than was the case at any point in the past. Valuable demographic resources are coming into the public domain; they must be enlisted in the harmonization effort.

While the globally harmonized approaches are being improved, it is vitally important not to overlook the country-specific systems already in place, however difficult they may be

to formally integrate. In countries such as India that have adopted jurisdictional criteria to define urban places, legal boundaries delineate the spaces within which local governments possess the authority to act, whether to improve environmental sustainability, safeguard health, or intervene in any number of areas that are high priorities in the Sustainable Development Goals. In the United States, which has fully embraced statistical conceptions of urbanization and expunged jurisdictional names and boundaries from its urban definitions, jurisdictions nevertheless play a central and enduring role in regional policies and governance (see, for example, [78] on core-city populations and their surrounding metropolitan areas). Jurisdictional spaces may never be detectable in remotely-sensed data, but they retain local and national significance. The harmonized measures derived from such data can, however, indicate where urban-like densities of population and structures spill across jurisdictional boundaries to present multiple local and higher-level governments with coordination challenges. As the balance of sustainable development effort shifts from problem description to action and intervention, both jurisdictional and harmonized measures will need to be kept firmly in view.

As our analysis has indicated, much population and land area is located in places that are not obviously either urban or rural. One study of spatial reclassification of urban and rural areas from 1990–2010 found that such areas (in the U.S. case, *rural but built-up* areas) are likely to be reclassified as urban in the next census [79]. We suspect that dynamics such as these will be found to characterize other countries and urban–rural classes. If the pattern seen in the U.S. proves to hold more generally, there may be a land-cover basis for devising spatial forecasts of urbanization for applications ranging from city planning to climate modeling.

In closing, we urge that rural and urban localities be studied not in isolation, but as members of systems, with recognition of and appreciation for their multiple inter-connections. A recent and otherwise sophisticated literature on the delineation of specific urban areas, focusing on the use of high-resolution satellite data in high-income, data-rich countries, has surprisingly restricted attention to the identification of individual cities rather than zooming out to survey the wider urban–rural context [80,81]. As the other papers in this Special Issue of *Remote Sensing* demonstrate, the bird’s-eye view provided by remote sensing can contribute not only new measures of urban–rural connections and change, but also new insights with the potential to shape a theory encompassing such disparate units [82–84]. In the lead-up to the adoption of the Sustainable Development Goals, the High-Level Panel [85] argued forcefully for a holistic view of urban–rural–environmental systems and warned of the risks of narrowly urban-specific or rural-specific perspectives:

The post-2015 agenda must be relevant for urban dwellers. Cities are where the battle for sustainable development will be won or lost. Yet the Panel also believes that it is critical to pay attention to rural areas, where three billion near-poor will still be living in 2030. The most pressing issue is not urban versus rural, but how to foster a local, geographic approach to the post-2015 agenda.

To advance toward this long-term goal, the necessary first steps are to achieve an accurate description of the constituent local units and to situate them along an urban–rural continuum—but these are only the first steps.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/rs13244973/s1>, Table S1: Total Area (km²) by Degree of Urbanization, using population data of varying spatial resolutions, India, Table S2: Total Population (000’s) by Degree of Urbanization, using population data of varying spatial resolutions, India, Table S3: Total Density (persons/km²) by Degree of Urbanization, using population data of varying spatial resolutions, India, Table S4: Total (km²) and percentage area by cross-classifications of official urban designations with GHSL, using alternative built-up, Table S5: Total (000s) and percentage population by cross-classifications of official urban designations with GHSL, using alternative built-up thresholds, Table S6: Population density by cross-classifications of official urban designations with GHSL, using alternative built-up thresholds, Table S7: Built-up density by cross-classifications of official urban designations with GHSL, using

alternative built-up thresholds, Table S8: Three-way cross-classifications of official urban–rural categories, GHSL–official classification and DoU classes, Mexico (official figures from 2010 census; GHSL estimate for 2014; DoU estimate for 2015), Table S9: Three-way cross-classifications of official urban–rural categories, GHSL–official classification and DoU classes, USA (official figures from 2010 census; GHSL estimate for 2014; DoU estimate for 2015), Table S10: Three-way cross-classifications of official urban–rural categories, GHSL–official classification and DoU classes, India. Upper panel reflects Global DoU based on subdistricts and lower panels reflect DoU using settlement-level administrative data (official figures from 2010 census; GHSL estimate for 2014; DoU estimate for 2015), Figure S1: Degree of urbanization, New Delhi and surrounding areas. Left panel shows DoU distribution using the global data product based on sub-district-level population data; Right panel shows DoU distribution produced with the settlement-level population data, Figure S2: Alluvial Plots India, 25% and 1% thresholds, Figure S3: Alluvial Plots Mexico, 25% and 1% thresholds, Figure S4: Alluvial Plots India, 25% and 1% thresholds, Figure S5: Alluvial Plots India using settlement-level population data, 50%, 25% and 1% thresholds, respectively. References [86,87] are cited in the Supplementary Materials.

Author Contributions: All authors contributed equally but in different ways to this work, as follows: conceptualization, D.B., S.L. and M.R.M.; methodology, D.B., S.L., H.E. and M.R.M.; programming, H.E. and M.R.M.; formal analysis, H.E., M.R.M., S.L. and D.B.; resources, D.B., S.L. and M.R.M.; data curation, H.E. and M.R.M.; writing—original draft preparation, D.B., S.L. and M.R.M.; writing—review and editing, M.R.M., D.B., S.L. and H.E.; visualization, H.E., M.R.M., D.B. and S.L.; supervision, D.B.; project administration, D.B.; funding acquisition, D.B., S.L. and M.R.M. All authors have read and agreed to the published version of the manuscript.

Funding: The work was funded, in large part, by the US National Science Foundation award #1416860 to the City University of New York, the Population Council, the National Center for Atmospheric Research (NCAR) and the University of Colorado at Boulder. Additional support was provided for author Balk at the City University of New York, via an Andrew Carnegie Fellowship (#G-F-16-53680) from the Carnegie Corporation of New York. Partial funding for author Leyk was provided through the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under award numbers R21 HD098717 01A1 and P2CHD066613. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Partial funding for this work was also provided through the Humans, Disasters and the Built Environment program of the National Science Foundation, award number 1924670. Author Montgomery received partial funding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under award number R03 HD096193-02 and also from internal funds supporting the Population Council’s Population, Environmental Risks and Climate Change (PERCC) research program. Funding for preparation of the India data received additional support from the European Commission Contract number 2018.CE.16.BAT.050.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study come from the publicly available sources cited throughout.

Acknowledgments: We thank Maria Brandao Sze for her research assistance.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. A Grid-Cell-Specific Diagnostic Test for Urban and Rural Land

As explained in the main text, the population densities and totals used in the seven-fold DoU model were constructed from the GHSL–BUILT estimates of built-up densities and GPW data on both administrative unit boundaries and total unit populations. The GHSL–BUILT estimates of the density of structures largely determine the spatial placement of GPW population into grid cells within each administrative unit and also determine the level of population density that is assigned to that unit’s built-up cells. To expose the role of

built-up density more clearly than is possible in the DoU, we devised a simple thresholding method for grid-cell classification that focuses on the cell's level of built-up density.

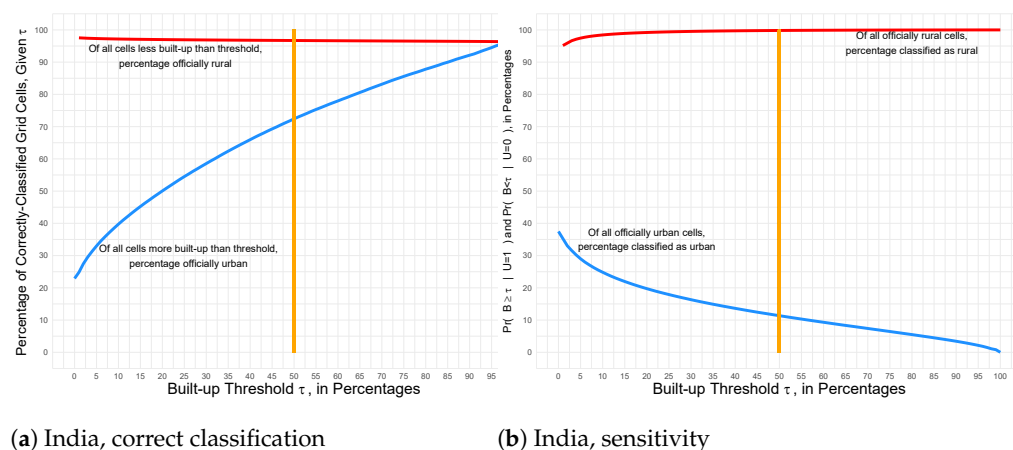
We view the method as a grid-cell-specific diagnostic test based on the percentage of the cell that is built-up. For any selected percentage threshold $\tau \in (0, 100)$, the built-up percentage B_i in grid cell i is compared with τ . If $B_i \geq \tau$, the cell is “diagnosed” as being officially urban; if $B_i < \tau$, it is diagnosed as officially rural. This test outcome is then checked against the official urban–rural status U_i of the cell (1 for urban, 0 for rural), using an overlay of urban vector data (polygons) on the GHSL built-up raster.

Table 2 in the main text provides verbal labels for the joint probability distribution of the built-up binary indicator and the official urban binary indicator (for example, the probability of *urban agreement* is $\Pr((B_i \geq \tau) \cap (U_i = 1))$). To go deeper in the exploration of τ values, it is helpful to recast the diagnostic problem in terms of conditional probabilities. A good test would have a high conditional probability $\Pr(U_i = 1|B_i \geq \tau)$ and also a high probability $\Pr(U_i = 0|B_i < \tau)$, both of which must depend on the level of the τ threshold that is used. These are the equivalents in the present case of what would be termed “true positives” and “true negatives” in a medical testing context. As the τ threshold increases, $\Pr(U_i = 1|B_i \geq \tau)$ is likely to increase, but $\Pr(U_i = 0|B_i < \tau)$ is apt to decrease. Thus, as τ is varied over a range of built-up densities, a generally negative relationship is traced out between the two probabilities.

Two additional probabilities also need consideration, which have to do with the “sensitivities” of the test, i.e., $\Pr(B_i \geq \tau|U_i = 1)$ (the likelihood that an officially urban grid cell is diagnosed as such by the threshold criterion) and $\Pr(B_i < \tau|U_i = 0)$ (the likelihood that an officially rural cell is diagnosed as rural via the threshold). This pair of probabilities also varies with the τ threshold, with urban sensitivity generally declining with τ (owing to the frequency of relatively low-density, below-the-threshold urban cells that become excluded from the $B_i \geq \tau$ diagnostic) and rural sensitivity increasing.

For these reasons, a test design (i.e., choice of τ) that exhibits high values for all four of these probabilities is simply unattainable. How, then, would a single best threshold τ be determined? This is not a statistical question as such, but rather a question about decision-making priorities. The optimal τ threshold clearly must depend on the positive weights attached by a decision-maker to each of the four relevant probabilities. There is no particular reason to think that decision-makers in different countries, experiencing different urban–rural contexts and paces of change, would necessarily adopt the same weights and arrive at the same choice for the τ threshold.

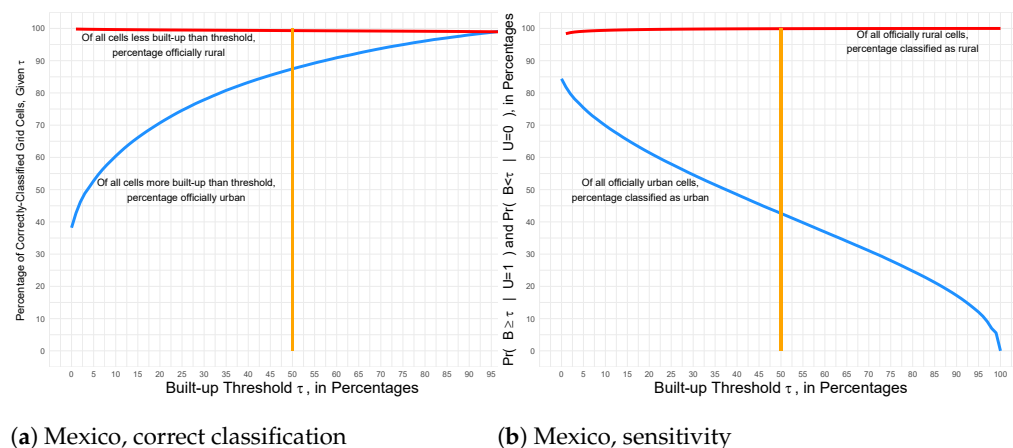
To estimate the four probabilities, we overlay the vector boundaries of officially urban land units on the GHSL-BUILT rasters of built-up density (here, as elsewhere in our analysis, water masks are used to restrict the comparison to habitable units of land). As is readily apparent in Figures A1 and A2, the choice of the threshold τ has little perceptible influence on the estimated $\Pr(U_i = 0|B_i < \tau)$, the probability of correct rural classification. This is due to two dominating facts about the distribution of land, namely, the percentages of rural land in the national totals for India and Mexico were extremely high—accounting for well above 90 percent of all habitable land—and the built-up percentages of rural land were very low, on average. Urban land misclassified as rural via the $B_i < \tau$ criterion certainly exists—see the main text for discussion of low built-up densities in India's officially urban settlements—and the total amount of such misclassified land does increase with τ , as would be expected; however, of all grid cells with $B_i < \tau$ the overwhelming majority remain officially rural even at high values for τ . A similar logic explains the lack of a τ effect on test sensitivity, $\Pr(B_i < \tau|U_i = 0)$. A large share of rural land has built-up densities below 1 percent, so that raising the threshold above that level has little impact on the percentage of all officially rural land diagnosed as rural by the $B_i < \tau$ criterion. In short, this kind of threshold test proves to be uninformative where rural land is concerned.



(a) India, correct classification

(b) India, sensitivity

Figure A1. *India.* Effects of varying τ percentage thresholds in the interval (1,99) on the likelihood of correct urban and rural land classifications and test sensitivity. By “correct”, we mean classifications that are in agreement with the official country-specific urban and rural designations. (a) India correct classification percentages by τ threshold. (b) India test sensitivity percentages by τ threshold.



(a) Mexico, correct classification

(b) Mexico, sensitivity

Figure A2. *Mexico.* Effects of varying τ percentage thresholds in the interval (1,99) on the likelihood of correct urban and rural land classifications and test sensitivity. By “correct”, we mean classifications that are in agreement with the official country-specific urban and rural designations. (a) Mexico correct classification percentages by τ threshold (b) Mexico test sensitivity percentages by τ threshold.

As these figures also show, however, the implications of the τ threshold for the correctness and sensitivity of urban land classification are far greater. A tradeoff between “true positives” correctness and sensitivity is clearly apparent; higher τ thresholds are associated with higher probabilities of correct classification given density (panel (a) of both figures) but produce lower test sensitivities (panel (b)), because raising τ has the effect of excluding relatively low-density urban cells. This presents would-be decision-makers with a dilemma: How should the two urban probabilities be balanced against each other in selecting one τ threshold?

Much as with the simple single-threshold diagnostic tool we here describe, issues of correct classification and sensitivity also arise in more heavily parameterized models such as the *Degree of Urbanization*, whose seven categories are based on multiple thresholds for both population density and size, as well as further parameters defining contiguity and built-up density. In this brief treatment, we cannot do justice to the complexities entailed in an analysis of threshold choice in such richer models. These issues must be left to future research.

References

1. United Nations, Department of Economic and Social Affairs, Population Division. *World Urbanization Prospects: The 2018 Revision*; Technical Report; ST/ESA/SER.A/420; United Nations: New York, NY, USA, 2019.
2. Rakodi, C. Economic development, urbanization, and poverty. In *Urban Livelihoods: A People-Centred Approach to Reducing Poverty*; Rakodi, C., Lloyd-Jones, T., Eds.; Earthscan: London, UK, 2014; Chapter 2.
3. Pandey, B.; Seto, K.C. Urbanization and agricultural land loss in India: Comparing satellite estimates with census data. *J. Environ. Manag.* **2015**, *148*, 53–66. [\[PubMed\]](#)
4. Bren d'Amour, C.; Reitsma, F.; Baiocchi, G.; Barthel, S.; Güneralp, B.; Erb, K.H.; Haberl, H.; Creutzig, F.; Seto, K.C. Future urban land expansion and implications for global croplands. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 8939–8944. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Bakker, V.; Verburg, P.H.; van Vliet, J. Trade-offs between prosperity and urban land per capita in major world cities. *Geogr. Sustain.* **2021**, *2*, 134–138. [\[CrossRef\]](#)
6. Forget, Y.; Shimoni, M.; Gilbert, M.; Linard, C. Mapping 20 Years of Urban Expansion in 45 Urban Areas of Sub-Saharan Africa. *Remote Sens.* **2021**, *13*, 525. [\[CrossRef\]](#)
7. Mitlin, D.; Satterthwaite, D. *Urban Poverty in the Global South*; Routledge: London, UK, 2013.
8. Fan, P.; Ouyang, Z.; Nguyen, D.D.; Nguyen, T.T.H.; Park, H.; Chen, J. Urbanization, economic development, environmental and social changes in transitional economies: Vietnam after Doimoi. *Landsc. Urban Plan.* **2019**, *187*, 145–155. [\[CrossRef\]](#)
9. Chen, M.; Sui, Y.; Liu, W.; Liu, H.; Huang, Y. Urbanization patterns and poverty reduction: A new perspective to explore the countries along the Belt and Road. *Habitat. Int.* **2019**, *84*, 1–14. [\[CrossRef\]](#)
10. Rajashekar, A.; Bower, J. *Densification without Contagion? Overcrowding and Pandemic Risk Hotspots in Rwanda*; Technical Report C19-20082-RWA-1; International Growth Center: London, UK, 2020.
11. Tian, H.; Hu, S.; Cazelles, B.; Chowell, G.; Gao, L.; Laine, M.; Li, Y.; Yang, H.; Li, Y.; Yang, Q.; et al. Urbanization prolongs hantavirus epidemics in cities. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 4707–4712. [\[CrossRef\]](#)
12. Montgomery, M.R. Urban Health in Low- and Middle-Income Countries. In *Oxford Textbook of Public Health*, 5th ed.; Detels, R., Beaglehole, R., Lansang, M.A., Gulliford, M., Eds.; Oxford University Press: Oxford, UK, 2009; Chapter 10.7, pp. 1376–1394.
13. Pinchoff, J.; Mills, C.W.; Balk, D. Urbanization and health: The effects of the built environment on chronic disease risk factors among women in Tanzania. *PLoS ONE* **2020**, *15*, e0241810. [\[CrossRef\]](#)
14. Moran, D.; Kanemoto, K.; Jiborn, M.; Wood, R.; Többen, J.; Seto, K.C. Carbon footprints of 13,000 cities. *Environ. Res. Lett.* **2018**, *13*, 064041. [\[CrossRef\]](#)
15. Creutzig, F.; Baiocchi, G.; Bierkandt, R.; Pichler, P.P.; Seto, K.C. Global typology of urban energy use and potentials for an urbanization mitigation wedge. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 6283–6288. [\[CrossRef\]](#)
16. Stewart, I.D.; Oke, T.R. Local Climate Zones for Urban Temperature Studies. *Bull. Am. Meteorol. Soc.* **2012**, *93*, 1879–1900. [\[CrossRef\]](#)
17. Revi, A.; Satterthwaite, D.E.; Aragón-Durand, F.; Corfee-Morlot, J.; Kiunsi, R.B.; Pelling, M.; Roberts, D.C.; Solecki, W. Urban Areas. In *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Field, C., Barros, V., Dokken, D., Mach, K., Mastrandrea, M., Bilir, T., Chatterjee, M., Ebi, K., Estrada, Y., Genova, R., et al., Eds.; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2014; pp. 535–612.
18. Solecki, W.; Seto, K.C.; Balk, D.; Bigio, A.; Boone, C.G.; Creutzig, F.; Fragkias, M.; Lwasa, S.; Marcotullio, P.; Romero-Lankao, P.; et al. A conceptual framework for an urban areas typology to integrate climate change mitigation and adaptation. *Urban Clim.* **2015**, *14*, 116–137. [\[CrossRef\]](#)
19. Demuzere, M.; Bechtel, B.; Middel, A.; Mills, G. Mapping Europe into local climate zones. *PLoS ONE* **2019**, *14*, e0214474. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Gao, J.; O'Neill, B.C. Mapping global urban land for the 21st century with data-driven simulations and Shared Socioeconomic Pathways. *Nat. Commun.* **2020**, *11*, 2302. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Buettner, T. Urban Estimates and Projections at the United Nations: The Strengths, Weaknesses, and Underpinnings of the World Urbanization Prospects. *Spat. Demogr.* **2015**, *3*, 91–108. [\[CrossRef\]](#)
22. Uchiyama, Y.; Mori, K. Methods for specifying spatial boundaries of cities in the world: The impacts of delineation methods on city sustainability indices. *Sci. Total Environ.* **2017**, *592*, 345–356. [\[CrossRef\]](#)
23. Ehrlich, D.; Freire, S.; Melchiorri, M.; Kemper, T. Open and Consistent Geospatial Data on Population Density, Built-Up and Settlements to Analyse Human Presence, Societal Impact and Sustainability: A Review of GHSL Applications. *Sustainability* **2021**, *13*, 7851. [\[CrossRef\]](#)
24. Henderson, J.V.; Liu, V.; Peng, C.; Storeygard, A. *Demographic and Health Outcomes by Degree of Urbanisation: Perspectives from a New Classification of Urban Areas*; Paper Prepared for European Commission, Directorate-General for Regional and Urban Policy; London School of Economics and Public Policy: London, UK, 2019.
25. Balk, D.; Montgomery, M.R.; Engin, H.; Lin, N.; Major, E.; Jones, B. Urbanization in India: Population and Urban Classification Grids for 2011. *Data* **2019**, *4*, 35. [\[CrossRef\]](#)
26. Leyk, S.; Uhl, J.H.; Connor, D.S.; Braswell, A.E.; Mietkiewicz, N.; Balch, J.K.; Gutmann, M. Two centuries of settlement and urban development in the United States. *Sci. Adv.* **2020**, *6*, aba2937. [\[CrossRef\]](#)

27. Balk, D.; Leyk, S.; Jones, B.; Montgomery, M.R.; Clark, A. Understanding Urbanization: A Study of Census and Satellite-derived Urban Classes in the United States, 1990–2010. *PLoS ONE* **2018**, *13*, e0208487. [[CrossRef](#)]
28. Leyk, S.; Balk, D.; Jones, B.; Montgomery, M.R.; Engin, H. The heterogeneity and change in the urban structure of metropolitan areas in the United States, 1990–2010. *Sci. Data* **2019**, *6*, 321. [[CrossRef](#)]
29. Dijkstra, L.; Poelman, H. *A Harmonised Definition of Cities and Rural Areas: The New Degree of Urbanisation*; Regional Working Paper, Directorate-General for Regional and Urban Policy; European Commission: Brussels, Belgium, 2014.
30. Freire, S.; Kemper, T.; Pesaresi, M.; Florczyk, A.J.; Syrris, V. *Combining GHSL and GPW to Improve Global Population Mapping*; Conference Paper; EC JRC Global Security and Crisis Management Unit, Joint Research Commission: Ispra, Italy, 2015.
31. Maffenini, L.; Schiavina, M.; Melchiorri, M.; Pesaresi, M.; Kemper, T. *GHSL-DUG User Guide*; Technical Report; Joint Research Commission, Publications Office of the European Union: Luxembourg, 2020. [[CrossRef](#)]
32. Dijkstra, L.; Florczyk, A.J.; Freire, S.; Kemper, T.; Melchiorri, M.; Pesaresi, M.; Schiavina, M. Applying the Degree of Urbanisation to the globe: A new harmonised definition reveals a different picture of global urbanisation. *J. Urban Econ.* **2020**, *125*, 103312. [[CrossRef](#)]
33. United Nations Statistical Commission. *Report on the Fifty-First Session (3–6 March 2020)*; Supplement No. 4, E/2020/24-E/CN.3/2020/37; Economic and Social Council Official Records; United Nations Statistical Commission: New York, NY, USA, 2020.
34. Statham, T.; Fox, S.; Wolf, L.J. Identifying urban areas: A new approach and comparison of national urban metrics with gridded population data. *Comput. Environ. Urban Syst.* **2021**, [[CrossRef](#)]
35. OECD; European Commission. *Cities in the World*; OECD and European Commission: Brussels, Belgium, 2020; p. 160. [[CrossRef](#)]
36. Coalition for Urban Transitions. *Climate Emergency/ Urban Opportunity*; Coalition for Urban Transitions c/o World Resources Institute: Washington, DC, USA, 2019.
37. Corbane, C.; Pesaresi, M.; Kemper, T.; Politis, P.; Florczyk, A.J.; Syrris, V.; Melchiorri, M.; Sabo, F.; Soille, P. Automated Global Delineation of Human Settlements from 40 Years of Landsat Satellite Data Archives. *Big Earth Data* **2019**, *3*, 140–169. [[CrossRef](#)]
38. Ratcliffe, M. A Century of Delineating a Changing Landscape: The Census Bureau's Urban and Rural Classification, 1910 to 2010. Presented at the Annual Meeting of the Social Science History Association, Baltimore, MD, USA, 14 November 2015. [[CrossRef](#)]
39. Redding, S.J. *Suburbanization in the United States 1970–2010*; Working Paper 28841; National Bureau of Economic Research (NBER): Cambridge, MA, USA, 2021. [[CrossRef](#)]
40. Allard, S.W.; Paisner, S.C. The Rise of Suburban Poverty. In *Oxford Handbooks Online*; Oxford University Press: Oxford, UK, 2016. [[CrossRef](#)]
41. Kim, Y.; Zangerling, B. (Eds.) *Mexico Urbanization Review: Managing Spatial Growth for Productive and Livable Cities in Mexico*; Directions in Development; World Bank: Washington, DC, USA, 2016. [[CrossRef](#)]
42. Khan, S. The Other Jawaharlal Nehru National Urban Renewal Mission: What Does It Mean for Small Town India? In *Subaltern Urbanisation In India: An Introduction to the Dynamics of Ordinary Towns*; Denis, E., Zerah, M.H., Eds.; Springer: New Delhi, India, 2017; Chapter 13, pp. 337–370. [[CrossRef](#)]
43. Mathur, O.P.; Naqvi, A.H.; Laroia, A.; Sayukta, V.S.; Verma, H. *State of the Cities: India*; Institute of Social Sciences: New Delhi, India, 2021.
44. Dasgupta, S.; Roy, S.N.; Bhol, A.; Raj, D. *Towards a New Research and Policy Paradigm: An Analysis of the Sanitation Situation in Large Dense Villages*; CPR Research Report; Centre for Policy Research: New Delhi, India, 2017. [[CrossRef](#)]
45. Bhol, A.; Dasguta, S.; Mukherjee, A.; Jain, A. *Sanitation in Large and Dense Villages of India: The Last Mile and Beyond*; Technical Report, CPR White Paper; Centre for Policy Research: New Delhi, India, 2019.
46. Denis, E.; Zerah, M.H. (Eds.) *Subaltern Urbanisation in India: An Introduction to the Dynamics of Ordinary Towns*; Exploring Urban Change in South Asia; Springer: New Delhi, India, 2017.
47. Onda, K.; Sinha, P.; Gaughan, A.E.; Stevens, F.R.; Kaza, N. Missing millions: Undercounting urbanization in India. *Popul. Environ.* **2019**, *41*, 126–150. [[CrossRef](#)]
48. Pesaresi, M.; Huadong, G.; Blaes, X.; Ehrlich, D.; Ferri, S.; Gueguen, L.; Halkia, M.; Kauffmann, M.; Kemper, T.; Lu, L.; et al. A Global Human Settlement Layer from Optical HR/VHR RS Data: Concept and First Results. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 2102–2131. [[CrossRef](#)]
49. Pesaresi, M.; Syrris, V.; Julea, A. A New Method for Earth Observation Data Analytics based on Symbolic Machine Learning. *Remote Sens.* **2016**, *8*, 399. [[CrossRef](#)]
50. Corbane, C.; Florczyk, A.; Pesaresi, M.; Politis, P.; Syrris, V. *GHSL-BUILT R2018A—GHSL Built-Up Grid, Derived from Landsat, Multitemporal (1975–1990–2000–2014)*; European Commission, Joint Research Centre (JRC): Ispra, Italy, 2018. [[CrossRef](#)]
51. Leyk, S.; Uhl, J.H.; Balk, D.; Jones, B. Assessing the accuracy of multi-temporal built-up land layers across rural-urban trajectories in the United States. *Remote Sens. Environ.* **2018**, *204*, 898–917. [[CrossRef](#)]
52. Liu, C.; Huang, X.; Zhu, Z.; Chen, H.; Tang, X.; Gong, J. Automatic extraction of built-up area from ZY3 multi-view satellite imagery: Analysis of 45 global cities. *Remote Sens. Environ.* **2019**, *226*, 51–73. [[CrossRef](#)]
53. Milesi, C.; Churkina, G. Measuring and Monitoring Urban Impacts on Climate Change from Space. *Remote Sens.* **2020**, *12*, 3494. [[CrossRef](#)]
54. Uhl, J.H.; Connor, D.S.; Leyk, S.; Braswell, A.E. A century of decoupling size and structure of urban spaces in the United States. *Communications Earth & Environment* **2021**, *2*, 20. [[CrossRef](#)]

55. CIESIN. *Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11*; Center for International Earth Science Information Network (CIESIN), Columbia University: Palisades, NY, USA, 2018.
56. Archila Bustos, M.F.; Hall, O.; Niedomysl, T.; Ernstson, U. A pixel level evaluation of five multitemporal global gridded population datasets: A case study in Sweden, 1990–2015. *Popul. Environ.* **2020**, *42*, 255–277. [[CrossRef](#)]
57. Reed, F.J.; Gaughan, A.E.; Stevens, F.R.; Yetman, G.; Sorichetta, A.; Tatem, A.J. Gridded Population Maps Informed by Different Built Settlement Products. *Data* **2018**, *3*, 33. [[CrossRef](#)] [[PubMed](#)]
58. Stevens, F.R.; Gaughan, A.E.; Nieves, J.J.; King, A.; Sorichetta, A.; Linard, C.; Tatem, A.J. Comparisons of two global built area land cover datasets in methods to disaggregate human population in eleven countries from the global South. *Int. J. Digit. Earth* **2020**, *13*, 78–100. [[CrossRef](#)]
59. Leyk, S.; Gaughan, A.E.; Adamo, S.B.; de Sherbinin, A.; Balk, D.; Freire, S.; Rose, A.; Stevens, F.R.; Blankespoor, B.; Frye, C.; et al. The spatial allocation of population: A review of large-scale gridded population data products and their fitness for use. *Earth Syst. Sci. Data* **2019**, *11*, 1385–1409. [[CrossRef](#)]
60. Calka, B.; Bielecka, E. GHS-POP Accuracy Assessment: Poland and Portugal Case Study. *Remote Sens.* **2020**, *12*, 1105. [[CrossRef](#)]
61. Florczyk, A.; Corbane, C.; Ehrlich, D.; Freire, S.; Kemper, T.; Maffenini, L.; Melchiorri, M.; Pesaresi, M.; Politis, P.; Schiavina, M.; et al. *GHS Data Package 2019*; Technical Report, EUR 29788EN; Publications Office of the European Union: Luxembourg, 2019. [[CrossRef](#)]
62. Markoff, J.; Shapiro, G. The Linkage of Data Describing Overlapping Geographical Units. *Hist. Methods Newsl.* **1973**, *7*, 34–46. [[CrossRef](#)]
63. Goodchild, M.F.; Lam, N.S.N. Areal Interpolation: A Variant of the Traditional Spatial Problem. *Geo-Processing* **1980**, *1*, 297–312.
64. Liu, Z.; He, C.; Zhou, Y.; Wu, J. How much of the world's land has been urbanized, really? A hierarchical framework for avoiding confusion. *Landsc. Ecol.* **2014**, *29*, 763–771. [[CrossRef](#)]
65. Mudau, N.; Mwaniki, D.; Tsoeleng, L.; Mashalane, M.; Beguy, D.; Ndugwa, R. Assessment of SDG Indicator 11.3.1 and Urban Growth Trends of Major and Small Cities in South Africa. *Sustainability* **2020**, *12*, 7063. [[CrossRef](#)]
66. Schiavina, M.; Melchiorri, M.; Corbane, C.; Florczyk, A.J.; Freire, S.; Pesaresi, M.; Kemper, T. Multi-Scale Estimation of Land Use Efficiency (SDG 11.3.1) across 25 Years Using Global Open and Free Data. *Sustainability* **2019**, *11*, 5674. [[CrossRef](#)]
67. Mennis, J. Dasymetric Mapping for Estimating Population in Small Areas. *Geogr. Compass* **2009**, *3*, 727–745. [[CrossRef](#)]
68. Zandbergen, P.A.; Ignizio, D.A. Comparison of Dasymetric Mapping Techniques for Small-Area Population Estimates. *Cartogr. Geogr. Inf. Sci.* **2010**, *37*, 199–214. [[CrossRef](#)]
69. Min, B. *Power and the Vote: Elections and Electricity in the Developing World*; Cambridge University Press: Cambridge, UK, 2015. [[CrossRef](#)]
70. Wang, P.; Huang, C.; Brown de Colstoun, E.C. Mapping 2000–2010 Impervious Surface Change in India Using Global Land Survey Landsat Data. *Remote Sens.* **2017**, *9*, 366. [[CrossRef](#)]
71. Koren, O.; Sarbahi, A.K. State Capacity, Insurgency, and Civil War: A Disaggregated Analysis. *Int. Stud. Q.* **2018**, *62*, 274–288. [[CrossRef](#)]
72. Hu, Y.; Yao, J. *Illuminating Economic Growth*; Technical Report, International Monetary Fund, Working Paper No. 19/77; IMF: Washington, DC, USA, 2019.
73. Small, C. International Earth Science Information Network (CIESIN). *VIIRS Plus DMSP Change in Lights*; Columbia University, Center for International Earth Sciences Information Network (CIESIN): Palisades, NY, USA, 2020.
74. Ch, R.; Martin, D.A.; Vargas, J.F. Measuring the size and growth of cities using nighttime light. *J. Urban Econ.* **2020**, *125*, 103254. [[CrossRef](#)]
75. Li, X.; Zhou, Y.; Zhao, M.; Zhao, X. A harmonized global nighttime light dataset 1992–2018. *Sci. Data* **2020**, *7*, 168. [[CrossRef](#)]
76. Mahtta, R.; Mahendra, A.; Seto, K.C. Building up or spreading out? Typologies of urban growth across 478 cities of 1 million+. *Environ. Res. Lett.* **2019**, *14*, 124077. [[CrossRef](#)]
77. Balk, D.L.; Nghiem, S.V.; Jones, B.R.; Liu, Z.; Dunn, G. Up and out: A multifaceted approach to characterizing urbanization in Greater Saigon, 2000–2009. *Landsc. Urban Plan.* **2019**, *187*, 199–209. [[CrossRef](#)]
78. Rowlands, D.W.; Loh, T.H. *Reinvesting in Urban Cores Can Revitalize Entire Regions*; Metropolitan Policy Program Report; The Brookings Institution: Washington, DC, USA, 2021.
79. Jones, B.; Balk, D.; Leyk, S. Urban Change in the United States, 1990–2010: A Spatial Assessment of Administrative Reclassification. *Sustainability* **2020**, *12*, 1649. [[CrossRef](#)]
80. Arribas-Bel, D.; Garcia-López, M.; Viladecans-Marsal, E. Building(s and) cities: Delineating urban areas with a machine learning algorithm. *J. Urban Econ.* **2019**, *125*, 103217. [[CrossRef](#)]
81. de Bellefon, M.P.; Combes, P.P.; Duranton, G.; Gobillon, L.; Gorin, C. Delineating urban areas using building density. *J. Urban Econ.* **2019**, *125*, 103226. [[CrossRef](#)]
82. Melchiorri, M.; Florczyk, A.J.; Freire, S.; Schiavina, M.; Pesaresi, M.; Kemper, T. Unveiling 25 Years of Planetary Urbanization with Remote Sensing: Perspectives from the Global Human Settlement Layer. *Remote Sens.* **2018**, *10*, 768. [[CrossRef](#)]
83. Yang, F.; Wang, Z.; Yang, X.; Liu, Y.; Liu, B.; Wang, J.; Kang, J. Using Multi-Sensor Satellite Images and Auxiliary Data in Updating and Assessing the Accuracies of Urban Land Products in Different Landscape Patterns. *Remote Sens.* **2019**, *11*, 2664. [[CrossRef](#)]
84. Brenner, N.; Schmid, C. The 'Urban Age' in Question. *Int. J. Urban Reg. Res.* **2014**, *38*, 731–755. [[CrossRef](#)]

-
85. United Nations. *A New Global Partnership: Eradicate Poverty and Transform Economics through Sustainable Development. The Report of the High-Level Panel of Eminent Persons on the Post-2015 Development Agenda*; United Nations: New York, NY, USA, 2013.
 86. Balk, D.; Montgomery, M.R.; McGranahan, G.; Todd, M. Understanding the Impacts of Climate Change: Linking Satellite and Other Spatial Data with Population Data. In *Population Dynamics and Climate Change*; Guzmán, J.M., Martine, G., McGranahan, G., Shensul, D., Tacoli, C., Eds.; United Nations Fund for Population Activities (UNFPA) and International Institute for Environment and Development (IIED): New York, NY, USA, 2009; Chapter 13, pp. 206–217.
 87. Tatem, A.J.; Campiz, N.; Gething, P.W.; Snow, R.W.; Linard, C. The effects of spatial population dataset choice on estimates of population at risk of disease. *Popul. Health Metrics* **2011**, *9*, 4. [[CrossRef](#)] [[PubMed](#)]

Supplement to “Global harmonization of urbanization measures: Proceed with care”

By Deborah Balk, Stefan Leyk, Mark R. Montgomery and Hasim Engin

This supplement illustrates ways in which the estimation is sensitivity to inputs, both the resolution of underlying inputs (in the administrative data) and in the choice of threshold of built-up levels. In the Appendix to the paper, we highlight areas of sensitivity, suggesting that of the officially urban locations, built-up levels of less than 50% may be quite reasonable. In this supplement, we consider two such lower levels, $\tau = 25\%$ and 1% built-up.

Implications of using higher resolution administrative inputs for India

The analysis reported in the main body of the paper uses the global Degree of Urbanization (DoU) data product (Florczyk et al., 2019). For India, as population inputs, it uses publicly-available subdistrict-level administrative data (about 5,500 units). In this appendix, to illustrate how the spatial resolution of the underlying census data in the DoU impacts the resulting classification, the DoU method was applied¹ using the settlement level (about 650,000 units) boundaries and associated population described in the main text (Balk et al., 2019). The resulting spatial distribution can be seen in Figure S2 and Tables S1-S3.

It is well known that spatial resolution matters and gains in resolution tend to lead to more accurate estimates of population distributions (Tatem et al. 2011, Leyk et al, 2018, Leyk et al., 2019, Balk et al., 2009). In this illustration, because we also assume that the finer resolution data will produce a more accurate rendering of the spatial distribution of the population, we also assume that the resulting reallocation in the GHS-Pop data, which is an intermediary input to the DoU classification, will also produce a more accurate description of the population distribution of India. However, as we point out elsewhere (Balk et al., 2019) the spatial precision and accuracy of the Indian settlement-level boundaries is to some extent unknown. Data used here were matched with district and state borders, but we made the fewest alterations to the underlying boundaries as possible due to a comprehensive lack of objective criteria upon which to do so. It is worth noting that this is a more general concern globally: fine-resolution spatial data simply because they include orders of magnitudes more units provide ample opportunities for errors to go undetected. Small errors in fine resolution data could be amplified when integrated with other data that is also spatially refined (such as the built-up area data). So, while we expect qualitative improvements in resulting classifications, we expect some errors resulting from inconsistencies arising from combining multiple high-resolution spatial data.

Apart from the underlying data, in the application of the DoU method, Dijkstra et al. (2020) report that the size of the unit -- that is, the grid cell -- has an influence on the population density estimation used as part of the criteria to create the classification (see fn 15, on India vs. US), with coarser units resulting in lower average densities.

The left-panel map, for New Delhi and surrounding areas, is based on the coarser, subdistrict inputs and thus classifies more land (and corresponding population) as urban centers. It also renders more land as very sparsely or low-density rural. In contrast, on the right panel, based on settlement-level data with more detailed distributions of the popu-

¹ We thank our colleagues in the GHSL Team of the European Commission, DG Joint Research Centre – JRC E.1 Group, Disaster Risk Management Unit, for producing GHS-Pop and applying the DoU method to them.

lation, there is less reallocation underway. This produces more land area classified as sub-urban or peri-urban, as indicated clearly in the regions on the outskirts and expanding from this capital region.

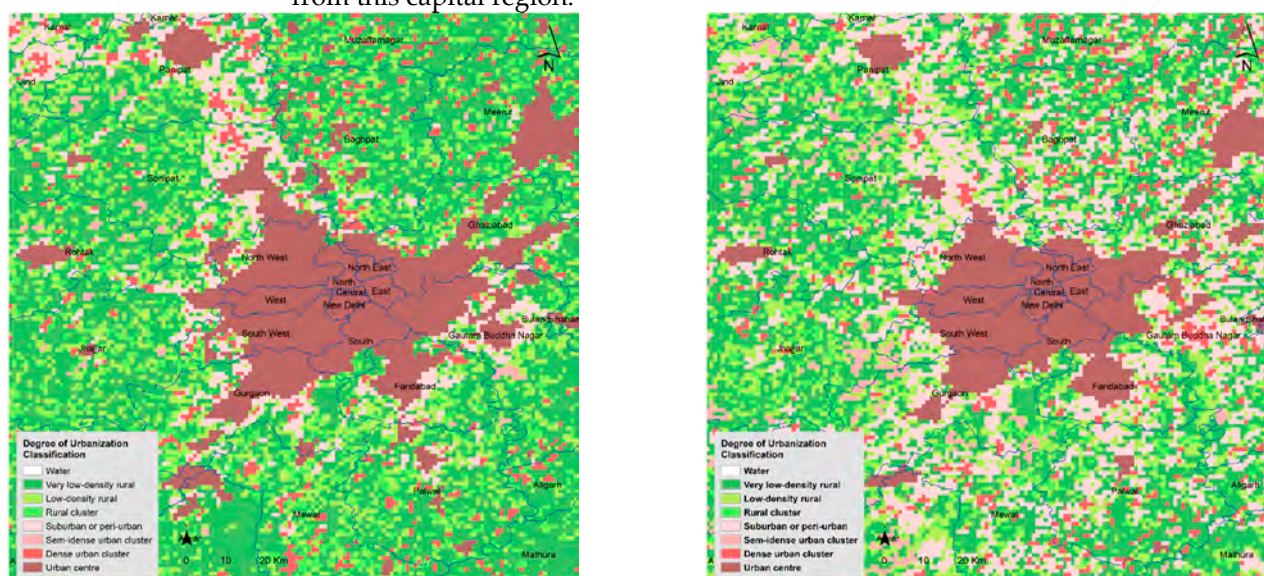


Figure S1. Degree of urbanization, New Delhi and surrounding areas (2015). Left panel shows DoU distribution using the global data product based on sub-district-level population data; Right panel shows DoU distribution produced with the settlement-level population data.

The pattern illustrated for New Delhi and surrounding areas is also reflected clearly in the national-level results, shown in Tables S1–S3. When using moderate resolution inputs, 2.3% of total land areas is estimated to be an urban center, in contrast to 1.6% when using the finer resolution inputs (Table S1). While these percentages are small, given that cities tend to occupy relatively small fractions of land in the first place, this difference is significant in relative terms. The estimated population living in urban centers, however, remains largely unchanged between the two approaches (Table S2).

Table S1. Total Area (km²) by Degree of Urbanization, using population data of varying spatial resolutions, India.

Degree of Urbanization Classification	Administrative-level used			
	Settlements		Subdistricts	
Urban centre	51,886	1.59%	75,094	2.31%
Dense urban	55,003	1.69%	51,522	1.58%
Semi-dense urban	54,909	1.68%	4,856	0.15%
Suburban or peri-urban	345,926	10.61%	41,477	1.27%
Rural	180,629	5.54%	234,587	7.21%
Low-density rural	821,116	25.19%	304,857	9.37%
Very low-density rural	1,750,424	53.70%	2,541,742	78.11%

One of most two notable differences is found in the suburban or peri-urban category, where the fine-resolution data estimate about 10 times more land area (Table S1) and 8 times more persons (Table S2) in that class. These very different urban forms, no doubt, would have significant implications for understanding urban life and future urbanization. The other notable difference that emerges is on the rural end of the continuum. The fine-resolution data estimate less land (Table S1) and population (Table S2) residing in all three rural classes, and perhaps most notably place 53.7% of India's land area and 23.5% of its population in "very low-density rural" areas. In contrast, the moderate-resolution inputs estimate 78.1% of India's land and 46.8% of its population live in these areas.

Table S2. Total Population (000) by Degree of Urbanization, using population data of varying spatial resolutions, India.

Degree of Urbanization Classification	Administrative-level used			
	Settlements		Subdistricts	
Urban centre	279,961	23.13%	296,221	24.52%
Dense urban	77,038	6.36%	45,516	3.77%
Semi-dense urban	25,168	2.08%	1,988	0.16%
Suburban or peri-urban	280,205	23.15%	38,501	3.19%
Rural	55,796	4.61%	128,965	10.68%
Low-density rural	207,360	17.13%	131,087	10.85%
Very low-density rural	284,868	23.54%	565,639	46.83%

Table S3 shows population densities, and leads to much different narratives. As the DoU based on the fine-resolution data reveals a smaller footprint for urban centers, the population density of such locations is even greater than estimated by moderate-level inputs. It is notably higher by more than 1000 persons/km². Similarly, densities of ‘dense urban’ and ‘semi-dense urban’ areas are also higher when using the fine-resolution data. Conversely, suburban/peri-urban areas and all of the rural classes are less densely populated (despite much less land area having been estimated in these classes), though they still remain high compared to these classes for Mexico or the United States.

Table S3. Total Density (persons/km²) by Degree of Urbanization, using population data of varying spatial resolutions, India.

Degree of Urbanization Classification	Administrative-level used	
	Settlements	Subdistricts
Urban centre	5,396	3,945
Dense urban	1,401	883
Semi-dense urban	458	409
Suburban or peri-urban	810	928
Rural	309	550
Low-density rural	253	430
Very low-density rural	163	223

On one hand, the value-added by the remote sensing data inputs is perhaps greatest where census administrative inputs are coarse, because they help to create necessary spatial refinements. But on the other hand, integrated data models such as the DoU stand to be most accurate when built on a baseline of fine-resolution census inputs. (This also raises a related question of whether there is an optimal spatial resolution of census inputs in DoU models.) This illustration suggests that DoU classifications are dependent on the spatial resolution of the key input data, and while much attention has gone to the critical review of the satellite data underlying the DoU model, the spatial resolution of the census units clearly matters as well, and deserves critical and systematic review beyond this example here.

Sensitivity to built-up threshold (τ).

Tables 4–7 show the results of cross-classification based on inclusive built-up levels, using $\tau = 25$ and 1 percent thresholds. As expected, more inclusive built-up levels lead to more land area (Table S4) and population (Table S5) for which there is urban agreement. Most notable is the case of India, for which 3 times the land area and more than twice the population falls into this category when going from a $\tau = 50\%$ to a $\tau = 1\%$. Substantial increases are also observed in Mexico and the US, capturing another approximately 25%

of the total population as being cross-classified as urban agreement from around 57%. This suggests that flexible τ parameters would be useful in capturing officially urban areas along the continuum of rural-urban locations.

Table S4. Total (km²) and percentage area by cross-classifications of official urban designations with GHSL, using alternative built-up thresholds.

Built-up threshold and agreement with census classification		Total Area (km ²), Count and Percentage					
		India (2011)		Mexico (2010)		USA (2010)	
50% Built-up	Urban agreement	12,580	0.4%	9,779	0.5%	113,853	1.5%
	Urban, not built-up	97,203	3.0%	13,164	0.7%	165,413	2.1%
	Rural, but built-up	5,014	0.2%	1,541	0.1%	13,618	0.2%
	Rural agreement	3,113,781	96.4%	1,940,730	98.8%	7,517,111	96.2%
25% Built-up	Urban agreement	19,858	0.6%	13,237	0.7%	165,061	2.1%
	Urban, not built-up	89,926	2.8%	9,705	0.5%	114,205	1.5%
	Rural, but built-up	17,346	0.5%	4,872	0.2%	46,211	0.6%
	Rural agreement	3,101,448	96.1%	1,937,399	98.6%	7,484,518	95.8%
1% Built-up	Urban agreement	40,767	1.3%	19,151	0.2%	240,984	3.1%
	Urban, not built-up	68,188	2.1%	3,791	1.0%	38,281	0.5%
	Rural, but built-up	140,077	4.3%	31,343	1.6%	341,479	4.4%
	Rural agreement	2,979,546	92.3%	1,910,929	97.2%	7,189,250	92.1%

However, lower τ parameters also lead to more land and population classified as rural, but built-up reassigning them from areas of rural agreement. More research could help ascertain whether these areas are indeed rural in character (and how they could be described to be more accurately classified) or places that are the cusp of becoming reclassified as urban (Jones et al., 2020).

Table S5. Total (000s) and percentage population by cross-classifications of official urban designations with GHSL, using alternative built-up thresholds.

Built-up threshold and agreement with census classification		Total Population (000) Count and Percentage					
		India (2011)		Mexico (2010)		USA (2010)	
50% Built-up	Urban agreement	129,362	10.7%	60,550	57.0%	177,569	57.9%
	Urban, not built-up	247,747	20.5%	25,018	23.5%	69,948	22.8%

25% Built-up	Rural, but built-up	5,521	0.5%	96	0.1%	2,287	0.7%
	Rural agreement	828,225	68.4%	20,589	19.4%	56,854	18.5%
	Urban agreement	167,836	13.9%	70,638	66.5%	211,077	68.8%
	Urban, not built-up	209,273	17.3%	14,930	14.1%	36,439	11.9%
	Rural, but built-up	15,707	1.3%	278	0.3%	6,396	2.1%
	Rural agreement	818,039	67.6%	20,406	19.2%	52,745	17.2%
	Urban agreement	239,729	19.8%	81,718	76.9%	239,956	78.2%
	Urban, not built-up	134,741	11.1%	3,850	3.6%	7,561	2.5%
	Rural, but built-up	85,833	7.1%	1,429	1.3%	18,812	6.1%
	Rural agreement	750,551	62.0%	19,256	18.1%	40,329	13.2%

Of course, population (Table S6) and built-up (Table S7) densities are also responsive to alternative specifications of τ . Declines in all cross-classifications are evident as τ is lowered. Most notable is that while areas of Urban Agreement are pretty similar in all three study countries, once τ is lowered, India looks much different from Mexico and the US. At a built-up threshold of 1% areas of Urban Agreement in India are only 35% built-up as opposed to close to 50% in Mexico and the US. Whether this reflects differences in urban form, levels of economic development or satellite detection across the three countries remains an open question.

Table S6. Population density by cross-classifications of official urban designations with GHSL, using alternative built-up thresholds.

Built-up threshold and agreement with census classification		Persons/km ²		
		India (2011)	Mexico (2010)	USA (2010)
50% Built-up	Urban agreement	10,283	6,192	1,560
	Urban, not built-up	2,549	1,901	423
	Rural, but built-up	1,101	62	168
	Rural agreement	266	11	8
25% Built-up	Urban agreement	8,452	5,336	1,279
	Urban, not built-up	2,327	1,538	319
	Rural, but built-up	905	57	138
	Rural agreement	264	11	7
1% Built-up	Urban agreement	5,881	4,267	996
	Urban, not built-up	1,976	1,016	198
	Rural, but built-up	613	46	55

Rural agreement	252	10	6
-----------------	-----	----	---

Table S7. Built-up density by cross-classifications of official urban designations with GHSL, using alternative built-up thresholds.

Built-up threshold and agreement with census classification		Mean Built-up %		
		India (2011)	Mexico (2010)	USA (2010)
50% Built-up	Urban agreement	78.3	82.0	79.5
	Urban, not built-up	4.7	14.4	16.7
	Rural, but built-up	66.8	66.5	67.5
	Rural agreement	0.4	0.2	0.4
25% Built-up	Urban agreement	63.0	70.27	66.4
	Urban, not built-up	2.1	6.5	7.5
	Rural, but built-up	44.0	44.8	44.3
	Rural agreement	0.3	0.1	0.6
1% Built-up	Urban agreement	35.3	51.8	49.0
	Urban, not built-up	0.0	0.1	0.1
	Rural, but built-up	11.3	12.7	12.1
	Rural agreement	0.0	0.0	0.0

Alluvial Plots: Sensitivity to alternative built-up thresholds, and to input-resolution of India census data.

The sensitivity to built-up threshold can also be seen in Figures S2 for India, S3 for Mexico and S4 for the U.S. showing $\tau = 25\%$ and 1% , respectively. It is clear that as τ is low, allowing for more inclusive criteria for considering built-up land as likely urban, we see many more officially *rural areas that are built-up*. In India, this class according to DoU is heterogeneous, but in the U.S., it is largely classified as low-density rural, suggesting that 1% built-up may have a more urban signature in some settings than in others. Similarly, the class of *urban, not built-up* declines as τ decreases. In all countries, this is heterogeneous class: when it diminishes, it places those persons into the *urban centre* class, which owing to its dominance of population for whom there is *urban agreement*, it is hard to disentangle the contribution of the alternative τ levels to that class.

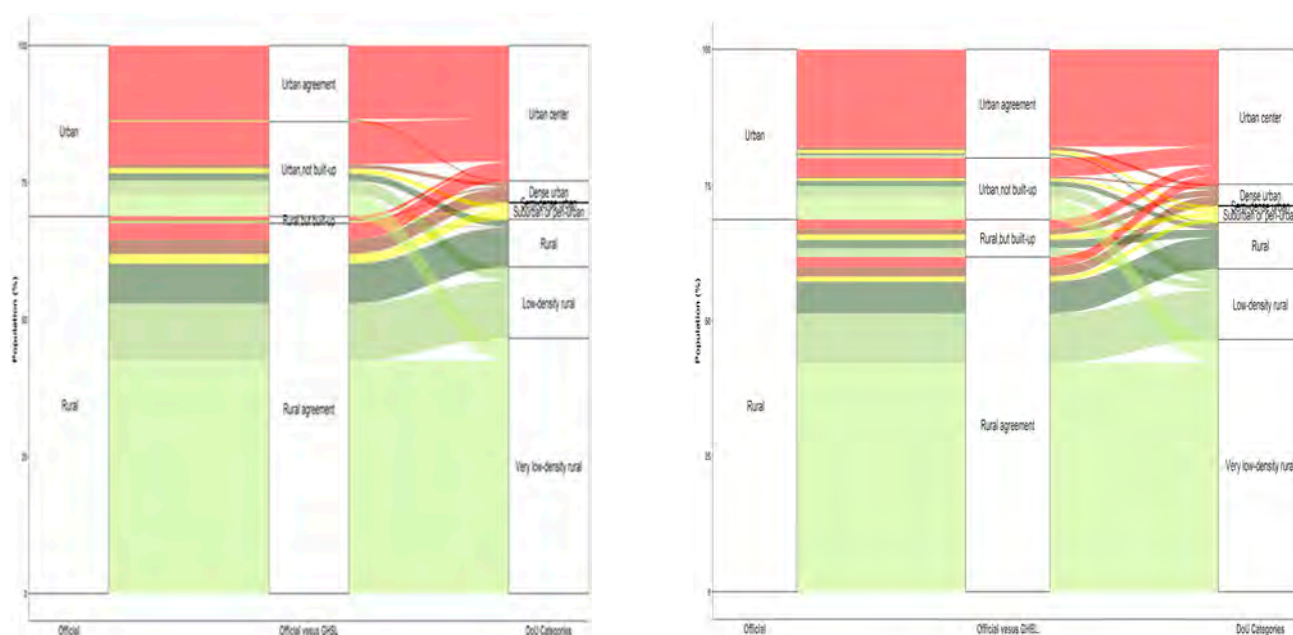


Figure S2. Alluvial Plots India, 25% and 1% thresholds.

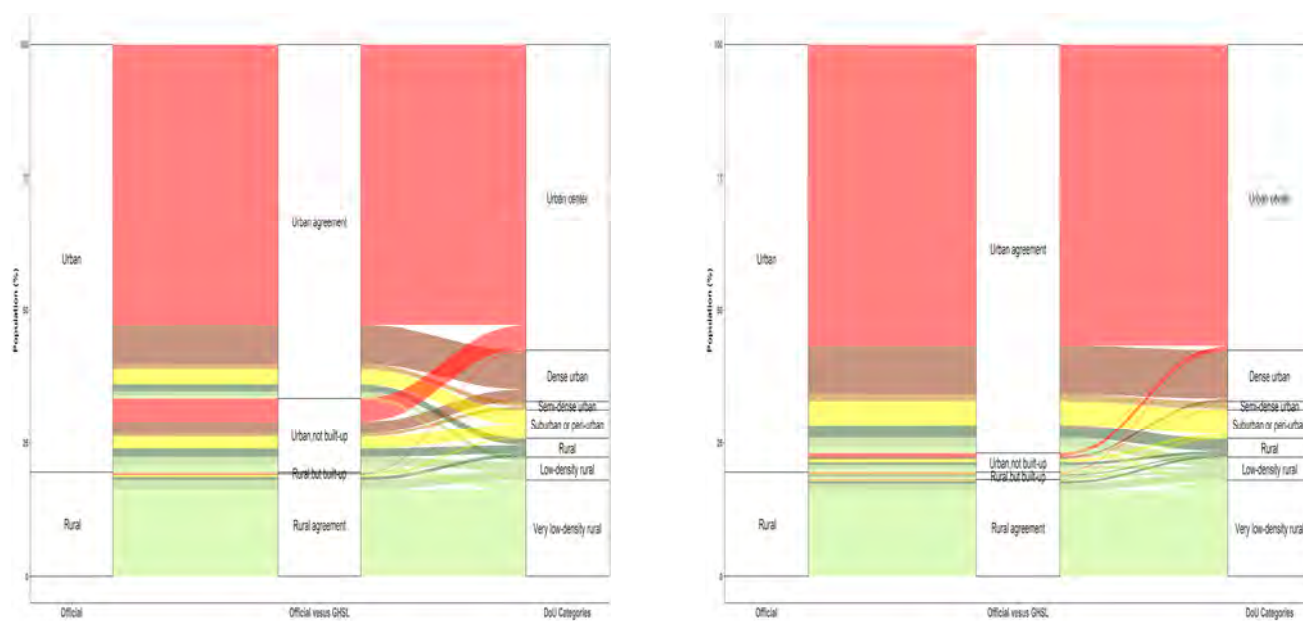


Figure S3. Alluvial Plots Mexico, 25% and 1% thresholds.

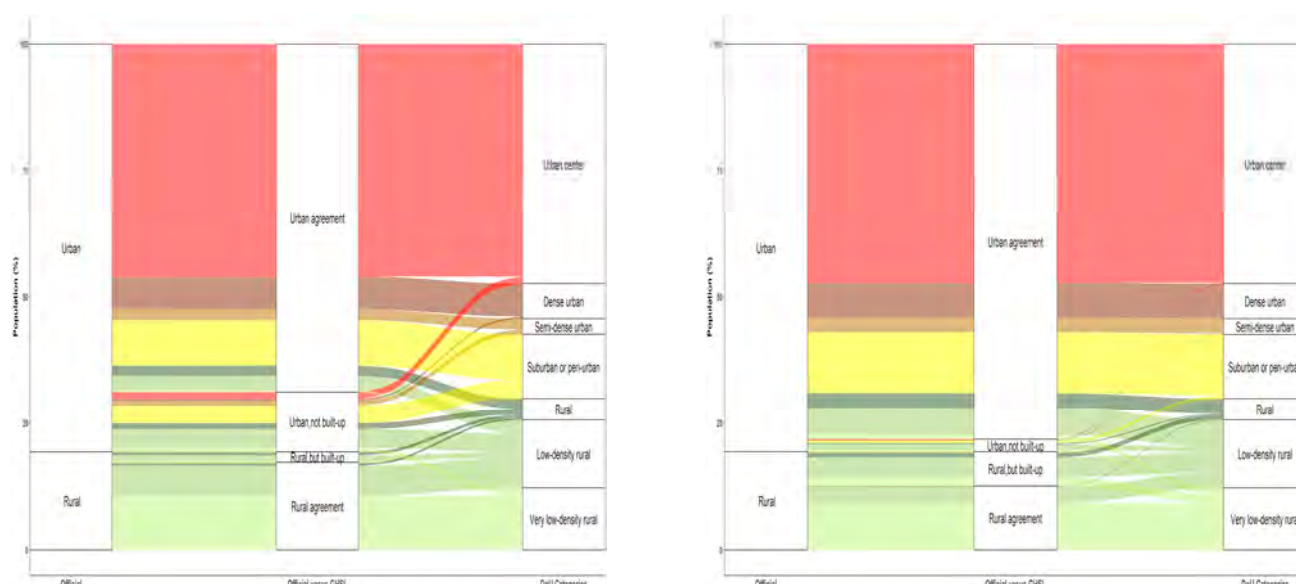


Figure S4. Alluvial Plots USA, 25% and 1% thresholds.

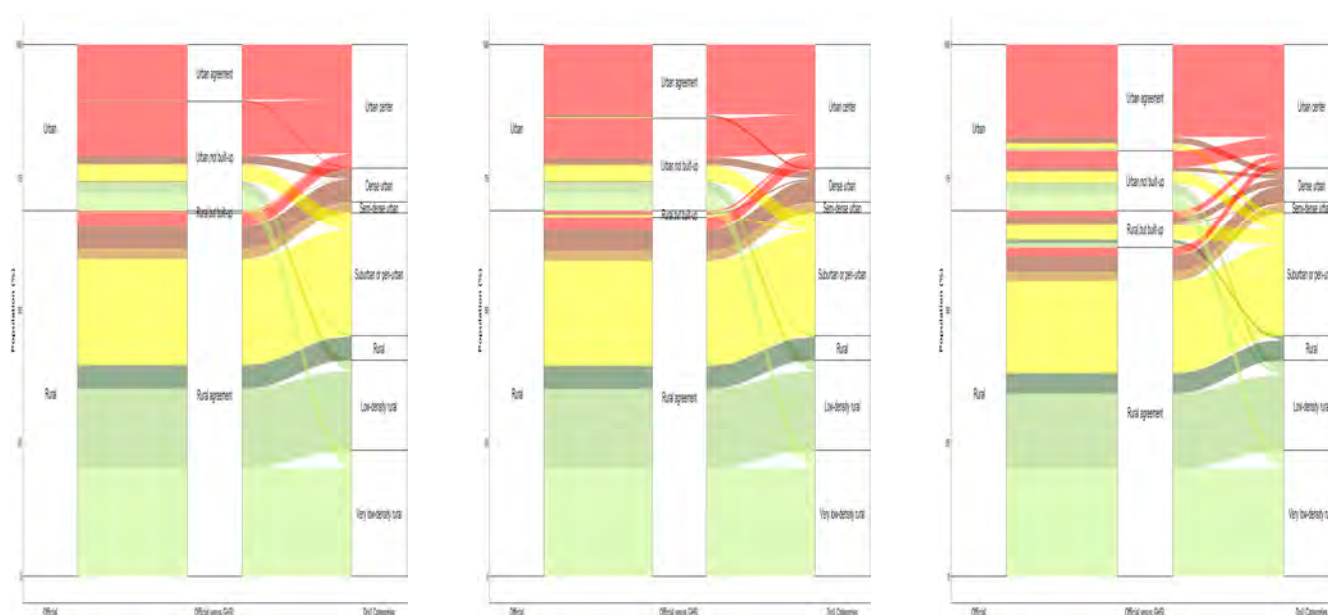


Figure S5. Alluvial Plots India using settlement-level population data, $\tau = 50\%$ 25% and 1% thresholds, respectively.

Data corresponding to Figure 7 in the main text are shown below in Table S8–S10 below. Note that Table S10 also includes the finer resolution data for India, as part of an additional sensitivity analysis.

Table S8. Three-way cross-classifications of official urban–rural categories, GHSL–official classification, and DoU classes, Mexico. (Official figures from 2010 census; GHSL estimate for 2014; DoU estimate for 2015.).

Rural - Urban Classification Schema			Population		Area	
Official	Official vs. GHSL ($\tau=50\%$)	Degree of Urbanization	Count	%	Km ²	%
Urban		Urban center	50,570,955	47.7	6,900	0.35

Urban agreement	Dense urban	5,918,316	5.6	1,379	0.07
	Semi-dense urban	589,437	0.6	225	0.01
	Suburban or peri-urban	1,814,486	1.7	602	0.03
	Rural	764,570	0.7	334	0.02
	Low-density rural	498,285	0.5	192	0.01
	Very low-density rural	268,959	0.3	104	0.01
	Urban center	10,069,152	9.5	3,076	0.16
	Dense urban	4,231,275	4.0	1,627	0.08
	Semi-dense urban	1,012,945	1.0	621	0.03
	Suburban or peri-urban	3,599,607	3.4	2,309	0.12
	Rural	2,306,478	2.2	1,624	0.08
	Low-density rural	2,081,867	2.0	2,011	0.10
	Very low-density rural	1,637,906	1.5	1,833	0.09
Rural	Urban center	27,181	0.0	202	0.01
	Dense urban	4,088	0.0	74	0.00
	Semi-dense urban	2,593	0.0	55	0.00
	Suburban or peri-urban	20,631	0.0	249	0.01
	Rural	20,401	0.0	444	0.02
	Low-density rural	15,303	0.0	342	0.02
	Very low-density rural	4,846	0.0	162	0.01
	Urban center	193,817	0.2	2,047	0.11
	Dense urban	80,893	0.1	1,660	0.09
	Semi-dense urban	74,287	0.1	1,637	0.08
	Suburban or peri-urban	383,091	0.4	5,849	0.30
	Rural	693,452	0.7	16,861	0.87

	Low-density rural	1,933,475	1.8	50,320	2.59
	Very low-density rural	17,113,828	16.2	1,842,629	94.72

Table S9. Three-way cross-classifications of official urban–rural categories, GHSL–official classification, and DoU classes, USA. (Official figures from 2010 census; GHSL estimate for 2014; DoU estimate for 2015.).

Rural - Urban Classification Schema			Population		Area	
Official	Official vs. GHSL ($\tau=50\%$)	Degree of Urbanization	Count	%	Km ²	%
Urban	Urban agreement	Urban center	130,623,637	42.8	67,091	0.87
		Dense urban	16,906,636	5.5	12,740	0.16
		Semi-dense urban	3,960,090	1.3	3,806	0.05
		Suburban or peri-urban	16,355,634	5.4	15,439	0.20
		Rural	3,759,080	1.2	4,066	0.05
		Low-density rural	4,025,917	1.3	7,282	0.09
		Very low-density rural	763,456	0.3	2,635	0.03
	Urban, not built-up	Urban center	13,776,616	4.5	22,040	0.29
		Dense urban	4,063,286	1.3	5,505	0.07
		Semi-dense urban	5,350,175	1.8	8,993	0.12
		Suburban or peri-urban	21,950,368	7.2	39,849	0.52
		Rural	5,901,485	1.9	11,903	0.15
		Low-density rural	15,508,722	5.1	56,759	0.73
		Very low-density rural	3,052,633	1.0	19,331	0.25
Rural	Rural, but built-up	Urban center	16,355	0.0	385	0.00
		Dense urban	22,352	0.0	275	0.00
		Semi-dense urban	22,569	0.0	212	0.00
		Suburban or peri-urban	49,851	0.0	599	0.01

Rural agreement	Rural	1,110,515	0.4	1,752	0.02
	Low-density rural	1,298,186	0.4	5,222	0.07
	Very low-density rural	265,278	0.1	5,019	0.06
	Urban center	40,879	0.0	1,491	0.02
	Dense urban	45,644	0.0	976	0.01
	Semi-dense urban	212,412	0.1	3,361	0.04
	Suburban or peri-urban	445,248	0.1	8,146	0.11
	Rural	2,133,724	0.7	13,701	0.18
	Low-density rural	20,114,613	6.6	334,946	4.33
	Very low-density rural	33,166,925	10.9	7,076,286	91.55

Table S10. Three-way cross-classifications of official urban–rural categories, GHSL–official classification, and DoU classes, India. Upper panel reflects Global DoU based on subdistricts, and lower panels reflect DoU using settlement-level administrative data. (Official figures from 2010 census; GHSL estimate for 2014; DoU estimate for 2015.).

Rural - Urban Classification Schema			Population		Area	
Official	Official vs. GHSL ($\tau=50\%$)	Degree of Urbanization	Count	%	Km ²	%
India (based on sub-district level administrative data for DoU)	Urban agreement	Urban center	126,969,757	10.5	11,707	0.36
		Dense urban	1,703,605	0.1	657	0.02
		Semi-dense urban	13,975	0.0	6	0.00
		Suburban or peri-urban	544,317	0.0	151	0.00
		Rural	52,913	0.0	22	0.00
		Very low-density rural	10,777	0.0	14	0.00
		Low-density rural	61,780	0.0	21	0.00
	Urban, not built-up	Urban center	129,225,897	10.7	28,416	0.88
		Dense urban	9,977,351	0.8	5,291	0.16
		Semi-dense urban	177,541	0.0	118	0.00

India (based on settlement level administrative data for	Rural		Suburban or peri-urban	12,429,389	1.0	5,137	0.16
			Rural	16,015,567	1.3	7,765	0.24
			Very low-density rural	49,809,393	4.1	34,978	1.09
			Low-density rural	29,506,245	2.4	15,095	0.47
		Rural, but built-up	Urban center	3,137,590	0.3	1,941	0.06
			Dense urban	1,390,631	0.1	1,485	0.05
			Semi-dense urban	58,610	0.0	138	0.00
			Suburban or peri-urban	543,592	0.0	610	0.02
			Rural	269,948	0.0	642	0.02
			Very low-density rural	11,680	0.0	18	0.00
			Low-density rural	61,913	0.0	107	0.00
		Rural agreement	Urban center	39,403,869	3.3	32,888	1.02
			Dense urban	33,518,468	2.8	44,242	1.37
			Semi-dense urban	1,869,384	0.2	4,371	0.14
			Suburban or peri-urban	22,436,401	1.9	31,044	0.96
			Rural	88,673,377	7.3	182,591	5.67
			Very low-density rural	512,760,264	42.5	2,474,191	76.85
			Low-density rural	127,232,327	10.5	335,695	10.43
	Urban	Urban agreement	Urban center	125,655,272	10.4	10,984	0.34
			Dense urban	3,280,001	0.3	1,314	0.04
			Semi-dense urban	12,184	0.0	18	0.00
			Suburban or peri-urban	381,337	0.0	241	0.01
			Rural	6,485	0.0	7	0.00
			Low-density rural	23,011	0.0	16	0.00

Rural	Urban, not built-up	Very low-density rural	3,605	0.0	2	0.00
		Urban center	121,030,785	10.0	22,295	0.69
		Dense urban	20,341,163	1.7	8,843	0.27
		Semi-dense urban	731,300	0.1	827	0.03
		Suburban or peri-urban	38,378,160	3.2	17,922	0.56
		Rural	2,010,488	0.2	1,618	0.05
		Low-density rural	24,248,291	2.0	14,435	0.45
		Very low-density rural	40,815,460	3.4	31,112	0.96
	Rural, but built-up	Urban center	2,689,663	0.2	1,331	0.04
		Dense urban	1,455,581	0.1	1,252	0.04
		Semi-dense urban	120,052	0.0	262	0.01
		Suburban or peri-urban	1,002,078	0.1	1,286	0.04
		Rural	191,295	0.0	606	0.02
		Low-density rural	54,952	0.0	205	0.01
		Very low-density rural	7,660	0.0	72	0.00
	Rural agreement	Urban center	30,585,257	2.5	17,198	0.53
		Dense urban	51,961,256	4.3	43,538	1.35
		Semi-dense urban	24,303,967	2.0	53,734	1.67
		Suburban or peri-urban	240,443,117	19.9	326,196	10.12
		Rural	53,587,378	4.4	178,201	5.53
		Low-density rural	183,033,742	15.1	805,075	24.97
		Very low-density rural	244,040,910	20.2	1,685,950	52.29

References

- Balk, D.; Montgomery, M.R.; Engin, H.; Lin, N.; Major, E.; Jones, B. Urbanization in India: Population and Urban Classification Grids for 2011. *Data* 2019, 4. <https://doi.org/10.3390/data4010035>.
- Balk, D. M. R. Montgomery, G. McGranahan, and M. Todd, 2009. "Understanding the Impacts of Climate Change: Linking Satellite and Other Spatial Data with Population Data," In G. Martine, J.M. Guzman, G. McGranahan, D. Schensul, and C. Tacoli (editors), *Population Dynamics and Climate Change*, New York: United Nations Population Fund and International Institute for the Environment and Development, pp. 206-217. http://www.environmentportal.in/files/pop_dynamics_climate_change.pdf#page=217
- Dijkstra, L.; Florczyk, A.J.; Freire, S.; Kemper, T.; Melchiorri, M.; Pesaresi, M.; Schiavina, M. Applying the Degree of Urbanisation to the globe: A new harmonised definition reveals a different picture of global urbanisation. *Journal of Urban Economics* 2020, p. 103312. doi:<https://doi.org/10.1016/j.jue.2020.103312>.
- Florczyk, A.; Corbane, C.; Ehrlich, D.; Freire, S.; Kemper, T.; Maffenini, L.; Melchiorri, M.; Pesaresi, M.; Politis, P.; Schiavina, M.; Sabo, F.; Zanchetta, L. GHSL Data Package 2019. Technical report, EUR 29788EN, Publications Office of the European Union, Luxembourg, 2019. doi:doi:10.2760/062975.
- Jones, B.; Balk, D.; Leyk, S. 2020. Urban Change in the United States, 1990–2010: A Spatial Assessment of Administrative Reclassification. *Sustainability*, 12, 1–20. <https://doi.org/10.3390/su12041649>
- Leyk, S., Gaughan, A. E., Adamo, S. B., de Sherbinin, A., Balk, D., Freire, S., Rose, A., Stevens, F. R., Blankespoor, B., Frye, C., Comenetz, J., Sorichetta, A., MacManus, K., Pistoiesi, L., Levy, M., Tatem, A. J., and Pesaresi, M.: The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use, *Earth Syst. Sci. Data*, 11, 1385–1409, <https://doi.org/10.5194/essd-11-1385-2019>, 2019.
- Tatem AJ, Campiz N, Gething PW, Snow RW, Linard C. The effects of spatial population dataset choice on estimates of population at risk of disease. *Popul Health Metr.* 2011 Feb 7;9:4. doi: 10.1186/1478-7954-9-4.

List of Supplemental Tables & Figures

- Table S1. Total Area (km²) by Degree of Urbanization, using population data of varying spatial resolutions, India.
- Table S3. Total Density (persons/km²) by Degree of Urbanization, using population data of varying spatial resolutions, India.
- Table S4. Total (km²) and percentage area by cross-classifications of official urban designations with GHSL, using alternative built-up.
- Table S5. Total (000s) and percentage population by cross-classifications of official urban designations with GHSL, using alternative built-up thresholds.
- Table S6. Population density by cross-classifications of official urban designations with GHSL, using alternative built-up thresholds.

Table S7. Built-up density by cross-classifications of official urban designations with GHSL, using alternative built-up thresholds.

Table S8. Three-way cross-classifications of official urban–rural categories, GHSL–official classification, and DoU classes, Mexico. (Official figures from 2010 census; GHSL estimate for 2014; DoU estimate for 2015.)

Table S9. Three-way cross-classifications of official urban–rural categories, GHSL–official classification, and DoU classes, USA. (Official figures from 2010 census; GHSL estimate for 2014; DoU estimate for 2015.)

Table S10. Three-way cross-classifications of official urban–rural categories, GHSL–official classification, and DoU classes, India. Upper panel reflects Global DoU based on subdistricts, and lower panels reflect DoU using settlement-level administrative data. (Official figures from 2010 census; GHSL estimate for 2014; DoU estimate for 2015.)

Figure S1. Degree of urbanization, New Delhi and surrounding areas. Left panel shows DoU distribution using the global data product based on sub-district-level population data; Right panel shows DoU distribution produced with the settlement-level population data.

Figure S2. Alluvial Plots India, 25% and 1% thresholds.

Figure S3. Alluvial Plots Mexico, 25% and 1% thresholds.

Figure S4. Alluvial Plots USA, 25% and 1% thresholds.

Figure S5. Alluvial Plots India using settlement-level population data, $\tau = 50\%$ 25% and 1% thresholds, respectively.