



**HAL**  
open science

## An Anatomy of Urbanization in Sub-Saharan Africa

Pierre-Philippe Combes, Clément Gorin, Shohei Nakamura, Mark Roberts,  
Benjamin Stewart

► **To cite this version:**

Pierre-Philippe Combes, Clément Gorin, Shohei Nakamura, Mark Roberts, Benjamin Stewart. An Anatomy of Urbanization in Sub-Saharan Africa. 2023. hal-04345529

**HAL Id: hal-04345529**

**<https://sciencespo.hal.science/hal-04345529>**

Preprint submitted on 14 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0  
International License

# AN ANATOMY OF URBANIZATION IN SUB- SAHARAN AFRICA

Pierre-Philippe Combes, Clément Gorin, Shohei Nakamura, Mark Roberts,  
and Benjamin Stewart

---

**SCIENCES PO ECONOMICS DISCUSSION PAPER**

---

No. 2023-06

# An Anatomy of Urbanization in Sub-Saharan Africa

Pierre-Philippe Combes, Clément Gorin,  
Shohei Nakamura, Mark Roberts and Benjamin Stewart\*

**JEL codes:** R12, R23, O55

**Keywords:** urbanization, Sub-Saharan Africa, dartboard approach, satellite imagery, population density

---

**Acknowledgments:** The authors gratefully acknowledge comments provided by participants at the 2021 and 2022 North American Urban Economics Association meetings and the 2021 SMU Urban & Regional Economics conference, as well as additional comments provided by Yue Li and Harris Selod. They further acknowledge the generous assistance provided by Andrew Tatem, Maksym Bondarenko and Alessandro Sorichetta from the University of Southampton's WorldPop group in generating constrained WorldPop gridded population data for 2015. This paper is part of a set of two Policy Research Working Papers that have been produced under the World Bank's "A Move to Consistency – Integrating the Consistent Measurement of Urbanization into Global Poverty Measurement" Advisory Services and Analytics project (P175622). The accompanying paper – entitled "Where is Poverty Concentrated? New Evidence Based on Internationally Consistent Urban and Poverty Measurements" – is available as Policy Research Working Paper No. WPS10620. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

\* Pierre-Philippe Combes is affiliated with the Department of Economics, Sciences Po, and CEPR, and Clément Gorin with the University of Paris 1. Shohei Nakamura, Mark Roberts, and Ben Stewart are affiliated respectively with the World Bank's Poverty & Equity Global Practice; Urban, Resilience, and Land Global Practice; and Geospatial Operations Support Team. Shohei Nakamura and Mark Roberts are the World Bank Task Team Leaders for the project. **Corresponding author:** Mark Roberts (mroberts1@worldbank.org).

## 1. Introduction

Over the last six decades, Sub-Saharan Africa (SSA) has been the world's fastest urbanizing region. Between 1960 and 2021, the share of its population living in officially classified urban areas increased almost 15-fold from around 33.3 million to just over 494 million, while the share of its population living in such areas climbed from 14.6 to 41.8 percent.<sup>1</sup> This rapid urbanization is, moreover, set to continue in the decades ahead. By 2050, more than 1.26 billion people will be living in SSA's towns and cities, more than 2.5 times the number today. Such rapid urbanization carries with it important economic, social, and environmental implications for the development of a region in which more than one-third of the population lived in extreme poverty in 2019.<sup>2</sup>

Despite its rapidity and importance, however, remarkably little is known about the detailed anatomy of patterns of urbanization across SSA countries. This neglect reflects both a lack of detailed official data on urban settlements and populations, and, partly also related to this, a relative lack, with some notable exceptions, of research interest among urban economists in the region's urbanization.<sup>3</sup> Against this backdrop, this paper aims to provide a detailed descriptive analysis, for the year *circa* 2015, of patterns of urbanization across SSA countries. In doing so, the paper distinguishes between two types of urban area – cities and towns. To identify these areas, the paper applies a modified version of an algorithm first developed by de Bellefon et al. (2021) to high-resolution gridded population data for the region, which is itself derived from digitized maps of the footprints of *all* buildings in the region from very high-resolution satellite imagery. This allows for a consistent definition of urban areas across all countries in the region, thereby overcoming the measurement problems that arise from relying on official definitions of urban areas, which vary widely across both SSA countries and countries globally (World Bank 2009, Roberts et al. 2017, Dijkstra et al. 2021).

As part of its analysis of urbanization for SSA, the paper presents evidence on key features of the urban landscape, both across and within urban areas. Looking across urban areas, the paper provides evidence on, *inter alia*, aggregate differences in levels of urbanization between countries, numbers of urban areas, the share of land they occupy, their population densities, and the structure of urban hierarchies as captured by, for example, Zipf's law, as well as on how these features of urbanization correlate with both a country's size (i.e., population) and level of development. Meanwhile, looking within cities, it provides evidence on key empirical regularities such as population density gradients and the number of (sub-)centers.

The paper contributes to three interrelated literatures. First, it contributes to a nascent literature that provides new evidence on patterns of urbanization in Africa by leveraging satellite data. This includes a recent paper by Henderson, Peng, and Venables (2022) which uses data from the European Union's Global Human Settlements built cover data set, *GHS-BUILT*, to examine patterns of settlement growth across SSA between 1975 and 2014. Henderson et al. identify a rapid growth in the number of settlements – from about 47,500 to around 111,500 – with evidence of complementary (competitive) patterns of growth for nearby settlements of different (similar) size.<sup>4</sup> They also empirically reject Gibrat's law, which underpins Zipf's law in Gabaix (1999), finding that built-up area growth has been fastest for the settlements that were initially smallest in 1975.

---

<sup>1</sup> The statistics on urbanization cited in this paragraph are based on data from the United Nations' World Urbanization Prospects: 2018 Revision (<https://population.un.org/wup/>).

<sup>2</sup> Extreme poverty is defined based on the global extreme poverty line of \$2.15 a day (2017 Purchasing Power Parity exchange rates). Data on extreme poverty is from the World Bank's World Development Indicators (<https://databank.worldbank.org/source/world-development-indicators>).

<sup>3</sup> Until relatively recently, most empirical urban economics research has been focused on the US and other developed countries, as well as on a few large developing countries such as Brazil, China, India, and Indonesia. Exceptions to the general neglect of SSA countries include several papers by Vernon Henderson and co-authors (Henderson and Turner, 2020; Henderson, Regan, and Venables, 2021; Henderson, Nigmatulina, and Kriticos, 2021).

<sup>4</sup> Henderson et al. do not explicitly distinguish between different types of settlement (i.e., cities versus towns) but instead examine settlement growth in general based on a binary classification of built-up versus non-built-up.

This nascent literature also includes work by OECD/SWAC (2020), which leverages data on populations and built-up area extents for urban agglomerations from the *Africapolis* database. In this database, urban agglomerations are identified as contiguous built-up areas from Google Earth and other satellite imagery, with a population of at least 10,000. Although they focus on a much narrower set of settlements than Henderson et al.,<sup>5</sup> the authors similarly provide evidence of a rapid growth in the number of urban agglomerations between 1950 and 2015 (see also Heinrigs 2021; OECD/UN ECA/AfDB 2022). They further document the emergence of larger networks of cities, some of which straddle national boundaries.

Whereas the above studies focus on describing urbanization dynamics over time, this paper instead focuses on providing a detailed empirical description of the anatomy of SSA's urbanization for a single year – *circa* 2015. A notable advantage of this is that it allows us to make use of much higher quality gridded population data. In doing so, we avoid relying on data for built-up areas only, which allows us to capture the intensive margin of urbanization relating to population density. In particular, as mentioned above, our analysis relies on high-resolution gridded population data that is derived, in part, from the digitization of the footprints of *all* buildings in the region from very high-resolution satellite imagery. Specifically, we use constrained *WorldPop* data from the University of Southampton's WorldPop group for SSA countries as an input into our algorithm for delineating urban areas. Unlike the more standard unconstrained WorldPop data, which allocates population more generally across grid cells, whether built-up or not, based on many spatial covariates of population, the constrained WorldPop algorithm restricts the spatial allocation of population from larger administrative units to only built-up grid cells, where built-up area is identified based on building footprints.<sup>6</sup> The building footprints, in turn, come from the Digitize Africa project, which applied Artificial Intelligence techniques to 50 cm resolution satellite imagery for *circa* 2015 to map all buildings in 51 SSA countries. Because of its precision, this built-up area data is far superior to, for example, the corresponding 2015 built-up area data from the GHS-BUILT data set on which Henderson et al. (2022) rely. Hence, relative to the Digitize Africa data for 2015, GHS-BUILT vastly under-detects built-up area.<sup>7</sup> Compared to Henderson et al. (2022) and OECD/SWAC (2022), focusing on a single time-period also allows us to go deeper into describing patterns of urbanization for that period. This includes providing empirical analysis of the internal spatial structures of urban areas, a topic about which little is known empirically, especially for SSA.

The second, related, literature to which our paper contributes is that on methods for the more accurate and consistent delineation of urban areas across countries. While there is a notable strand of the urban economics literature that has long been concerned with the questions of how to rigorously define urban and metropolitan areas (see, for example, Berry 1960; Fox and Kumar 1965; Berry et al. 1969; Kanemoto and Kurima 2005; Duranton 2015) and how to delineate urban areas across countries in a comparable manner (Hall and Hay 1980; Cheshire and Hay 1989), the development and large-scale application of algorithms for the more accurate and consistent delineation of urban areas across countries was, for a long time, constrained by both data and computing costs. However, with greater availability of increasingly high-resolution and low-cost satellite imagery, improvements in remote sensing techniques for extracting useful data from satellite images, and declining computational costs

---

<sup>5</sup> OECD/SWAC (2020) identify 7,617 urban agglomerations as of 2015. In addition to SSA, their analysis covers North Africa.

<sup>6</sup> For a more detailed description of the differences between *WorldPop*'s constrained and unconstrained data sets see [https://www.worldpop.org/methods/top\\_down\\_constrained\\_vs\\_unconstrained/](https://www.worldpop.org/methods/top_down_constrained_vs_unconstrained/). In both cases, population is allocated to grid cells using machine learning methods described in Stevens et al. (2015).

<sup>7</sup> In the GHS-BUILT data set, the quality of the satellite imagery used to derive built-up area is also likely to be lower for earlier years which introduces an additional source of measurement error for any dynamic analysis of SSA's urbanization process. Similar considerations apply to OECD/SWAC's (2020) dynamic analysis of urbanization patterns in Africa based on the *Africapolis* data set. Hence, while Google Earth provides the underlying source of imagery for 2015 for this data set, it is unclear, from the descriptions of the data in OECD/SWAC (2020) and Heinrigs (2021), what the underlying sources of imagery are for the earlier years, going back to 1950.

for applying algorithms to those data at regional and global scales, there has been a recent upsurge of research interest in urban delineation. This has led to the development and application of several new algorithms for the more accurate and consistent delineation of urban areas across countries (see, *inter alia*, Uchida and Nelson 2009; Dijkstra and Poelman 2014; Zhou, Hubacek, and Roberts 2015; Ch, Martin, and Vargas 2021; Dijkstra et al. 2021).

One shortcoming that most of these algorithms share, however, is that they rely on one or more, essentially arbitrary, thresholds to delineate an urban area. For example, the widely used “degree of urbanization” approach (Dijkstra and Poelman 2014; Dijkstra et al. 2021) relies on two such thresholds to identify urban settlements – namely, an overall population threshold for the settlement and a population density threshold for each 1 km<sup>2</sup> population grid cell that belongs to it.<sup>8,9</sup>

To overcome this shortcoming, we instead apply a modified version of the de Bellefon et al. (2021) “dartboard” algorithm. This algorithm has the advantage that it grounds the identification and delineation of urban areas in a statistical approach which avoids the *a priori* specification of arbitrary, and globally uniform, population (density) thresholds. In essence, and after performing smoothing, the algorithm identifies urban areas as clusters of statistically significant peaks of population or built-up density. This is done by comparing the actual spatial distribution of population or built-up density with the counterfactual spatial distributions that would result if population / built-up density were instead randomly distributed across a country’s potentially inhabitable (or “livable”) area.

Whereas de Bellefon et al. originally applied their algorithm to France using detailed data on every building footprint in the country as input, we instead apply the algorithm on a much larger spatial scale – i.e., to the entire SSA region – using constrained WorldPop data for 250 m\*250 m grid cells as input. We also modify the algorithm to exclude deserts from the definition of a country’s potentially inhabitable area. While excluding deserts is not important for France, it is for SSA countries such as Mauritania, Niger, Mali, and Chad. For all four of these countries, more than 50 percent of their land areas are desert.<sup>10</sup> Not excluding deserts would mechanically decrease the counterfactual density and make these countries appear more urbanized overall than they really are. Our exclusion of deserts, and, more generally, of uninhabitable areas, also distinguishes our application of the “dartboard” algorithm from the version applied by Henderson et al. (2022) in their analysis of SSA urbanization dynamics.<sup>11</sup>

Another notable advantage of the “dartboard” algorithm over other approaches for delineating urban areas is that the counterfactual density threshold to be considered significantly urban is specific to each cell and, most importantly, relative to a country’s own overall density instead of being common to all countries. Therefore, the approach much better discriminates urban areas. This is especially so for either very densely populated countries, which appear as very highly urbanized, or very sparsely populated countries, which appear little urbanized partly by construction in the degree of urbanization and similar approaches.

The third, and final, literature to which our paper contributes is the descriptive literature on the structures of both urban systems and the internal structures of urban areas. This includes the extensive

---

<sup>8</sup> The degree of urbanization uses a variety of such thresholds to distinguish different types of urban area. In “level 1” of the approach, a distinction is drawn between urban areas (“towns and suburbs”) that satisfy an overall population threshold of 5,000 and a population density threshold of 300 people per km<sup>2</sup>, and urban centers (“cities”) that satisfy thresholds of 50,000 people and 1,500 people per km<sup>2</sup> for overall population and population density respectively (Dijkstra et al. 2021).

<sup>9</sup> As mentioned above, the Africapolis data utilized by OECD/SWAC (2020) uses a population threshold of 10,000 to identify urban areas.

<sup>10</sup> The dartboard algorithm also excludes water bodies and both areas at extreme elevations and which have extreme slopes from the definition of potentially inhabitable area.

<sup>11</sup> Henderson et al. (2022) also impose thresholds that are constant over time, fixing them to those of the last year of observation, 2014, in their data. Given that built-up area has generally expanded overtime, this mechanically increases the number of settlements that are identified over time. The full dartboard approach would have allowed for an assessment relative to each year’s overall development.

literature on Zipf's law - see, among many others, Gabaix (1999), Eeckhout (2004), Dobkins and Ioannides (2001), Black and Henderson (2003), Duranton (2007), Desmet and Rappaport (2017), Bosker and Buringh (2017), de Bellefon et al. (2021), Jedwab and Storeygard (2021), Düben and Krause (2021), and Henderson et al. (2022). It also includes papers which estimate population density gradients for cities and empirically analyze the degree to which urban areas are mono- versus polycentric and the number of (sub-)centers that any given urban area has (see McDonald 1987; Giuliano and Small 1991; McMillen 2001; Zhou, Hubacek, and Roberts 2015; Ellis and Roberts 2016; Goswami and Lall 2016; Roberts 2018; De Bellefon et al. 2021).

The structure of the rest of the paper is as follows. Section 2 describes the data and methodology – i.e., the modified “dartboard” algorithm – that we apply to delineate urban areas across SSA. Section 3 presents descriptive analysis of overall urbanization patterns across SSA countries, which includes analysis of Zipf's law. Section 4 examines how these patterns correlate with both a country's population size and level of development. Section 5 zooms in to provide an overview of patterns relating to the internal structures of cities. Section 6 briefly assesses the sensitivity of our results to the alternative use of the degree of urbanization approach to delineating urban areas, as well as alternative data choices. Section 7 concludes.

## 2. Methodology and Data

### 2.1. Dartboard Algorithm

The methodology that we apply to identify and delineate urban areas across 47 SSA countries is the de Bellefon et al. (2021) “dartboard” algorithm. Urban areas are identified as statistically significant peaks of (slightly smoothed) population density by comparing the actual spatial distribution of population across 250 m\*250 m grid cells within each individual country against a counterfactual population density distribution for that country that is derived under the assumption of randomness. This implies that urban areas are based on country-specific, endogenously derived, population density thresholds. Because, *inter alia*, some cells are not considered inhabitable, these thresholds are furthermore specific to each cell. One notable advantage of such a relative approach is that it avoids a country being classified as almost entirely urban simply because it has a high average population density, which can be a problem for absolute threshold approaches (Bosker, Park and Roberts 2021).

Our application of the dartboard algorithm proceeds in four steps. In the first step, a country's gridded population at the 250 m\*250 m resolution is smoothed using a bi-square kernel with a bandwidth of 10 grid cells.<sup>12</sup> Following this, in the second step, 3,000 random reshuffles of a country's (smoothed) populated grid cells are performed over the sub-set of its grid cells which are potentially habitable. This leads to a counterfactual population density distribution for the country under the assumption of randomness. In the third step, a country's actual (smoothed) population distribution is compared to its counterfactual distribution and grid cells that have an actual population that is above the 95<sup>th</sup> percentile of the counterfactual distribution are classed as urban cells.<sup>13</sup> Finally, in the fourth step, contiguous sets of urban cells are identified and delineated as urban areas.<sup>14, 15</sup>

While the above four steps allow us to identify and delineate a country's urban areas, they do not allow us to distinguish between different types of urban areas. We therefore proceed, again for each country,

---

<sup>12</sup> Results are robust to the exact choice of smoothing bandwidth – for example, the urban delineations derived using a bandwidth of 5 are very similar to those derived using a bandwidth of 10. Results using bandwidths other than 10 are available on request.

<sup>13</sup> While the 95<sup>th</sup> percentile may be considered an arbitrary choice, it is, nevertheless, less arbitrary than the choices of population (density) thresholds that characterize absolute threshold approaches to delineating urban areas. This is because of the universality of use of the 5 percent level as a criterion for statistical significance.

<sup>14</sup> Contiguity is defined as Queen's contiguity (cells share common sides or corners).

<sup>15</sup> Two urban cells are also identified as belonging to the same urban area when they are separated by no more than one non-urban cell or by no more than two non-urban cells when one of those cells is uninhabitable. This allows for the filling of “holes” within urban areas that arise because of, for example, water bodies or large green areas.

to perform a second random reshuffling of all grid cells that belong to urban areas to generate a counterfactual population density distribution within those areas. Contiguous urban area grid cells that have an actual population that is above the 95<sup>th</sup> percentile of this second counterfactual distribution are “urban cores”. Any given urban area may have either zero or one or more cores. Urban areas that possess one or more cores are “cities”, while urban areas that lack a core are “towns”.

In applying the above methodology, the criteria used to designate grid cells as potentially habitable are important. Following de Bellefon et al. (2021), we define cells as potentially habitable if they do not: (i) intersect with a water body; (ii) are not at high altitude, defined here as below the 99<sup>th</sup> percentile of grid cells with at least one inhabitant in terms of altitude; and (iii) are not on a very steep slope, defined here as below the 99<sup>th</sup> percentile of grid cells with at least one inhabitant in terms of steepness. We also define as uninhabitable those grid cells which intersect with a desert. We do this for the 14 SSA countries in our sample which have more than 0.5 percent of their land area covered by desert.<sup>16</sup>

## 2.2. Data

The main data that we use as input into the dartboard algorithm is constrained WorldPop gridded population data from the University of Southampton’s WorldPop group for the year 2015.<sup>17</sup> This data has an underlying resolution of approximately 100 m\*100 m, but to avoid excessive compute times, we aggregate the data into 250 m\*250 m grid cells. This aggregation also ensures compatibility with other spatial layers that we use to classify cells as potentially habitable. As with other “top down” gridded population data sets, the constrained WorldPop algorithm derives a grid by taking population data from national censuses for larger sub-national administrative units and distributing this across the cells that fall within each unit. In the case of constrained WorldPop, population is spread unevenly across cells within a given admin unit based on weights derived from a Machine Learning (ML) model (Stevens et al. 2015), only allocating population to cells that are built-up. In turn, the classification of cells as built-up is based on the extremely detailed map of all building footprints in SSA that was derived under the Gates Foundation sponsored Digitize Africa project.<sup>18</sup> The quality and resolution of this building footprint data is comparable to, if not better than, the building footprint data that de Bellefon et al. (2021) use for France in their original application of the dartboard algorithm.

The spatial data used to define a country’s potentially habitable areas come from two sources. That used to identify water bodies and deserts come from the European Space Agency’s GlobCover product (Bontemps et al. 2013).<sup>19</sup> Meanwhile, the data used to measure elevation and slope come from SRTM30, which is a near-global digital elevation model (DEM) comprising a combination of data from the Space Shuttle Radar Topography Mission flown in February 2000 and the U.S. Geological Survey’s GTOPO30 data set.<sup>20</sup>

## 3. Descriptive Analysis of Overall Urbanization Patterns

### 3.1. Aggregate differences across countries

Table 1 presents summary statistics that describe key urbanization and other characteristics across SSA countries, while table 2 identifies the countries that have the three largest and smallest values for each variable presented in table 1. Before commenting on the urbanization characteristics, it is worth noting that SSA countries vary considerably in population, area, and average (overall) population density. Hence, population at the 90<sup>th</sup> percentile of the distribution across countries is 46.8 times that at the 10<sup>th</sup> percentile, while the corresponding ratios for land area and population density are 273.5 and 22.4 (table 1). The particularly large variation for area reflects the fact that SSA consists of both geographically

---

<sup>16</sup> The 14 countries are: Botswana, Cabo Verde, Chad, Eritrea, Ethiopia, Kenya, Mali, Mauritania, Mozambique, Namibia, Niger, Somalia, South Africa, and Eswatini.

<sup>17</sup> We use the “UN adjusted” version of this data in which the national population counts from summing across grid cells match national population counts as reported by the United Nations.

<sup>18</sup> See <https://ui.adsabs.harvard.edu/abs/2019AGUFMIN11D0688H/abstract>.

<sup>19</sup> [http://due.esrin.esa.int/page\\_globcover.php](http://due.esrin.esa.int/page_globcover.php).

<sup>20</sup> See [https://icesat.gsfc.nasa.gov/icesat/tools/SRTM30\\_Documentation.html](https://icesat.gsfc.nasa.gov/icesat/tools/SRTM30_Documentation.html) for more information.



large countries such as the Democratic Republic of Congo, Sudan and Chad, and tiny island nations such as Comoros, São Tomé and Príncipe, and Seychelles (table 2). Meanwhile, we would expect the large variation in average population density to, almost mechanically, translate into higher urban shares of the population under an absolute approach to defining urban areas such as the degree of urbanization. However, given that it is a relative approach, these variations in average absolute population density will have no *mechanical* effect on urban shares of the population as calculated using the dartboard approach.<sup>21</sup>

**Table 1: Distribution of urbanization characteristics over 47 SSA countries**

Variable	Unit	Mean	Std. Error	Percentile				
				P10	P25	P50	P75	P90
Population ( <i>mil.</i> )	Country	21.0	32.0	1.1	2.5	11.5	23.9	51.5
Area ( <i>'000s km<sup>2</sup></i> )	Country	517.0	541.1	4.6	58.0	325.9	831.1	1,258.0
Pop. density ( <i>per km<sup>2</sup></i> )	Country	93.0	121.0	9.0	20.0	51.0	115.0	202.0
Official urban (%)	Country	44.0	17.4	20.5	29.6	43.4	56.1	67.8
Pop. share (%)	Urban Areas	73.8	14.9	61.0	68.0	75.7	83.0	88.6
	Cities	55.1	14.1	38.6	47.4	55.4	64.9	74.1
	Towns	18.7	6.9	10.7	14.8	18.2	23.9	26.4
	Urban Cores	38.4	12.9	22.1	31.2	38.7	44.6	56.0
Area share (%)	Urban Areas	5.8	3.5	2.0	3.5	4.7	7.9	11.2
	Cities	3.3	2.7	0.9	1.4	2.0	4.4	8.1
	Towns	2.5	1.3	0.9	1.6	2.4	3.0	4.0
	Urban Cores	0.7	0.6	0.2	0.3	0.5	0.9	1.3
Pop. density ( <i>per km<sup>2</sup></i> )	Urban Areas	913.0	803.0	277.0	405.0	638.0	1,249.0	1,525.0
	Cities	1,374.0	1,202.0	339.0	588.0	1,131.0	1,727.0	2,687.0
	Towns	535.0	560.0	147.0	195.0	370.0	695.0	1,083.0
	Urban Cores	3,959.0	2,480.0	1,171.0	2,383.0	3,632.0	5,387.0	6,070.0
Area elasticity	Urban Areas	0.90	0.07	0.82	0.85	0.92	0.95	0.97
	Cities	0.86	0.13	0.70	0.77	0.87	0.91	0.96
	Towns	0.98	0.07	0.87	0.94	0.99	1.02	1.05
	Urban Cores	0.93	0.14	0.79	0.84	0.95	0.99	1.02
Built-up elasticity	Urban Areas	1.23	0.35	0.87	1.01	1.22	1.38	1.52
	Cities	1.27	0.31	1.04	1.09	1.21	1.41	1.53
	Towns	1.06	0.39	0.73	0.84	1.06	1.29	1.50
	Urban Cores	1.43	0.53	1.07	1.21	1.43	1.51	1.61

**Notes:** Country, Urban areas, Cities, Towns, Urban cores indicate the geographical level for which the statistics are computed. *Pop. share* (*Area share, resp.*): Population (area, resp.) share of cities, urban areas, and urban cores within the country. *Pop. density*: Average population density. *Area elasticity*: Elasticity of area with respect to population. *Built-up elasticity*: Elasticity of built-up with respect to population. Statistics are computed over the cross-section of 47 SSA countries. PX corresponds to the X<sup>th</sup> centile of the distribution.

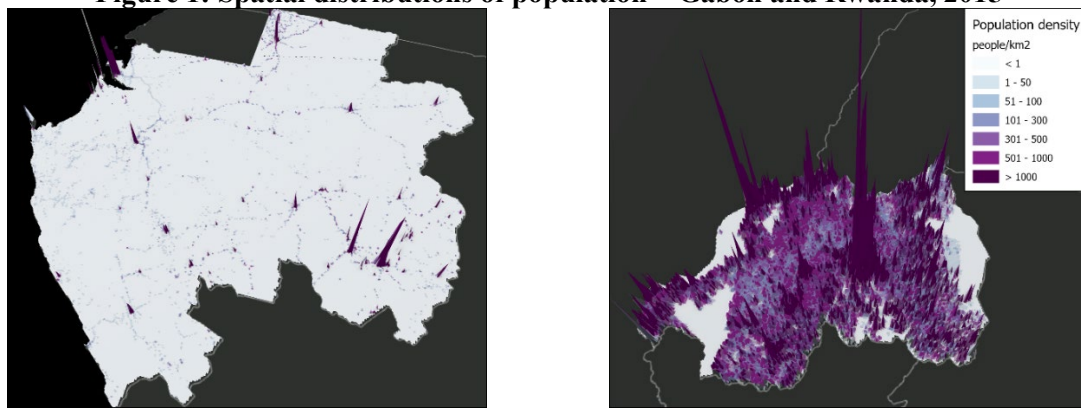
Turning to the urbanization characteristics which are derived from the dartboard approach, both tables 1 and 2 present statistics for four types of urban area. **Urban areas** refer to all areas identified as urban (i.e., all areas which consist of contiguous sets of cells whose population is above the 95 percentile of the counterfactual distribution); **Cities** refer to all urban areas that possess at least one **core**; and **Towns** to urban areas that do not possess a core. Cities may also be thought of as representing a stricter definition of urban than towns. As can be seen from table 1, countries are, on average, much more urbanized than official statistics would have us believe – across the 47 SSA countries, the mean

<sup>21</sup> Section 4 examines the correlation between a country's urban share of the population, as well as other key urbanization characteristics, and its population density and area.

(median) urban “dartboard” share of the population is 73.8 percent (75.7 percent), which is around 30 percentage points higher than the mean (median) official share. Even if we restrict ourselves just to cities, SSA countries appear more urbanized than official statistics suggest. On average, just over 55 percent of a SSA country’s population lives in cities according to the dartboard approach, while official statistics imply that just less than 45 percent live in urban areas on average. In cities, most people, on average across countries, live in core areas. This follows from the fact that the mean share of a country’s population that lives in cores is 38.4 percent. Cores are more densely populated than the cities of which they are a part, which, in turn, are more densely populated than towns. The mean population density of towns, in turn, exceeds that for countries overall.

Around these higher than official average levels of urbanization, however, there is also considerable variation across countries. This is the case regardless of whether we compare urban population shares at the 90<sup>th</sup> percentile with those at the 10<sup>th</sup> percentile (table 1) or the countries with the three largest urban population shares with the countries with the three smallest (table 2). Thus, for urban areas, population shares range from 61.0 percent at the 10<sup>th</sup> percentile to 88.6 percent at the 90<sup>th</sup> percentile. If we focus just on cities, the inter-decile range in population shares is even larger – from 38.6 percent to 74.1 percent. Again, it is important to note that this variation, which is also mirrored in the variation in the share of a country’s land area that its urban areas and cities occupy, is not driven by differences in average population densities across countries given the relative nature of the dartboard approach. Rather, the variation reflects differences in the spatial concentration of a country’s own population or, to put it another way, the unevenness of the spatial distribution of a country’s population. Hence, from table 2, Gabon has a particularly uneven spatial distribution of population, while São Tomé and Príncipe, Burundi and even Rwanda have more even distributions (figure 1). The population densities of different types of urban area also vary widely across countries. In some cases, such as Namibia, Botswana, and South Sudan, average population densities are quite low – below 300 people per km<sup>2</sup> – even for cities (table 2).

**Figure 1: Spatial distributions of population – Gabon and Rwanda, 2015**



**(a) Gabon**

**(b) Rwanda**

**Notes:** Maps constructed using constrained WorldPop gridded population data.

**Table 2: Highest and lowest urbanization characteristics**

Variable	Unit	Largest	2nd Largest	3rd Largest	3rd Lowest	2nd Lowest	Lowest
Population ( <i>mil.</i> )	Country	Nigeria (185.3)	Ethiopia (90.9)	Congo, Dem. Rep. (89.1)	Cabo Verde (0.5)	São Tomé Príncipe (0.2)	Seychelles (0.1)
Area ( <i>'000s km<sup>2</sup></i> )	Country	Congo, Dem. Rep. (2,350.1)	Sudan (1,859.4)	Chad (1,275.0)	Comoros (1.9)	São Tomé Príncipe (1.1)	Seychelles (0.8)
Pop. density ( <i>per km<sup>2</sup></i> )	Country	Mauritius (574)	Rwanda (442)	Comoros (398)	Botswana (4)	Mauritania (4)	Namibia (3)
Official urban (%)	Country	Gabon (90.4)	São Tomé Príncipe (75.1)	Botswana (71.6)	Rwanda (17.6)	Niger (16.8)	Burundi (14.1)
Pop. share (%)	Urban Areas	Gabon (94.4)	Congo Brazzaville (94.0)	Central African Republic (91.6)	Rwanda (50.0)	Burundi (42.2)	São Tomé Príncipe (5.2)
	Cities	Gabon (79.4)	Kenya (78.6)	Botswana (76.0)	Rwanda (34.1)	Burundi (30.6)	São Tomé Príncipe (4.5)
	Towns	Mali (35.1)	Comoros (31.6)	Côte d'Ivoire (28.2)	Botswana (8.8)	Kenya (4.8)	São Tomé Príncipe (0.7)
	Urban Cores	Gabon (70.2)	Congo Brazzaville (67.6)	Botswana (60.8)	Rwanda (16.6)	Burundi (15.9)	São Tomé Príncipe (2.5)
Area share (%)	Urban Areas	Mauritius (14.6)	Rwanda (14.5)	Eswatini (13.8)	Botswana (1.5)	Gabon (1.4)	Mauritania (0.8)
	Cities	Mauritius (10.8)	Rwanda (8.6)	Uganda (8.4)	Congo Brazzaville (0.5)	Mali (0.5)	Mauritania (0.2)
	Towns	Eswatini (6.0)	Rwanda (5.9)	Burkina Faso (5.0)	Mauritania (0.6)	Botswana (0.5)	São Tomé Príncipe (0.3)
	Urban Cores	Mauritius (3.4)	Seychelles (1.9)	Rwanda (1.7)	Mali (0.1)	Namibia (0.1)	Mauritania (0.1)
Pop. density ( <i>per km<sup>2</sup></i> )	Urban Areas	Comoros (4,561)	Mauritius (3,024)	Nigeria (2,375)	Chad (246)	Botswana (206)	Namibia (64)
	Cities	Comoros (7,498)	Mauritius (3,288)	Nigeria (2,993)	South Sudan (296)	Botswana (270)	Namibia (84)
	Towns	Comoros (3,162)	Mauritius (2,263)	Nigeria (1,318)	Central African Republic (128)	Botswana (68)	Namibia (24)
	Urban Cores	Comoros (15,246)	Nigeria (8,241)	Senegal (7,198)	South Sudan (1,046)	Eswatini (983)	Botswana (875)
Area elasticity	Urban Areas	Sierra Leone (1.01)	São Tomé Príncipe (1.00)	Rwanda (0.99)	Gabon (0.75)	Central African Republic (0.73)	Botswana (0.72)
	Cities	São Tomé Príncipe (1.47)	Eswatini (0.99)	Rwanda (0.98)	Congo Brazzaville (0.70)	Mali (0.70)	Gabon (0.68)
	Towns	Sierra Leone (1.16)	Côte d'Ivoire (1.09)	Gambia, The (1.05)	Central African Republic (0.84)	Congo Brazzaville (0.84)	Botswana (0.81)
	Urban Cores	São Tomé Príncipe (1.66)	Mali (1.04)	Chad (1.03)	Ethiopia (0.78)	Uganda (0.78)	Lesotho (0.73)
Built-up elasticity	Urban Areas	São Tomé Príncipe (2.93)	Gambia, The (1.74)	Côte d'Ivoire (1.58)	Mauritania (0.77)	Comoros (0.73)	Cabo Verde (0.58)
	Cities	São Tomé Príncipe (2.93)	Cabo Verde (1.74)	Liberia (1.64)	Madagascar (0.96)	Eswatini (0.92)	South Sudan (0.81)
	Towns	Gambia, The (1.91)	Côte d'Ivoire (1.71)	Ghana (1.63)	Mauritania (0.18)	Cabo Verde (0.14)	Eritrea (-0.01)
	Urban Cores	São Tomé Príncipe (4.64)	Eritrea (1.80)	Mali (1.70)	Mauritius (1.02)	Gambia, The (1.01)	Cabo Verde (0.58)

**Notes:** The table lists the top 3 and bottom 3 countries for each variable with the value taken for the country in brackets. The definition of variables is the same as in Table 1.

Finally, tables 1 and 2 present summary statistics for two estimated elasticities – the elasticity of *overall* area with respect to population across different types of urban area and the elasticity of *built-up* area with respect to population across different types of urban area, where a settlement’s built-up area may be less than its overall area due to open spaces. For the overall area elasticities, these are estimated to be less than 1 for most countries for all types of urban area. This implies that a doubling of an urban area’s population is associated with a less than doubling of its overall area, hence a higher density. However, for the built-up area elasticities, these are estimated to be larger than 1 for a large majority of countries for all types of urban area. Indeed, the median (mean) estimated elasticity of built-up area with respect to population is 1.22 (1.23) for all urban areas and 1.21 (1.27) for cities. Hence, for the median country, a doubling of an urban area or city’s population is associated with a 121-122 percent increase in its built-up area. The finding of estimated built-up area elasticities that exceed 1 is in marked contrast with results for developed countries and cities globally. Ahfledt and Pietrostefani (2019) estimate an elasticity of approximately 0.60 for a large global sample of cities, while de Bellefon et al. (2021) estimate an elasticity of 0.84 for all French urban areas, as defined using the dartboard algorithm, with a population greater than 100 and 0.89 for the 500 largest French urban areas.

One possible explanation for the difference in our estimated built-up area elasticities for SSA countries and those from the literature for developed countries is related to possible differences in the costs of vertical construction, and the ability of households and firms to cover those costs, in SSA versus developed countries.<sup>22</sup> It has been argued by, for instance, Lall et al. (2021) and Mukim and Roberts (2023), that dysfunctional land markets and, in some cases, misguided planning restrictions – for example, excessively stringent building height restrictions and floor area ratios – inflate the costs of vertical development in developing country cities. At the same time, because productivity and hence income levels are much lower in developing country cities, households and firms are less able to cover the costs of vertical construction (i.e., the development of mid- and high-rise buildings), which is much more capital intensive than low-rise construction that leads to more horizontal development of an urban area. For these reasons, Lall et al. (2021) describe cities in SSA and many other developing countries as expanding like “pancakes” rather than developing as “pyramids”, in which case horizontal expansion of a city is accompanied by vertical layering.

An alternative, not necessarily mutually exclusive, explanation for the difference in our estimated elasticities for SSA countries (greater than 1) vis-à-vis the estimates (less than 1) for developed countries relates to the potential composition of built-up area. Hence, built-up area includes not only residential built-up area, but also built-up area associated with industrial and commercial uses. If industry is more intensive in its use of land than housing and more populous cities in SSA countries dedicate a relatively larger share of their land to industrial than to residential uses, this could help explain the high estimated elasticities for these countries. This is especially the case if such compositional changes in built-up area across urban areas of different size are less evident in developed countries, as we might expect given both their more service-oriented economies and the tendency of manufacturing to gravitate towards smaller cities, where land tends to be cheaper, in these countries (World Bank 2009).<sup>23</sup>

### **3.2. Spatial concentration across urban areas**

Having described aggregate differences in basic urbanization characteristics across countries, we now turn to the description of urban *area* characteristics and the spatial concentration of population across urban areas. This includes a focus on characteristics of the size-distributions of urban areas, as captured both by estimated Zipf’s coefficients and the share of a country’s population that lives in its largest urban area, which is a measure of urban primacy.

---

<sup>22</sup> This includes both the absolute costs of vertical construction and the costs relative to the price of land. In the case of illegal squatter development, the price of land to the household occupying it is zero.

<sup>23</sup> The fact that estimated elasticities of overall area with respect to population tend to be less than 1 for most SSA countries while estimated elasticities of built-up area tend to be more than 1 suggests that larger SSA cities tend to have more in-fill development.

**Table 3: Distribution of city characteristics and concentration between cities across 47 SSA countries**

				Percentiles				
	Units	Mean	Std. Error	P10	P25	P50	P75	P90
No. units	Urban Areas	4,265	4,579	46	700	2,672	5,688	10,638
	Cities	365	459	8	60	172	472	1,271
	Towns	3,900	4,176	36	662	2,535	5,506	9,447
	Urban Cores	542	720	12	66	226	774	1,892
Average pop.	Urban Areas	4,657	4,194	1,504	1,861	2,684	6,562	10,362
	Cities	41,807	31,018	10,128	16,298	35,034	63,701	89,660
	Towns	1,244	1,239	386	505	984	1,400	2,359
	Urban Cores	21,588	16,856	4,928	7,963	17,037	32,079	45,216
Average area	Urban Areas	5.5	3.1	2.7	3.6	5.1	6.1	9
	Cities	37.5	30.4	13.4	18.7	29.7	47.6	55
	Towns	2.5	0.7	1.8	2.2	2.5	2.8	3.5
	Urban Cores	5.3	3.3	2.5	3.1	4.2	6.4	9.1
Average den.	Urban Areas	612	605	178	205	454	806	1,279
	Cities	1,078	1,042	353	552	854	1,301	1,739
	Towns	568	559	151	196	412	765	1,235
	Urban Cores	2,963	2,197	938	1,833	2,676	3,724	4,710
Pop. P75/P25	Urban Areas	5	2.6	3.1	3.7	4.4	5.1	7.4
	Cities	29.9	101.1	2.8	3.1	3.7	5	73.1
	Towns	3.8	1.2	2.7	3	3.6	4.2	5.4
	Urban Cores	9.5	20.1	3	3.7	4.4	6.1	18.5
Pop. P90/P10	Urban Areas	24.7	18.8	9.6	14.1	19.5	29.6	41.5
	Cities	114.4	238.1	8.4	11	15.6	30.7	387.4
	Towns	13.9	7.5	7.2	9.1	12.1	16.4	19.6
	Urban Cores	81.8	225.9	12	17.8	29.3	66.1	136.9
Den. P75/P25	Urban Areas	1.5	0.2	1.3	1.4	1.4	1.5	1.7
	Cities	1.6	0.4	1.3	1.4	1.5	1.7	2
	Towns	1.4	0.2	1.3	1.3	1.4	1.4	1.6
	Urban Cores	1.4	0.3	1.2	1.3	1.3	1.4	1.6
Den. P90/P10	Urban Areas	2.4	0.6	1.8	2.1	2.3	2.5	2.9
	Cities	2.9	1.6	1.9	2.1	2.4	3	4.1
	Towns	2	0.3	1.7	1.9	2	2.2	2.4
	Urban Cores	2.2	0.9	1.5	1.7	1.9	2.4	3.7
Share largest	Urban Areas	21.3	12.3	7.9	11.7	18.7	29	36.2
	Cities	21.3	12.3	7.9	11.7	18.7	29	36.2
	Towns	0.7	2.5	0.1	0.1	0.2	0.4	1.2
	Urban Cores	15.5	8.6	5.9	9	14.2	19.7	28.2
Zipf's coef.	Urban Areas	-0.8	0.2	-1	-0.9	-0.8	-0.7	-0.6
	Cities	-0.8	0.2	-1.1	-0.9	-0.8	-0.8	-0.4
	Towns	-1.2	0.3	-1.4	-1.3	-1.1	-1	-0.9
	Urban Cores	-0.9	0.1	-1.1	-0.9	-0.9	-0.8	-0.7

**Notes:** *No. units*: Number of units delineated. *Average pop.*: Average population of the units, in persons. *Average area*: Average area of the units, in km<sup>2</sup>. *Pop. P75/P25*: Inter-quartile ratio for population. *Pop. P90/P10*: Inter-decile ratio for population. *Share largest*: Population share of the largest unit. *Zipf's coef.*: Zipf's coefficient. Statistics are computed over the cross-section of 47 SSA countries, except for the Zipf's Law coefficient for which Seychelles is excluded due to its too low number of cities (3). PX corresponds to the X<sup>th</sup> percentile of the distribution.

On average, an SSA country's number of urban areas is large – the mean number of urban areas per country is 4,265 (table 3). This partly reflects the fact that, unlike many other approaches, the dartboard approach

imposes no minimum population threshold on the definition of urban areas. Hence, at the bottom end of the population and area distributions, urban areas can be very small. At the 10<sup>th</sup> percentile, the population of an urban area is just over 1,500, while area is 2.7 km<sup>2</sup>. In Namibia, the average population of urban areas is just 447, while in São Tomé and Príncipe, the average area is just 2.2 km<sup>2</sup> (table 4). If we adopt a stricter definition of urban by focusing just on cities, then, unsurprisingly, the numbers of cities identified are much smaller with population sizes and areas that are, on average, much larger. The mean number of cities per country is therefore 365, which is less than one-tenth the mean number of urban areas per country. Meanwhile, mean city population is 41,807 and mean city area is 37.5 km<sup>2</sup> (table 3). Irrespective of whether we consider all urban areas or just cities, the mean number of units per country and the mean population are noticeably larger than the corresponding medians, implying a positive skew to the distributions of these variables. Interestingly, the mean (median) number of cores per country exceeds the mean (median) number of cities, implying that the average city tends to have more than one core. As well as being larger than urban areas in both population and area, cities are also more densely populated (table 3).

Once again, we also observe considerable variation in the data around their averages. This is particularly true for the populations of cities – the inter-quartile ratio for city population across all SSA cities is 29.9, while the inter-decile ratio is 114.4. The equivalent ratios for population density are much smaller – 1.6 and 2.9 – but still large (table 3). While this variation describes the entire size distribution of cities across all 47 SSA countries, the variation is also very large *within* many individual SSA countries. In Lesotho, city population at the 90<sup>th</sup> percentile is more than 1,000 times that at the 10<sup>th</sup> percentile (table 4).

These differences in population between cities, whether for SSA as a whole or individual SSA countries, imply that population is not evenly distributed across cities, but, rather, is much more concentrated in some cities than in others. Related to this, the mean (median) share of a country’s population that lives in its largest city is 21.3 (18.7) percent, while more than three quarters of SSA countries have more than 11.7 percent of their populations living in their largest city (table 3). A quarter of SSA countries have their largest city accommodating close to a third of their population. In the most extreme case of the Seychelles, almost two-thirds of the population lives in the largest city (table 4). Despite this, levels of urban primacy in SSA do not appear, on average, to be higher than those in developed countries. For the 44 countries and territories that the World Bank classifies as high-income and provides the necessary data for, the mean (median) share of the population that lived in the largest city, as officially defined, in 2021 was 27.1 (21.1) percent.<sup>24</sup> This suggests that, if anything, levels of urban primacy in SSA are, on average, *lower* than for developed countries.<sup>25</sup>

More generally, city-size distributions for SSA countries conform quite well to Zipf’s law, which implies that a country’s largest city is twice as populous as its second largest, three times as populous as its third largest, and so on. The mean estimated Zipf’s coefficient for cities across SSA countries is -0.8, which is the same as for urban areas overall.<sup>26</sup> Again, this is close to estimates of Zipf’s coefficient for developed countries – de Bellon et al., for example, estimate 0.81 for French cities.<sup>27</sup> A Zipf coefficient of 0.8 implies slightly more spatial concentration across cities than is predicted by the law. In general, this is due to a long tail of small cities. As indicated in table 4, however, there are some countries that deviate radically from Zipf’s law. Lesotho and Eswatini have estimated Zipf’s coefficients for cities of -0.3, while Rwanda has an estimated coefficient of -0.4. This implies that these countries have greater disparities in population between cities than predicted by Zipf’s law.

---

<sup>24</sup> This calculation excludes both Hong Kong SAR, China, and Macao SAR, China.

<sup>25</sup> Based on the dartboard approach, 17 percent of France’s population lives in its largest city, Paris (de Bellefon et al. 2021).

<sup>26</sup> For each country, we estimate Zipf’s coefficient separately for urban areas, cities, towns, and cores using Gabaix and Ibragimov’s (2011) “Rank – ½” estimator – i.e., for each country, we estimate  $\log(Rank_i - 0.5) = a + b \cdot \log(Size_i) + \varepsilon_i$  where  $Rank_i$  denotes an urban area’s rank in the size distribution and  $Size_i$  its population.

<sup>27</sup> In two of the most comprehensive studies of Zipf’s law, Soo (2005) and Brakman, Garretsen and van Marrewijk (2009) report mean estimated Zipf’s coefficients of 0.90 and 0.88 respectively for large samples of countries when defining cities as “cities proper”. When cities are instead defined as urban agglomerations, they obtain mean estimated Zipf’s coefficients of 1.17 and 1.05 respectively.

**Table 4: Highest and lowest number, size, and size dispersion of cities**

Variable	Unit	Largest	2nd Largest	3rd Largest	3rd Lowest	2nd Lowest	Lowest
No units	Urban Areas	Congo, Dem. Rep. (22,195)	Ethiopia (13,104)	Sudan (12,490)	Cabo Verde (40)	São Tomé Príncipe (12)	Seychelles (11)
	Cities	Ethiopia (1,753)	Congo, Dem. Rep. (1,557)	Sudan (1,468)	Comoros (4)	São Tomé Príncipe (4)	Seychelles (3)
	Towns	Congo, Dem. Rep. (20,637)	Ethiopia (11,351)	Sudan (11,022)	Mauritius (33)	Seychelles (8)	São Tomé Príncipe (8)
	Urban Cores	Ethiopia (3,198)	Sudan (2,245)	Congo, Dem. Rep. (2,085)	Comoros (6)	Seychelles (4)	São Tomé Príncipe (4)
Average pop.	Urban Areas	Mauritius (20,995)	Kenya (15,579)	Comoros (12,586)	Mauritania (1,059)	São Tomé Príncipe (819)	Namibia (447)
	Cities	Gambia, The (143,268)	Togo (101,557)	Kenya (94,181)	Botswana (9,438)	Chad (6,162)	São Tomé Príncipe (2,127)
	Towns	Comoros (6,565)	Mauritius (5,648)	Rwanda (3,270)	Botswana (234)	São Tomé Príncipe (165)	Namibia (57)
	Urban Cores	Gambia, The (66,584)	Angola (61,420)	Togo (59,124)	Niger (4,487)	Chad (2,693)	São Tomé Príncipe (1,183)
Average area	Urban Areas	Kenya (21.2)	South Sudan (10.9)	Ethiopia (9.5)	Guinea-Bissau (2.3)	Sierra Leone (2.2)	São Tomé Príncipe (2.2)
	Cities	Namibia (183.9)	Kenya (122.3)	Zimbabwe (89.5)	Mali (9.6)	Comoros (8.9)	São Tomé Príncipe (5.7)
	Towns	South Africa (4.5)	Eswatini (4.0)	South Sudan (3.8)	Sierra Leone (1.5)	Seychelles (1.5)	São Tomé Príncipe (0.4)
	Urban Cores	Zimbabwe (19.7)	Angola (12.3)	Gambia, The (11.6)	Chad (2.3)	Niger (2.1)	São Tomé Príncipe (0.8)
Average den.	Urban Areas	Comoros (3,394)	Mauritius (2,367)	Nigeria (1,619)	Central African Republic (151)	Botswana (82)	Namibia (32)
	Cities	Comoros (6,953)	Nigeria (2,625)	Mauritius (2,556)	Eswatini (283)	Chad (283)	Botswana (213)
	Towns	Comoros (2,998)	Mauritius (2,293)	Nigeria (1,489)	Central African Republic (116)	Botswana (54)	Namibia (25)
	Urban Cores	Comoros (14,067)	Nigeria (7,675)	Côte d'Ivoire (5,444)	Botswana (711)	Eswatini (570)	Namibia (535)
Pop. P75/P25	Urban Areas	Botswana (17.9)	Seychelles (11.6)	Eswatini (9.2)	Liberia (2.7)	Guinea-Bissau (2.6)	Sierra Leone (2.5)
	Cities	Seychelles (653.3)	Lesotho (184.7)	Eswatini (178.6)	Côte d'Ivoire (2.5)	Mali (2.1)	Mauritius (2.0)
	Towns	Botswana (7.8)	Seychelles (7.3)	Eswatini (6.9)	Guinea-Bissau (2.5)	Sierra Leone (2.4)	São Tomé Príncipe (1.5)
	Urban Cores	Seychelles (136.6)	Cabo Verde (34.0)	Lesotho (33.4)	São Tomé Príncipe (2.5)	Mali (2.4)	Comoros (1.8)
Pop. P90/P10	Urban Areas	Botswana (120.3)	Eswatini (58.3)	South Africa (57.6)	Guinea-Bissau (9.3)	Sierra Leone (7.2)	Liberia (7.1)
	Cities	Lesotho (1,016.3)	Eswatini (880.6)	Uganda (730.7)	Côte d'Ivoire (7.8)	Niger (7.6)	Mali (5.2)
	Towns	Botswana (44.8)	Eswatini (33.8)	South Africa (28.5)	Liberia (6.1)	Sierra Leone (5.9)	São Tomé Príncipe (3.1)
	Urban Cores	Cabo Verde (1,547.5)	Seychelles (270.9)	Lesotho (226.1)	Niger (8.7)	Mali (8.3)	São Tomé Príncipe (4.1)
Den. P75/P25	Urban Areas	Botswana (2.4)	São Tomé Príncipe (2.2)	Central African Republic (1.9)	Eswatini (1.3)	Mauritius (1.3)	Seychelles (1.2)
	Cities	Namibia (3.5)	Zimbabwe (3.4)	Angola (2.0)	Ghana (1.3)	Mauritius (1.3)	Cabo Verde (1.3)

	Towns	São Tomé Príncipe (2.2)	Botswana (1.7)	Central African Republic (1.6)	Malawi (1.3)	Mauritius (1.2)	Seychelles (1.2)
	Urban Cores	Seychelles (2.8)	São Tomé Príncipe (2.3)	Lesotho (2.1)	Guinea (1.2)	Senegal (1.2)	Togo (1.2)
Den. P90/P10	Urban Areas	Botswana (5.3)	Central African Republic (3.6)	Kenya (3.1)	Malawi (1.8)	Mauritius (1.7)	Seychelles (1.4)
	Cities	Namibia (11.9)	Zimbabwe (6.4)	Gambia, The (4.6)	Cabo Verde (1.7)	Seychelles (1.6)	Ghana (1.6)
	Towns	Botswana (3.0)	Central African Republic (2.5)	Namibia (2.5)	Lesotho (1.7)	Malawi (1.6)	Mauritius (1.4)
	Urban Cores	Cabo Verde (6.7)	Rwanda (3.9)	Mauritius (3.8)	Benin (1.4)	Senegal (1.4)	Comoros (1.4)
	Urban Areas	Seychelles (65.9)	Gambia, The (50.0)	Mauritius (43.8)	Nigeria (6.7)	South Sudan (6.5)	São Tomé Príncipe (3.0)
Share largest	Cities	Seychelles (65.9)	Gambia, The (50.0)	Mauritius (43.8)	Nigeria (6.7)	South Sudan (6.5)	São Tomé Príncipe (3.0)
	Towns	Comoros (17.0)	Seychelles (2.8)	Cabo Verde (2.3)	Côte d'Ivoire (0.1)	Mali (0.1)	Nigeria (0.0)
	Urban Cores	Congo Brazzaville (35.2)	Gambia, The (32.5)	Gabon (31.5)	Ethiopia (3.8)	South Sudan (3.1)	São Tomé Príncipe (1.4)
	Urban Areas	Cabo Verde (-0.4)	Gambia, The (-0.5)	Lesotho (-0.5)	Chad (-1.1)	Niger (-1.3)	São Tomé Príncipe (-1.3)
Zipf's coef.	Cities	Lesotho (-0.3)	Eswatini (-0.3)	Rwanda (-0.4)	Côte d'Ivoire (-1.1)	Niger (-1.1)	Mali (-1.2)
	Towns	Botswana (-0.7)	Eswatini (-0.8)	Comoros (-0.8)	Liberia (-1.6)	Sierra Leone (-1.6)	São Tomé Príncipe (-2.5)
	Urban Cores	Botswana (-0.5)	Cabo Verde (-0.6)	Eswatini (-0.7)	Liberia (-1.1)	Sierra Leone (-1.1)	Mali (-1.2)
<b>Notes:</b> <i>No units:</i> Number of units delineated. <i>Average pop.:</i> Average population of the units, in persons. <i>Average area:</i> Average area of the units, in km <sup>2</sup> . <i>Pop. P75/P25:</i> Inter-quartile ratio for population. <i>Pop. P90/P10:</i> Inter-decile ratio for population. <i>Dens. P90/P10:</i> Inter-decile ratio for population density. <i>Share largest:</i> Population share of the largest unit. <i>Zipf's coef.:</i> Zipf's coefficient.							



#### 4. Correlates of Urbanization Characteristics across Countries

To gain further insights into how urbanization characteristics vary across SSA countries, we estimate, for each characteristic, the following three regressions:

$$UC_i = a + b.\ln(POP_i) + u_{1i} \quad [1]$$

$$UC_i = c + d.\ln(DEN_i) + e.\ln(AREA_i) + u_{2i} \quad [2]$$

$$UC_i = f + g.\ln(DEN_i) + h.\ln(AREA_i) + j.\ln(GDP\_PC_i) + u_{3i} \quad [3]$$

where  $UC_i$  denotes the urbanization characteristic in question for country  $i$ ,  $POP_i$  a country's population size,  $DEN_i$  a country's overall average population density,  $AREA_i$  a country's land area, and  $GDP\_PC_i$  a country's level of GDP per capita. Equation [1] relates the urbanization characteristic to a country's overall size, as measured by its population. Equation [2] then decomposes the "effect" of a country's population on the urbanization characteristic between the intensive (average overall population density) and extensive (area) margins of a country's size. Finally, equation [3] looks at the correlation between the urbanization characteristic and a country's GDP per capita while controlling for both the intensive and extensive margins of its size. We estimate all three equations as cross-sectional OLS regressions for 2015. Given that a country's urbanization characteristics may also potentially affect its GDP per capita, as well as other potential endogeneity problems, these regressions should not be interpreted as capturing causal effects, merely (partial) correlations in the data.

**Table 5: OLS regression results - relationship between country size and urbanization characteristics, cities only**

Dependent variable	(1)		(2)			(3)			
	Pop.	R2	Density	Area	R2	Density	Area	GDP per cap.	R2
Official urb.	-3.38**	0.10	-7.67***	-3.55**	0.18	-4.64**	-1.24	10.39***	0.41
Pop. share	2.26*	0.07	-3.03	2.04*	0.27	-0.67	3.85***	8.11***	0.48
Area share	-0.02	0.00	1.38***	0.04	0.39	1.52***	0.15	0.48	0.41
Pop. density	-94.33	0.02	449.27***	-72.15	0.31	497.71***	-35.03	166.49	0.32
Area elas.	-0.02	0.05	0.03*	-0.02	0.28	0.03	-0.02*	-0.02	0.29
Built-up elas.	-0.08***	0.17	-0.05	-0.08***	0.18	-0.06	-0.09***	-0.04	0.19
No. units	191.97***	0.44	149.42***	190.23***	0.46	140.02**	183.04***	-32.29	0.46
Average pop.	3895.02	0.04	15837.43***	4382.45*	0.25	19351.29***	7074.94***	12076.20**	0.35
Average area	3.13	0.03	-2.57	2.90	0.08	0.56	5.30*	10.75*	0.16
Pop. P75/P25	-28.67***	0.20	-15.90	-28.15***	0.23	-7.95	-22.06**	27.32	0.27
Pop. P90/P10	-36.31*	0.06	15.07	-34.21	0.13	19.40	-30.89	14.90	0.13
Den. P75/P25	-0.04	0.02	-0.17***	-0.04	0.16	-0.15**	-0.02	0.08	0.18
Den. P90/P10	-0.09	0.01	-0.65***	-0.11	0.18	-0.59**	-0.07	0.18	0.19
Share largest	-4.22***	0.30	-3.35**	-4.18***	0.31	-1.67	-2.89***	5.77***	0.45
Zipf's coef.	-0.03	0.05	0.03	-0.03	0.21	0.04	-0.03	0.02	0.21

**Notes:** Estimation results for equations [1]-[3] for different urbanization characteristics (listed in the notes of Tables 1 and 3). All regressions based on 47 countries except those for Zipf's coefficient, which exclude Seychelles due to it only having three cities.

Table 5 presents the results from estimating equations [1]-[3] for each urbanization characteristic. For brevity, the table includes only the results for cities – results for other types of urban area are presented in Annex Table A1. Several interesting findings regarding the correlates of SSA urbanization characteristics emerge. First, there is a strong and statistically significant negative relationship between an SSA country's overall population size and the share of its population that lives in officially defined urban areas – i.e., more populous SSA countries tend to have lower official urban population shares (col. [1]). This negative relationship between a country's official level of urbanization and its size exists at both the intensive and extensive margins but is stronger at the former margin (col. [2]). Controlling for size, a country's official urbanization level is then also strongly correlated with its GDP per capita, confirming the well-known existence of a positive relationship between a country's level of urbanization and its level of development (col. [3]). Including GDP per capita in the regression, does, however, lead to a loss of significance of area in the relationship with a country's official urbanization level. Overall, therefore, SSA countries that are more developed and less densely populated have higher official urbanization levels.

Controlling for size, the existence of a positive relationship between a country’s level of GDP per capita and its level of urbanization continues to hold when, instead of using official definitions of urban areas, we define urban areas as cities using the dartboard approach. However, the share of population that lives in cities tends to be significantly *larger* for more populous SSA countries, the opposite of what we found for official definitions of urban areas. This tendency for more populous SSA countries to have a larger share of their populations living in cities is driven by the extensive margin – i.e., by area rather than by population density. GDP per capita, area and overall average population density alone account for almost half of the variation in the shares of population that live in cities across SSA countries.

For the share of a country’s land area that is occupied by cities, the results are very different. Hence, neither a country’s level of development nor its area is correlated with this share (col. [3]). However, countries that are more densely populated on average tend to have larger shares of their areas dedicated to cities, suggesting that cities are more “sprawling” for these countries, perhaps because of stronger congestion forces that bid-up the price of land and push cities outward. Consistent with this, more densely populated countries also tend to have more densely populated cities, irrespective of whether we control for a country’s level of development (cols. [2] and [3]). Unsurprisingly, countries that are more populous at either the intensive or extensive margins tend to have more cities, although there is no correlation between a country’s level of development and its number of cities. This suggests that, as SSA countries become richer, they also become urbanized not because of the emergence of new cities, but because the existing set of cities become more populous. Consistent with this, both the average population and average area of a country’s cities are significantly positively correlated with its GDP per capita, controlling for country size (col. [3]).

In terms of the size distribution of cities, more developed and geographically smaller SSA countries tend to exhibit higher levels of urban primacy (i.e., have larger shares of their populations living in the largest city) (col. [3]). The positive relationship with GDP per capita, conditional on controlling for a country’s overall size, is consistent with SSA countries being on the upward sloping section of a spatial Kuznets curve in which population concentration at first increases with GDP per capita at low levels of development before starting to decline above some critical level of development. Such a relationship is, in turn, rationalized by the tendency of population and resources to first concentrate in a few cities as countries start to develop before later deconcentrating to secondary and other smaller cities as, *inter alia*, rising land costs and improved transportation networks incentivize relatively land intensive industries to deconcentrate to smaller cities (World Bank 2009). We also observe that more populous countries tend to have a smaller spread of population sizes across cities, regardless of whether we look at the inter-quartile or inter-decline ratio. There is, however, no correlation between a country’s estimated Zipf coefficient and any of the three country characteristics – average population density, area, or GDP per capita – examined.

## 5. Spatial Concentration within Cities

Having described patterns of urbanization across SSA countries, at both the aggregate and urban area levels in Sections 3 and 4, we now turn to describe features of the *internal structures* of cities and how these differ across SSA countries and cities. Given that comprehensive evidence is generally lacking on the internal structures of cities, especially for SSA, this arguably represents the most novel part of our analysis. Tables 6-8 present our results. Tables 6 and 7 provide descriptive evidence on population and built-up density gradients within cities respectively. Meanwhile, Table 8 provides descriptive evidence on the numbers of centers within cities. For brevity, we, again, only focus on cities in this section of the paper. This is because the internal structures of cities are likely to be more complicated than those of towns.<sup>28,29</sup>

### 5.1. City population density gradients

The first row of Table 6 presents summary statistics for triplets of  $R^2$  values for the following three sets of *grid cell level* regressions, which are estimated separately for each country, pooling all cities within a country (i.e., the first-row reports [ $R^2_{eqn.[4]}$ ,  $R^2_{eqn.[5]}$ ,  $R^2_{eqn.[6]}$ ):

$$\ln(DEN_i) = \alpha + \theta_{c(i)} + \varepsilon_{1i} \quad [4]$$

<sup>28</sup> Results have also been generated for towns and are available on request.

<sup>29</sup> Table A4 in the Annex reports results from OLS regressions which examine the correlation, across SSA countries, between average within-city characteristics and country size. We do not focus on these results in the main text because we lack clear theory on these relationships.

$$\ln(DEN_i) = \alpha + \theta_{c(i)} + \beta \ln(d_i) + \varepsilon_{2i} \quad [5]$$

$$\ln(DEN_i) = \alpha + \theta_{c(i)} + \beta_{c(i)} \ln(d_i) + \varepsilon_{3i} \quad [6]$$

where  $DEN_i$  denotes the average population density of grid cell  $i$ ,  $\theta_{c(i)}$  a city-specific fixed effect for city  $c$  where cell  $i$  is, and  $d_i$  the distance from the centroid of cell  $i$  to the city's barycenter of all its cells weighted by their populations (i.e., the center of the city's population "mass").<sup>30</sup> The difference between equations [5] and [6] is that while the former constrains the population density gradient to be the same for all cities within a country, the latter allows it to differ across cities within a country.

For any given country, the  $R^2$  from the estimation of equation [4] measures how much of the variation in population density across the grid cells that belong to its cities can be explained by differences in population density *between* its cities. Correspondingly,  $(1 - R^2)$  measures the variation that is attributable to cell level variation *within* cities. As can be seen from table 6, both the mean and median  $R^2$  across all SSA countries are 0.07. Hence, on average for SSA countries, variation across cities only accounts for 7 percent of the variation in population density across the cells that belong to those cities, implying that 93 percent of the variation occurs within cities. Even at the 90<sup>th</sup> percentile of the distribution of estimated  $R^2$  values, only 12 percent of the variation in population density across cells is explained by the variation across cities.

**Table 6: Distribution of within-city population density gradients over 43 SSA countries**

Row		Mean	Std. Error	P10	P25	P50	P75	P90
1	$R^2$ pop. [ $R^2_{eqn.[4]}$ , $R^2_{eqn.[5]}$ , $R^2_{eqn.[6]}$ ]	0.07/ 0.16/0.20	0.03/ 0.09/0.09	0.03/ 0.05/0.08	0.04/ 0.08/0.13	0.07/ 0.13/0.19	0.10/ 0.24/0.27	0.12/ 0.28/0.32
2	Pop. grad. sh. <0	0.81	0.18	0.54	0.74	0.88	0.93	0.95
3	Pop. grad. sh. ns	0.18	0.18	0.04	0.07	0.11	0.24	0.44
4	Pop. grad. sh. >0	0.01	0.02	0	0	0.01	0.02	0.03
5	Pop. grad. Av.	-0.89	0.5	-1.53	-1.17	-0.83	-0.49	-0.26
6	Pop. grad. P10	-2.96	0.79	-3.91	-3.63	-2.84	-2.39	-2.06
7	Pop. grad. P25	-2.4	0.72	-3.23	-2.98	-2.36	-1.81	-1.6
8	Pop. grad. P50	-1.73	0.64	-2.42	-2.11	-1.73	-1.28	-0.97
9	Pop. grad. P75	-1.11	0.48	-1.67	-1.44	-1.09	-0.79	-0.45
10	Pop. grad. P90	-0.61	0.38	-1.03	-0.8	-0.62	-0.41	-0.22

**Notes:** Row 1:  $R^2$  for the estimation of equations [4] – [6] for population density. *Pop. grad. sh. <0*: Share of cities with significant (10% threshold) negative population density gradient. *Pop. grad. sh. ns*: Share of cities with non-significant (10% threshold) population density gradient. *Pop. grad. sh. >0*: Share of cities with significant (10% threshold) positive population density gradient. *Pop. grad. Av.*: average (over all cities) population density gradient (when significantly (10% threshold) different from zero). *Pop. grad. PXX*: Population density gradient for the city at the XX<sup>th</sup> centile for the gradient value. Statistics are computed over the cross-section of 43 SSA countries, Comoros, Madagascar, Seychelles and São Tomé Príncipe being excluded for the within-city statistics given their to low number of cities or of centers in these cities PX corresponds to the X<sup>th</sup> percentile of the distribution.

By adding  $\ln(d_i)$  to equation [4], equation [5] allows a grid cell's average population density not only to vary across a country's cities, but also with the distance between that cell and the barycenter of the city of which it is a part. As can be seen, the mean  $R^2$  of 0.16 from the estimation of equation [5] for each country is more than double the mean  $R^2$  of 0.07 from the estimation of equation [4] for each country.<sup>31</sup> Hence, distance to a city's barycenter is an important source of population density variation at the grid cell level. The importance of distance further increases when we allow distance gradients to vary across a country's cities, as in equation [6]. This follows from the fact that both the mean and median  $R^2$  values from the estimation of equation [6] are greater than the corresponding  $R^2$  values from the estimation of equation [5]. At the 90<sup>th</sup> percentile across countries, the city-specific fixed effects and distance to a city's barycenter can, when allowing for heterogeneity in distance gradients across cities, account for almost a third of the variation in population density across individual grid cells. This increase in explanatory power is to be

<sup>30</sup> The concept of a city's barycenter should not be confused with the concept of city centers that is discussed in Section 5.3 below. Hence, while a city's barycenter is unique, a city may have one or more centers, and, under special circumstances, it may even have no statistically identifiable center at all.

<sup>31</sup> Given that equations [4]-[6] are estimated at the grid cell level for each country, each is based on an extremely large number of observations. As such, whether the  $R^2$  is adjusted or not for the number of degrees of freedom makes no difference to the discussion.

expected given that urban economic theory – i.e., the Alonso-Muth-Mills model – implies that a city’s population density gradient will depend on commuting costs to its center, which include both monetary costs and the opportunity cost of the time spent commuting (Alonso 1964; Mills 1967; Muth 1969). These costs are likely to vary across cities due to, for example, variations in the quality of transport infrastructure.

For most cities in most SSA countries, the population density gradients estimated from equation [5] are also negative (rows 2-4, Table 6). Hence, on average across countries, 81 percent of cities have density gradients that are significantly negative at the 10 percent level, while 18 percent have gradients that are either insignificantly different from zero or which cannot be estimated because the city is too small. It follows that only 1 percent of cities have gradients that are significantly positive. For these cities, population density tends to increase with distance to the barycenter, implying that the location of a city’s weighted population center does not coincide with its location of peak population density.

This dominance of negative density gradients, which is again consistent with the Alonso-Muth-Mills model, suggests that most SSA cities are monocentric. When population density gradients are constrained to be the same across all a country’s cities, as in the estimation of equation [4], the mean (median) estimated gradient across countries is -0.89 (-0.83) (row 5, Table 6). This implies that a doubling of distance to a city’s barycenter is associated with, on average, a 46 (44) percent decline in population density.<sup>32</sup> Across countries, the gradient, when constrained to be the same across cities, varies from -1.53 at the 10<sup>th</sup> percentile to -0.26 at the 90<sup>th</sup> percentile. Hence, there is significant variation in the estimated gradient across countries. As Table A2 in the Annex shows, the three countries that have the steepest density gradients are Mali (-2.26), Cote d’Ivoire (-2.06) and Congo Brazzaville (-1.74). Meanwhile, the three countries that have the shallowest gradients are Eswatini (-0.12), Malawi (-0.23), and South Sudan (-0.24). Within countries, when allowing for heterogeneity, mean density gradients vary from -2.96 at the 10<sup>th</sup> percentile to -0.61 at the 90<sup>th</sup> percentile, as can be seen by comparing the results in rows 6-10 of Table 6.

## 5.2. Built-up density gradients

Table 7 follows the same structure as Table 6 except that instead of reporting results for grid cell population densities within cities, it reports results for grid cell *built-up densities* within cities. These results are again obtained from estimating equations [4]-[6], except that, this time,  $DEN_i$  denotes a cell’s built-up density instead of its population density. As can be seen from row 1 of the table, on average across countries, much more of the cell-level variation in built-up density, compared to the cell-level variation in population density, within cities can be accounted for by differences between cities. Hence, the mean (median)  $R^2$  across countries for the regressions that only include city fixed effects is 0.29 (0.28). This is not to say, however, that a cell’s distance to a city’s barycenter is not important. Quite the opposite – when a cell’s built-up density is also allowed to depend on this distance, the estimated mean (median)  $R^2$  across countries increases to 0.47 (0.49). When built-up density gradients are further allowed to vary in slope across a country’s cities, the mean (median)  $R^2$  becomes 0.53 (0.54). Hence, 53-54 percent of the variation in built-up density across city cells can be “explained” by the combination of the distance to a city’s barycenter and average differences in built-up density across cities. At the 90<sup>th</sup> percentile across countries, these two factors account for 65 percent of the variation, while, even at the 10<sup>th</sup> percentile, they account for 40 percent.

---

<sup>32</sup>  $2^{-0.89} - 1 = 0.460$  and  $2^{-0.83} - 1 = 0.437$ .

**Table 7: Distribution within-city built-up density gradients over 43 SSA countries**

Row		Mean	Std. Error	P10	P25	P50	P75	P90
1	R <sup>2</sup> built-up [R <sup>2</sup> <sub>eqn.[4]</sub> , R <sup>2</sup> <sub>eqn.[5]</sub> , R <sup>2</sup> <sub>eqn.[6]</sub> ]	0.29/ 0.47/0.53	0.11/ 0.13/0.12	0.15/ 0.31/0.40	0.23/ 0.41/0.46	0.28/ 0.49/0.54	0.33/ 0.55/0.62	0.41/ 0.60/0.65
2	Built-up grad. sh. <0	0.76	0.22	0.45	0.56	0.82	0.95	0.97
3	Built-up grad. sh. ns	0.22	0.21	0.02	0.03	0.16	0.38	0.54
4	Built-up grad. sh. >0	0.02	0.02	0	0.01	0.01	0.02	0.03
5	Built-up grad. Av.	-1.35	0.38	-1.86	-1.67	-1.31	-1.1	-0.87
6	Built-up grad. P10	-2.45	0.39	-2.87	-2.66	-2.52	-2.26	-2.13
7	Built-up grad. P25	-1.94	0.3	-2.3	-2.1	-1.97	-1.82	-1.72
8	Built-up grad. P50	-1.55	0.26	-1.79	-1.71	-1.62	-1.43	-1.29
9	Built-up grad. P75	-1.14	0.23	-1.37	-1.34	-1.17	-1.01	-0.85
10	Built-up grad. P90	-0.73	0.26	-1.01	-0.95	-0.75	-0.54	-0.4

**Notes:** Row 1: R<sup>2</sup> for the estimation of equation [4] – [6] for built-up density. *Built-up grad. sh. <0*: Share of cities with significant (10% threshold) negative built-up density gradient. *Built-up grad. sh. ns*: Share of cities with non-significant (10% threshold) built-up density gradient. *Built-up grad. sh. >0*: Share of cities with significant (10% threshold) positive built-up density gradient. *Built-up grad. Av.*: average (over all cities) built-up density gradient (when significantly (10% threshold) different from zero). *Built-up grad. PXX*: Built-up density gradient for the city at the XX<sup>th</sup> centile for the gradient value. Statistics are computed over the cross-section of 43 SSA countries, Comoros, Madagascar, Seychelles and São Tomé Príncipe being excluded for the within-city statistics given their to low number of cities or of centers in these cities. PX corresponds to the X<sup>th</sup> percentile of the distribution.

As with population density, most cities within most countries have negative estimated built-up density gradients (rows 2-4, Table 7) with the mean (median) estimated gradient across countries being -1.35 (-1.31) when gradients are constrained to being the same for all cities within a country (row 5). Estimated built-up density gradients tend to be steeper than estimated population density gradients – i.e., built-up density declines faster with distance from the barycenter than does population density. One potential explanation for this is that city centers tend to have relatively more commercial and relatively less residential buildings than do more “suburban” areas of cities.

### 5.3. Numbers of (sub-)centers within cities

The above discussion of density gradients has focused on the concept of a city’s barycenter, which is its gridded population weighted center. A distinct, but related concept to this, is the concept of a city’s center as a significant peak of population density. If a city is monocentric and has a single such peak, then the location of this center will line-up well with its barycenter. However, if the city is polycentric and has several, geographically distinct, significant peaks of density then the barycenter may fall between those peaks/centers. This is analogous to, for example, the center of mass, i.e., the barycenter, of two roughly equally sized planets falling in-between those two planets. To identify a city’s center(s) – i.e., its significant peaks of population density – we perform 3,000 random reshuffles of its cells within the city. This yields a counterfactual distribution of population *within* the city under the assumption of randomness. A center is then a contiguous set of cells within the city whose population density is above the 95<sup>th</sup> percentile of the city’s counterfactual distribution.<sup>33</sup>

In the SSA context, it is important to note that a city center(s) identified in this way may or may not represent a commercial or business district. This is because the areas of greatest population density within a city may be slum neighborhoods. To empirically distinguish between a city’s business/commercial centers and its residential population centers would require spatially fine-grained data on employment, which is, unfortunately, not available, in addition to our current spatial data on population.

As noted in the previous section, the fact that most SSA cities have negative estimated population density gradients with respect to their barycenters suggests that the typical SSA city is monocentric. Consistent with this, Table 8 shows that the mean (median) number of centers per city across countries is 1.65 (1.51),

<sup>33</sup> It is important to note here the distinction between a city’s “center” and its “core.” As described in Section 2.1, “cores” are identified through a second-order random reshuffling of urban grid cells across *all urban areas* of the country and, by definition, all cities must have at least one core. By contrast, centers are identified by the random reshuffling of cells *within each individual city*.

while the mean (and median) share of cities that have a single center is 80 percent (rows 1 and 3 respectively). Moreover, in 90 percent of countries, 90 percent of cities have a single center (row 3).<sup>34</sup>

Notwithstanding the dominance of monocentric cities, however, the above results still imply that, on average, one-fifth of an SSA country's cities have multiple centers. For one-half of these cities, the number of centers is two (row 4, table 8). Thus, most cities are either monocentric or, if not, duo centric in terms of their spatial population distributions. The largest cities in SSA, however, have many centers. Hence, the mean number of centers in a country's largest city is 46, while the median is 4 (row 7). The fact that the mean is far higher than the median implies there is a strong positive skew to the distribution with a few countries having largest cities with very many centers. Thus, for 10 percent of SSA countries, the largest city has a remarkable 195 or more centers. As per the discussion above, many of these "centers" are likely to be densely populated slum areas, many of which may be quite small. As Table A3 in the Annex shows, Kenya, Ethiopia, and Niger are the three countries whose largest cities (Nairobi, Addis Ababa, and Niamey respectively) have the most centers.<sup>35</sup>

**Table 8: Distribution of city center characteristics over 43 SSA countries**

		Mean	Std. Error	P10	P25	P50	P75	P90
1	No. centers	1.65	0.58	1.11	1.21	1.51	1.81	2.22
2	Sh. 0 cen.	0.01	0.02	0	0	0	0.02	0.05
3	Sh. 1 cen.	0.8	0.09	0.69	0.74	0.8	0.85	0.9
4	Sh. 2 cen.	0.1	0.05	0.05	0.07	0.1	0.13	0.16
5	Sh. 3-5 cen.	0.06	0.04	0.01	0.03	0.06	0.08	0.13
6	Sh. >6 cen.	0.03	0.03	0	0	0.02	0.04	0.05
7	#cen. largest	45.98	112.29	1	2	4	14	195
8	Sh. pop. lar. cen.	0.68	0.08	0.57	0.61	0.67	0.73	0.77
9	Sh. bui. lar. cen.	0.57	0.07	0.49	0.52	0.57	0.62	0.65
10	Sh. area lar. cen.	0.26	0.08	0.21	0.22	0.24	0.27	0.38
11	lar/2nd pop.	12.86	15.79	3.96	5.24	6.54	12.37	35.29
12	lar/2nd bui.	1852.1	9456.56	13.56	25.76	54.98	331.01	1405.52
13	lar/2nd area	11.25	14.19	3.97	5.05	6.32	10.96	27.81
14	Zipf centers	-0.57	0.12	-0.72	-0.67	-0.58	-0.5	-0.4
15	Elas. #cen. pop.	0.2	0.12	0.04	0.1	0.17	0.29	0.35

**Notes:** *No. centers:* Number of centers per city. *Sh. 0 cen.:* Share of cities without center. *Sh. 1 cen.:* Share of cities with one center. *Sh. 2 cen.:* Share of cities with two centers. *Sh. 2-3 cen.:* Share of cities with two to five centers. *Sh. >6 cen.:* Share of cities with 6 or more centers. *#cen. largest:* Number of centers of the largest city. *Sh. pop. lar. cen.:* City population share of the largest center. *Sh. bui. lar. cen.:* City built share of the largest center. *Sh. area lar. cen.:* City area share of the largest center. *lar/2nd pop.:* Relative population of the largest over the second largest center. *lar/2nd bui.:* Relative built of the largest over the second largest center. *lar/2nd area:* Relative area of the largest over the second largest center. *Zipf centers:* Within city population Zipf's coefficient for city centers. *Elas. #cen. pop.:* elasticity of the number of centers in the city with respect to its population. Statistics are computed over the cross-section of 43 SSA countries, Comoros, Madagascar, Seychelles and São Tomé Príncipe being excluded for the within-city statistics given their low number of cities or of centers in these cities. PX corresponds to the X<sup>th</sup> percentile of the distribution.

On average across countries, a city's largest, and, typically, only, center, occupies 26 percent of its land area but accounts for 57 percent of its built-up density and 68 percent of its population (rows 8-10, Table 8). This is consistent with the fact that, by definition, centers are contiguous sets of grid cells with very high population densities relative to all other cells within a city. For cities that have multiple centers, the largest center is typically much larger than the second largest center. Thus, across countries, the population of a city's largest center is, on average, almost 13 times that of its second largest center, while the area occupied by a city's largest center is, on average, more than 11 times that occupied by its second largest center (rows

<sup>34</sup> As shown in row 2 of table 8, there is a very small minority of cities that have no statistically identifiable centers. A city can have no statistically significant population density peak if its population is very evenly distributed across its extent.

<sup>35</sup> Even though the largest cities have many centers, for 72 percent of SSA countries, the city with the most centers is, somewhat surprisingly, *not* its largest city. And, across countries, the mean number of centers in the city with the most centers is 61.

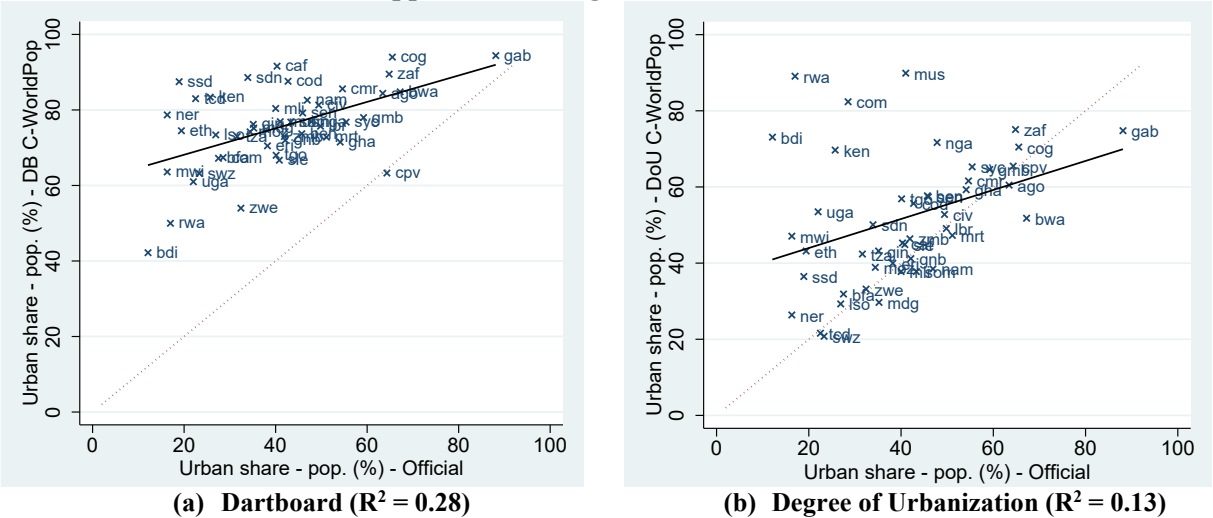
11 and 13).<sup>36</sup> Consistent with this, when we estimate Zipf’s law for the populations of centers for cities with multiple centers, we find an estimated Zipf coefficient that is, on average across cities and countries, negative (row 14), with even lower absolute values than those presented above for cities. By contrast, the estimated elasticity of the number of centers with respect to a city’s population is, on average across cities and countries, positive (row 15), although these estimates, as with the estimated Zipf coefficients, need to be treated with care given that most cities only have a single center.

**6. Comparison with Degree of Urbanization**

Our analysis so far has relied on the dartboard approach to both identify and delineate urban areas for SSA countries. However, there are other approaches to urban area identification and delineation. One of these is the “degree of urbanization” approach of Dijkstra and Poelman (2014) and Dijkstra et al. (2021). While, compared to the dartboard approach, this approach has the shortcoming of being more reliant on arbitrary thresholds to identify and delineate urban areas, it is, nevertheless, an approach that has been widely applied both in academic papers (see, for example, Henderson et al. 2019) and policy reports by international agencies, including by the World Bank (Ferreya and Roberts 2018; Lall et al. 2021; Mukim and Roberts 2023). Given this, in this section we provide a brief comparison of our results from the dartboard approach with those obtained from the alternative use of the degree of urbanization. In doing so, we focus, for brevity, on results for differences in levels of urbanization across countries.

The degree of urbanization defines an urban area as a contiguous set of 1 km<sup>2</sup> gridded population cells in which each cell has a population density of at least 300 people per km<sup>2</sup> and the overall population of the set of cells is at least 5,000. As such, unlike the dartboard approach, which is a relative approach, the degree of urbanization is an absolute approach to urban area identification and delineation.<sup>37</sup> It will also be noted that the use of a 1 km<sup>2</sup> gridded population layer is part of the definition of the degree of urbanization. In applying the degree of urbanization, we therefore aggregate the constrained WorldPop data to this resolution rather than the 250 m\*250 m resolution that we use for the dartboard approach.

**Figure 2: Correlation with official urban population shares across 46 SSA countries – dartboard approach and degree of urbanization**



**Note:** In each graph, the x-axis shows the share of a country’s population that lived in officially defined urban areas in 2015 based on data from the United Nations World Urbanization Prospects: 2018 Revision database (<https://population.un.org/wup/>). The y-axis in part (a) shows a country’s urbanization rate based on the dartboard approach, while the y-axis in part (b) shows a country’s urbanization rate based on the degree of urbanization. Results exclude São Tomé Príncipe.

As Figure 2 shows, as with the dartboard approach, estimates of urbanization rates – i.e., of the share of a country’s population that is urban – derived from the degree of urbanization tend to be higher than urbanization rates based on official national definitions of urban areas. Therefore, according to both

<sup>36</sup> Consistent with this, the shares of population and area accounted for by all centers within a city are, on average across countries, not much higher than the shares accounted for by a city’s largest center (results available on request).

<sup>37</sup> A related difference between the two approaches is that our application of the dartboard approach takes account of the presence of significant desert areas, which the degree of urbanization does not. Hence, the dartboard approach identifies urban grid cells based on their density *relative to the density of all cells within a country that are potentially habitable*, where the potentially habitable areas exclude deserts.

approaches, most SSA countries are more urbanized than official definitions would have us believe, although the number of countries for which this is the case is less for the degree of urbanization (33 out of 46 countries examined) than it is for the dartboard approach (45 out of 46 countries). Estimates of urbanization rates from both approaches are also positively correlated with official urbanization rates. As can be seen from the respective  $R^2$  values in Figure 2, however, the strength of this positive correlation is stronger for the dartboard approach than it is for the degree of urbanization. This is mainly because of four countries – Burundi (BDI), Comoros (COM), Mauritius (MUS), and Rwanda (RWA) – that have relatively low official urbanization rates, but for which the degree of urbanization generates very high estimated rates of urbanization. It is no coincidence that these four countries are among the most densely populated in SSA – indeed, as shown in Table 2, Mauritius, Rwanda, and Comoros are the three most densely populated in the region. Because it is an absolute approach to urban area identification and delineation, unlike with the dartboard approach, these high average population densities almost mechanically translate into high estimated urbanization rates under the degree of urbanization. Once Burundi, Comoros, Mauritius, and Rwanda are dropped from the estimation, the correlation of the urbanization rates derived from the degree of urbanization with the official rates becomes much stronger (the  $R^2$  increases from 0.13 to 0.49).

Although urbanization rates derived from both the dartboard and degree of urbanization approaches are positively correlated across countries with official rates, the urbanization rates from the two approaches are uncorrelated with each other – the  $R^2$  from regressing the dartboard urbanization rate on the corresponding degree of urbanization rate is zero. This lack of correlation, however, is again driven by the same four highly densely populated SSA countries – Burundi, Comoros, Mauritius, and Rwanda – for which the degree of urbanization generates very high estimated rates of urbanization. Once these four countries are excluded, the correlation between the urbanization rates from the two approaches becomes positive and statistically significant at the 5 percent level ( $R^2 = 0.13$ ).

Unsurprisingly given the above results, Rwanda, Burundi, Comoros, and Mauritius are the four SSA countries for which the urbanization rate based on the degree of urbanization most exceeds the urbanization rate based on the dartboard approach (Table 9). In the case of Rwanda, the country is almost 40 percentage points more urbanized according to the degree of urbanization than it is according to the dartboard approach (see also part (a) of Figure 3). Conversely, however, the dartboard approach implies that Chad, Niger, South Sudan, the Central African Republic, and Madagascar are all much more urbanized than the degree of urbanization implies (see also part (b) of Figure 3 for an illustration of this for Chad). These differences are again related to the fact that the dartboard approach is a relative approach to urban area identification and delineation, while the degree of urbanization is an absolute approach. The high urbanization rates for Chad, Niger, South Sudan, the Central African Republic, and Madagascar generated by the dartboard approach compared to the degree of urbanization reflect highly uneven spatial distributions of population even though, compared to Burundi, Comoros, Mauritius and Rwanda, average population densities are relatively low.<sup>38</sup>

**Table 9: Largest differences in urban population shares between the dartboard (DB) and degree of urbanization (DoU) approaches**

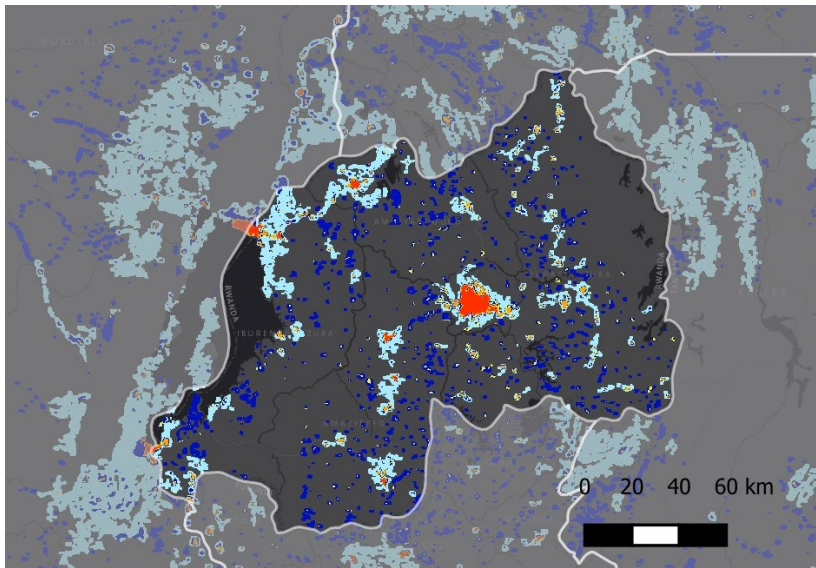
Top 5 – DB > DoU				Top 5 – DoU > DB			
Country	DB	DoU	DB - DU	Country	DB	DoU	DB - DU
Chad	83	21.6	61.4	Rwanda	50	89.1	-39.1
Niger	78.7	26.4	52.3	Burundi	42.2	73.1	-30.9
South Sudan	87.5	36.5	51	Comoros	67.4	82.4	-15
Central African Republic	91.6	45.3	46.3	Mauritius	76.9	89.9	-13
Madagascar	75.2	29.7	45.5	Cabo Verde	63.3	65.5	-2.2

**Note:** Results based on 46 countries and exclude São Tomé Príncipe.

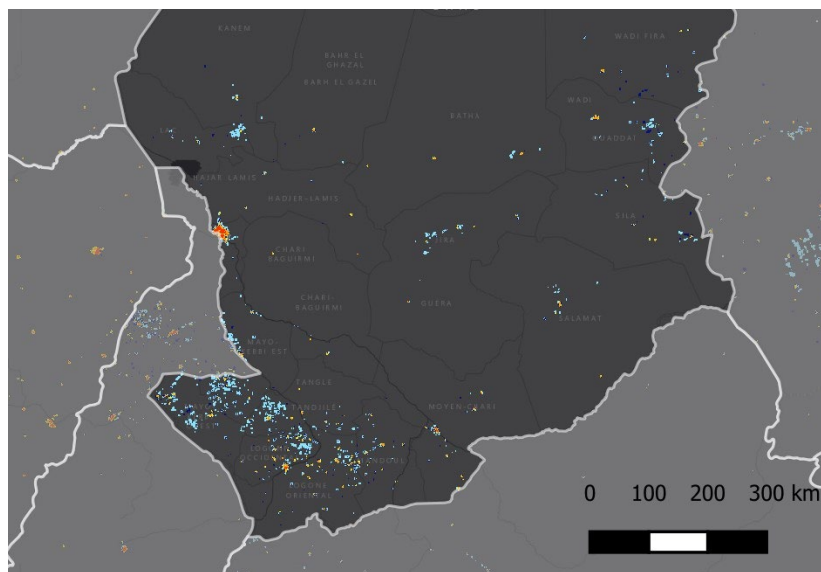
**Figure 3: Maps of urban areas in Rwanda and Chad – dartboard -v- degree of urbanization**

<sup>38</sup> A large share of land area in Chad (50.1 percent) and Niger (66.9 percent) is desert. If we were to calculate average population densities only over the “potentially habitable” areas of these countries, these would be much higher.





(a) Rwanda

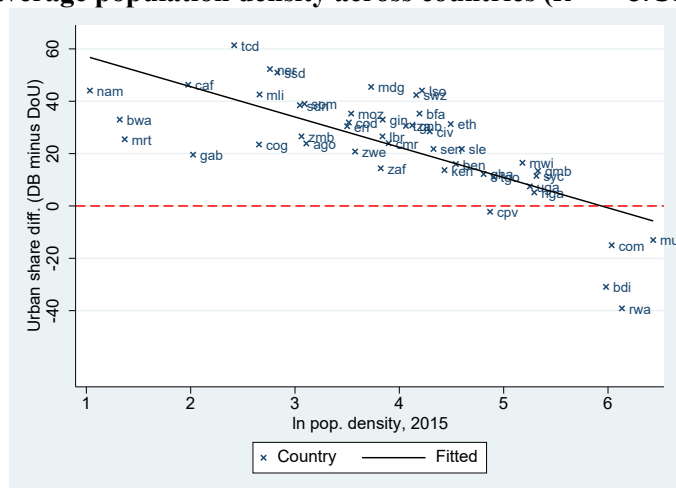


(b) Chad

**Note:** The two maps are color coded according to whether the dartboard and degree of urbanization approaches agree or disagree on the classification of areas as urban. Red indicates areas that both approaches agree are cities, yellow indicates areas that both agree are urban but not cities, dark (light) blue indicates areas that the dartboard approach classifies as urban (rural) and the degree of urbanization as rural (urban), and orange indicates areas that both approaches agree are urban but for which there is disagreement as to whether they are cities.

Finally, and consistent with the above results, there is a strong, and statistically highly significant correlation, at the 1 percent level, between the difference in the urbanization rates generated by the dartboard and degree of urbanization approaches (i.e., dartboard *minus* degree of urbanization) and a country's average population density (Figure 4). This, again, reflects the, almost mechanical, effect of a country's average population density on its level of urbanization as implied by the degree of urbanization.

**Figure 4: Correlation between difference in urban shares and average population density across countries ( $R^2 = 0.53$ )**



**Notes:** Data on average population densities taken from the World Bank’s World Development Indicators (WDI) database (<https://databank.worldbank.org/source/world-development-indicators>). Results exclude São Tomé Príncipe.

It follows from the above discussion that while the dartboard and degree of urbanization approaches agree on some of the broad patterns of urbanization in SSA – most notably that most SSA countries are more urbanized than official data would have us believe – they can produce, what appear to be, dramatically different results for individual countries. These differences, however, reflect the different perspectives – absolute versus relative – on urbanization that underpin the two approaches, both of which, especially when taken together, yield useful information. As such, the two approaches should be viewed as complements rather than substitutes.

## 7. Summary and Conclusion

Sub-Saharan Africa has, over the last six decades, been among the world’s fastest urbanizing regions. And, by 2050, the number of people living in its towns and cities is projected to expand by a further 0.95 billion. Despite this, however, remarkably little is known about the detailed anatomy of patterns of urbanization across the region’s countries. In this paper, we have attempted to remedy this by applying the dartboard algorithm to gridded population data derived from very high-resolution satellite data for all building footprints in the region, thereby allowing us to delineate urban areas for 46 SSA countries. Using this data, we have provided a detailed descriptive analysis of patterns of urbanization within the region for *circa* 2015, including both patterns across and within urban areas.

Our results reveal, *inter alia*, that SSA countries are generally more urbanized than data on officially defined urban areas would have us believe. Despite this, however, there remain large differences in the share of people who live in urban areas across countries with the urban share of a country’s population being positively correlated with both its level of development and area. Regardless of the type of urban area – i.e., cities, towns, or both – considered, most countries, have built-up area elasticities that are greater than one, implying that a doubling of an urban area’s population is associated with a more than doubling of its built-up area. In this regard, SSA countries appear to differ from high-income countries, which tend to have built-up area elasticities that are less than unity. One explanation for this difference may be higher costs of vertical development in SSA countries related to dysfunctional land and property markets, as well as failures in planning. On other dimensions, however, urbanization patterns within SSA countries appear more like those in developed countries. Hence, SSA countries exhibit levels of urban primacy that, on average, are similar to those in high-income countries, while Zipf’s law generally holds quite well. As for the internal structures of cities, most cities in most SSA countries are monocentric with only a single statistically significant peak of population density and both negative population and built-up density gradients. However, this is not the case for all cities and the very largest cities in the region have many centers of dense population concentration, some of which may reflect densely populated informal neighborhoods. Moreover, within cities with multiple centers, there is a very strong rank-size relationship among centers, with the largest center having, on average, almost 13 times the population of its second largest center, while occupying 11 times as much land.

Despite the new insights into patterns of urbanization both across and within SSA countries generated by our analysis, there are important limitations that should be noted. Hence, our analysis is limited to a single cross-section in time, *circa* 2015. This is because of the limitations of the underlying gridded population data that we rely on as input into our analysis. Thus, the detailed map of all building footprints in the region that underpins the derivation of the gridded population data we use is only available for *circa* 2015. Moreover, even for this year, although it is the best available, the quality of the population data that we use is ultimately limited by the quality of the underlying population data available from national population Censuses in the region. For some SSA countries, this data is quite outdated, while, for many, it is only available for large administrative regions as opposed to, for example, small census tracts, which explains the need for top-down gridding of the data in the first place. In this sense, improving the quality and spatial resolution of available population Census data for SSA countries is crucial to further progress on the understanding of urbanization in the region.

## References

- Ahfeldt, G. M., and E. Pietrostefani. 2019. "The economic effects of density: A synthesis." *Journal of Urban Economics* 111: 93-107.
- Alonso, W. 1964. *Location and Land Use: Towards a General Theory of Land Rent*. Cambridge, Mass.: Harvard University Press.
- Berry, B.J.L. 1960. "The impact of expanding metropolitan communities upon the central place hierarchy." *Annals of the Association of American Geographers* 50(2):112–116.
- Berry, B., J. Lobley, P.G. Goheen, and H. Goldstein. 1969. *Metropolitan Area Definition: A Re-evaluation of Concept and Statistical Practice*. Washington, D.C.: US Bureau of the Census.
- Black, D. and V. Henderson. 2003. "Urban evolution in the USA." *Journal of Economic Geography*, 3(4): 343–372.
- Bontemps S., P. Defourny, J. Radoux, E. Van Bogaert, C. Lamarche, F. Achard, P. Mayaux, M. Boettcher, C. Brockmann, G. Kirches, M. Zülkhe, V. Kalogirou, O. Arino. 2013. "Consistent Global Land Cover Maps for Climate Modeling Communities: Current Achievements of the ESA's Land Cover." CCI ESA Living Planet Symposium 9 - 13 September 2013, Edinburgh, United Kingdom.
- Bosker, M. and E. Buringh. 2017 "City seeds: Geography and the origins of the European city system." *Journal of Urban Economics* 98: 139–157. Urbanization in Developing Countries: Past and Present.
- Bosker, M., J. Park, and M. Roberts. 2021. "Definition matters. Metropolitan areas and agglomeration economies in a large-developing country." *Journal of Urban Economics* 125.
- Brakman, S., H. Garretsen and C. van Marrewijk. 2009. *The New Introduction to Geographical Economics*. Cambridge : Cambridge University Press.
- Ch, R., D. A. Martin, and J. F. Vargas. 2021. "Measuring the size and growth of cities using nighttime light." *Journal of Urban Economics* 125.
- Cheshire, P.C. and D. Hay. 1989. *Urban Problems in Western Europe: An Economic Analysis*. London: Unwin Hyman.
- de Bellefon, M. P, P-P. Combes, G. Duranton, L., Gobillon, and C. Gorin. 2021. "Delineating urban areas using building density." *Journal of Urban Economics* 125.
- Desmet, K. and J. Rappaport. 2017. "The settlement of the United States, 1800–2000: The long transition towards Gibrat's law." *Journal of Urban Economics* 98: 50–68.
- Dijkstra, L., and H. Poelman. 2014. "A harmonised definition of cities and rural areas: The new degree of urbanization." Regional Working Paper, Directorate-General for Regional and Urban Policy. Brussels: European Commission.
- Dijkstra, L., A. J. Florczyk, S. Freire, T. Kemper, M. Pesaresi, and M. Schiavina. 2021. "Applying the Degree of Urbanization to the Globe: A New Harmonized Definition Reveals a Different Picture of Global Urbanization." *Journal of Urban Economics* 125.
- Dobkins, L.H. and Y.M Ioannides. 2001. "Spatial interactions among U.S. cities: 1900–1990," *Regional Science and Urban Economics* 31(6): 701–731.
- Düben, C. and M. Krause. 2021. "The Emperor's Geography - City Locations, Nature, and Institutional Optimization," SSRN Electronic Journal.
- Duranton, G. 2007. "Urban Evolutions: The Fast, the Slow, and the Still." *American Economic Review* 97(1): 197–221.
- Duranton, G. 2015a. "A Proposal to Delineate Metropolitan Areas in Colombia." *Desarrollo y Sociedad* 15: 223-64.
- Eeckhout, J. 2004. "Gibrat's Law for (All) Cities." *American Economic Review* 49(5): 1429–1451.
- Ellis, P., and M. Roberts. 2016. *Leveraging Urbanization in South Asia: Managing Spatial Transformation for Prosperity and Livability*. Washington, D.C.: World Bank.
- Ferreira, M.M. and M. Roberts. 2018. *Raising the Bar for Productive Cities in Latin America and the Caribbean*. Washington, D.C.: World Bank.
- Fox, K.A. and T.K. Kumar. 1965. "The functional economic area: Delineation and implications for economic analysis and policy." *Papers of the Regional Science Association* 15(1):57–85.
- Gabaix, X. 1999. "Zipf's Law for Cities: An Explanation." *Quarterly Journal of Economics* 114(3): 739–767.
- Gabaix, X. and R. Ibragimov. 2011. "Rank - 1/2: A simple way to improve the OLS estimation of tail exponents." *Journal of Business Economics and Statistics* 29(1):24–39.
- Giuliano, G. and K.A. Small. 1991. "Subcenters in the Los Angeles region." *Regional Science and Urban Economics* 21(2): 163–182.
- Goswami, A.G. and S.V. Lall. 2016. "Jobs in the city: Explaining urban spatial structure in Kampala." World Bank Policy Research Working Paper No. 7655. Washington, D.C.: World Bank.

- Hall, P.G. and D. Hay. 1980. *Growth centres in the European urban system*. London: Heinemann Educational Books.
- Heinrigs, P. 2021. "Africapolis: understanding the dynamics of urbanization in Africa." *Field Actions Science Reports*, Special Issue 22.
- Henderson, J. V., D. Nigmatulina, and S. Kriticos. 2021. "Measuring Urban Economic Density." *Journal of Urban Economics* 125(C).
- Henderson, J.V., C. Peng, and A.J. Venables. 2022. "Growth in the African Urban Hierarchy." CEPR Discussion Paper No. 17493.
- Henderson, J.V., T. Regan and A.J. Venables. 2021. "Building the City: From Slums to a Modern Metropolis [New Estimates of the Elasticity of Substitution of Land for Capital]." *Review of Economic Studies* 88(3): 1157-1192.
- Henderson, J.V. and M.A. Turner. 2020. "Urbanization in the Developing World: Too Early or Too Slow?" *Journal of Economic Perspectives* Summer 34(3): 150-173.
- Jedwab, R. and A. Storeygard. 2021. "The Average and Heterogeneous Effects of Transportation Investments: Evidence from Sub-Saharan Africa 1960-2010." *Journal of the European Economic Association*.
- Kanemoto, Y. and R. Kurima. 2005. "Urban employment areas: Defining Japanese metropolitan areas and constructing the statistical database for them." In Okabe, A. (ed.) *GIS Based Studies in the Humanities and Social Sciences*. Boca Raton: Taylor & Francis, 85–97.
- Lall, S.V., M. Lebrand, H. Park, D. Sturm, and A.J. Venables. 2021. *Pancakes to Pyramids: City Form to Promote Sustainable Growth*. Washington, D.C.: World Bank.
- McDonald, J.F. 1987. "The identification of urban employment sub-centers." *Journal of Urban Economics* 21(2): 242–258.
- McMillen, D.P. 2001. "Nonparametric employment sub-center identification." *Journal of Urban Economics* 50(3): 448–473.
- Mills, E.S. 1967. "An Aggregative Model of Resource Allocation in a Metropolitan Area." *American Economic Review* 57(2): 197-210.
- Mukim, M. and M. Roberts. 2023. *Thriving: Making Cities Green, Resilient, and Inclusive in a Changing Climate*. Washington, D.C.: World Bank.
- Muth, R. F. 1969. *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*. Chicago and London: The University of Chicago Press.
- OECD/SWAC. 2020. *Africa's Urbanisation Dynamics 2020: Africapolis, Mapping a New Urban Geography*, West African Studies, Paris: OECD Publishing.
- OECD/UN ECA/AfDB. 2022. *Africa's Urbanisation Dynamics 2022: The Economic Power of Africa's Cities*, West African Studies, Paris: OECD Publishing.
- Roberts, M., B. Blankespoor, C. Deuskar, and B. Stewart. 2017. "Urbanization and development: Is Latin America and the Caribbean different from the rest of the world?" World Bank Policy Research Working Paper 8019. Washington, D.C.: World Bank.
- Roberts, M. 2018. "The Many Dimensions of Urbanization and the Productivity of Cities in Latin America and the Caribbean." In *Raising the Bar for Productive Cities in Latin America and the Caribbean*, edited by M.M. Ferreyra and M. Roberts, 49-85. Washington, D.C.: World Bank.
- Stevens F.R., A.E. Gaughan, C. Linard, and A.J. Tatem. 2015. "Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data." *PLoS ONE* 10(2).
- Soo, K.T. 2005. "Zipf's law for cities: A cross-country investigation." *Regional Science and Urban Economics* 35: 239-263.
- Uchida, H., and A. Nelson. 2009. "Agglomeration Index: Towards a New Measure of Urban Concentration." Background paper to the 2009 World Development Report. Washington, D.C.: World Bank.
- World Bank. 2009. *World Development Report, 2009: Reshaping Economic Geography*. Washington, D.C.: The World Bank.
- Zhou, N., K. Hubacek, and M. Roberts. 2015. "Analysis of Spatial Patterns of Urban Growth across South Asia Using DMSP-OLS Nighttime Lights Data." *Applied Geography* 63: 292-303.

Table A1: OLS regression results - relationship between country size and urbanization characteristics, all types of urban area

		(1)		(2)			(3)			
Dependent variable	Type of area	Pop.	R2	Density	Area	R2	Density	Area	GDP per cap.	R2
Pop. share	Urban Areas	2.92**	0.10	-3.32*	2.67**	0.35	-1.91	3.75***	4.86**	0.42
	Towns	0.67	0.02	-0.30	0.63	0.05	-1.24	-0.09	-3.23***	0.19
	Urban Cores	1.39	0.03	-5.06***	1.13	0.39	-3.51**	2.31**	5.32***	0.50
Area share	Urban Areas	0.02	0.00	2.06***	0.11	0.47	2.09***	0.13	0.12	0.47
	Towns	0.04	0.00	0.67***	0.07	0.35	0.57***	-0.02	-0.36*	0.40
	Urban Cores	-0.09	0.06	0.25***	-0.07*	0.53	0.30***	-0.03	0.18**	0.59
Pop. density	Urban Areas	-69.97	0.02	391.73***	-51.13	0.49	438.19***	-15.53	159.66	0.52
	Towns	-76.16	0.05	259.13***	-62.48*	0.56	279.46***	-46.90	69.89	0.57
	Urban Cores	91.86	0.00	1146.98***	134.93	0.26	1217.46***	188.94	242.23	0.27
Area elas.	Urban Areas	-0.00	0.00	0.04***	0.00	0.41	0.03***	-0.00	-0.01	0.44
	Towns	0.01	0.06	0.03***	0.01*	0.21	0.03***	0.01	-0.02	0.24
	Urban Cores	-0.01	0.01	-0.02	-0.01	0.02	-0.02	-0.01	-0.02	0.03
Built-up elas.	Urban Areas	-0.02	0.01	0.09*	-0.01	0.16	0.09*	-0.01	-0.01	0.16
	Towns	0.05	0.03	0.17***	0.05	0.19	0.19***	0.06	0.07	0.22
	Urban Cores	-0.08*	0.06	-0.06	-0.08*	0.06	-0.09	-0.11**	-0.12	0.10
No. units	Urban Areas	1999.27***	0.49	1228.06**	1967.79***	0.53	1172.07**	1924.88***	-192.44	0.53
	Towns	1807.28***	0.48	1078.61**	1777.54***	0.52	1032.01**	1741.83***	-160.16	0.52
	Urban Cores	293.95***	0.42	242.26***	291.84***	0.43	233.91**	285.45***	-28.69	0.43
Average pop.	Urban Areas	103.39	0.00	2517.71***	201.93	0.48	3124.35***	666.76**	2084.85***	0.63
	Towns	-81.86	0.01	649.42***	-52.01	0.51	718.20***	0.69	236.37	0.53
	Urban Cores	1235.21	0.01	5342.38**	1402.84	0.10	6881.85**	2582.45	5290.72*	0.16
Average area	Urban Areas	0.46	0.05	0.57	0.46	0.06	0.92*	0.73**	1.21**	0.15
	Towns	0.16**	0.13	-0.02	0.15**	0.23	0.01	0.18***	0.13	0.25
	Urban Cores	0.15	0.01	0.03	0.15	0.01	0.38	0.42	1.21*	0.09
Pop. 75/P25	Urban Areas	-0.52**	0.10	-0.90**	-0.54**	0.13	-0.62	-0.32	0.96**	0.22

	Towns	-0.14	0.04	-0.36*	-0.15	0.08	-0.24	-0.06	0.39*	0.15
	Urban Cores	-6.35***	0.25	-3.80	-6.25***	0.28	-2.03	-4.89***	6.09*	0.33
Pop. 90/P10	Urban Areas	-1.18	0.01	-3.48	-1.28	0.03	-1.73	0.06	6.02*	0.10
	Towns	-0.40	0.01	-1.86	-0.46	0.06	-1.14	0.09	2.47*	0.13
	Urban Cores	-51.99**	0.13	-21.98	-50.76**	0.16	-11.03	-42.37*	37.64	0.18
den. P75/P25	Urban Areas	-0.03	0.05	-0.09***	-0.03*	0.16	-0.08**	-0.03	0.01	0.16
	Towns	-0.03**	0.11	-0.05**	-0.04**	0.13	-0.05**	-0.04**	-0.00	0.13
	Urban Cores	-0.11***	0.38	-0.08**	-0.11***	0.40	-0.07*	-0.10***	0.06	0.42
den. P90/P10	Urban Areas	-0.01	0.00	-0.26***	-0.02	0.28	-0.25***	-0.01	0.06	0.28
	Towns	-0.03	0.02	-0.14***	-0.03	0.27	-0.14***	-0.03	-0.00	0.27
	Urban Cores	-0.23***	0.16	-0.06	-0.23***	0.21	-0.02	-0.20**	0.14	0.23
Share largest	Urban Areas	-4.22***	0.30	-3.35**	-4.18***	0.31	-1.67	-2.89***	5.77***	0.45
	Towns	-0.56**	0.13	0.07	-0.54**	0.22	0.05	-0.55**	-0.06	0.22
	Urban Cores	-2.33***	0.19	-3.10**	-2.36***	0.20	-2.64**	-2.00**	1.59	0.22
Zipf's coef.	Urban Areas	-0.03*	0.07	0.01	-0.03*	0.14	0.03	-0.02	0.07**	0.22
	Towns	0.04	0.04	-0.01	0.04	0.08	0.01	0.05	0.05	0.10
	Urban Cores	-0.03**	0.09	-0.02	-0.03**	0.09	-0.00	-0.01	0.06**	0.20
<b>Notes:</b> Estimation results for equations [1]-[3] for different urbanisation characteristics (listed in the notes of Tables 1 and 3). All regressions based on 47 countries except those for Zipf's coefficient, which exclude Seychelles due to it only having three cities.										

**Table A2: Highest and lowest values for within-city density gradients**

	<b>Largest</b>	<b>2nd Largest</b>	<b>3rd Largest</b>	<b>3rd Lowest</b>	<b>2nd Lowest</b>	<b>Lowest</b>
<b>Population density gradients</b>						
Pop. grad. sh. <0	Cabo Verde (1.00)	Gambia, The (1.00)	Togo (0.97)	Rwanda (0.42)	Lesotho (0.30)	Eswatini (0.28)
Pop. grad. sh. ns	Eswatini (0.72)	Lesotho (0.70)	Malawi (0.54)	Togo (0.03)	Gambia, The (0.00)	Gambia, The (0.00)
Pop. grad. sh. >0	Mauritius (0.08)	Burundi (0.06)	Rwanda (0.05)	Gabon (0.00)	Togo (0.00)	Togo (0.00)
Pop. grad. Av.	Eswatini (-0.12)	Malawi (-0.23)	South Sudan (-0.24)	Congo Brazzaville (-1.74)	Côte d'Ivoire (-2.06)	Mali (-2.26)
Pop. grad. P10	Lesotho (-1.43)	Cabo Verde (-1.47)	Rwanda (-1.48)	Niger (-4.10)	Côte d'Ivoire (-4.22)	Mali (-4.89)
Pop. grad. P25	Rwanda (-0.95)	Cabo Verde (-1.10)	Lesotho (-1.23)	Niger (-3.52)	Côte d'Ivoire (-3.61)	Mali (-4.20)
Pop. grad. P50	Rwanda (-0.48)	Eswatini (-0.50)	Lesotho (-0.82)	Niger (-2.63)	Côte d'Ivoire (-2.87)	Mali (-3.47)
Pop. grad. P75	Rwanda (-0.23)	Mauritius (-0.34)	Malawi (-0.44)	Botswana (-1.76)	Côte d'Ivoire (-2.20)	Mali (-2.36)
Pop. grad. P90	Mauritius (0.60)	Rwanda (0.05)	Eswatini (-0.12)	Sierra Leone (-1.20)	Mali (-1.45)	Côte d'Ivoire (-1.52)
<b>Built-up density gradients</b>						
Built grad. sh. <0	Cabo Verde (1.00)	Gambia, The (1.00)	Cabo Verde (1.00)	Chad (0.36)	Eritrea (0.31)	South Sudan (0.27)
Built grad. sh. ns	South Sudan (0.72)	Eritrea (0.68)	Chad (0.64)	Cabo Verde (0.00)	Cabo Verde (0.00)	Cabo Verde (0.00)
Built grad. sh. >0	Eswatini (0.06)	Burundi (0.06)	Rwanda (0.05)	Mauritius (0.00)	Mauritius (0.00)	Sierra Leone (0.00)
Built grad. Av.	Mauritius (-0.44)	Eswatini (-0.69)	Niger (-0.83)	Congo Brazzaville (-1.91)	Mali (-1.98)	Burkina Faso (-2.14)
Built grad. P10	Mauritius (-0.74)	Rwanda (-1.69)	South Africa (-2.09)	Senegal (-2.95)	Cabo Verde (-2.98)	Mauritania (-3.10)
Built grad. P25	Mauritius (-0.72)	Cabo Verde (-1.38)	Rwanda (-1.49)	Gabon (-2.34)	Angola (-2.45)	Guinea-Bissau (-2.49)
Built grad. P50	Mauritius (-0.47)	Rwanda (-1.05)	Eswatini (-1.06)	Gabon (-1.81)	Angola (-1.88)	Guinea-Bissau (-2.05)
Built grad. P75	Mauritius (-0.35)	Eswatini (-0.59)	Rwanda (-0.71)	Mauritania (-1.39)	Guinea (-1.39)	Burkina Faso (-1.45)
Built grad. P90	Eswatin (0.15)	Rwanda (-0.28)	Mauritius (-0.31)	Senegal (-1.07)	Congo Brazzaville (-1.08)	Guinea (-1.10)
<p><b>Notes:</b> <i>Pop. grad. sh. &lt;0</i>: Share of cities with significant (10% threshold) negative population density gradient. <i>Pop. grad. sh. ns</i>: Share of cities with non-significant (10% threshold) population density gradient. <i>Pop. grad. sh. &gt;0</i>: Share of cities with significant (10% threshold) positive population density gradient. <i>Pop. grad. Av.</i>: average (over all cities) population density gradient (when significantly (10% threshold) different from zero). <i>Built-up grad. sh. &lt;0</i>: Share of cities with significant (10% threshold) negative built-up density gradient. <i>Built-up grad. sh. ns</i>: Share of cities with non-significant (10% threshold) built-up density gradient. <i>Built-up grad. sh. &gt;0</i>: Share of cities with significant (10% threshold) positive built-up density gradient. <i>Built-up grad. Av.</i>: average (over all cities) built-up density gradient (when significantly (10% threshold) different from zero). <i>Built-up grad.</i></p>						



**Table A3: Highest and lowest values for city center characteristics**

	<b>Largest</b>	<b>2nd Largest</b>	<b>3rd Largest</b>	<b>3rd Lowest</b>	<b>2nd Lowest</b>	<b>Lowest</b>
# centers	Namibia (3.98)	Kenya (2.97)	South Sudan (2.69)	Mali (1.07)	Côte d'Ivoire (1.07)	Guinea-Bissau (1.05)
Sh. 0 cen.	Rwanda (0.11)	Uganda (0.07)	Malawi (0.06)	Cameroon (0.00)	Gabon (0.00)	Cameroon (0.00)
Sh. 1 cen.	Guinea-Bissau (0.95)	Mali (0.95)	Liberia (0.95)	South Sudan (0.63)	Zimbabwe (0.63)	Eswatini (0.63)
Sh. 2 cen.	Cabo Verde (0.20)	Togo (0.18)	Zambia (0.18)	Liberia (0.04)	Mali (0.03)	Burundi (0.00)
Sh. 3-5 cen.	Zimbabwe (0.15)	Ghana (0.13)	Gambia, The (0.13)	Congo Brazzaville (0.00)	Guinea-Bissau (0.00)	Guinea-Bissau (0.00)
Sh. >6 cen.	Zimbabwe (0.11)	Eswatini (0.09)	South Sudan (0.08)	Liberia (0.00)	Côte d'Ivoire (0.00)	Guinea-Bissau (0.00)
#cen. largest	Kenya (499)	Ethiopia (432)	Niger (287)	Central African R. (1)	Cameroon (1)	Chad (1)
Sh. pop. lar. cen.	Congo Brazzaville (0.81)	Mali (0.81)	Gabon (0.78)	Cabo Verde (0.56)	Eswatini (0.56)	Zimbabwe (0.55)
Sh. bui. lar. cen.	Angola (0.67)	Zambia (0.67)	Congo Brazzaville (0.67)	Niger (0.43)	Cabo Verde (0.42)	South Africa (0.37)
Sh. area lar. cen.	Lesotho (0.63)	Rwanda (0.45)	Eswatini (0.43)	Central African R. (0.19)	Namibia (0.18)	South Sudan (0.18)
lar/2nd pop.	Togo (83.6)	Senegal (55.9)	Congo Brazzaville (38.2)	Niger (3.4)	Burundi (2.7)	Sierra Leone (2.2)
lar/2nd bui.	Congo Brazzaville (62135.6)	Cameroon (5006.4)	South Sudan (3004.8)	Liberia (11.1)	Mauritius (8.6)	Sierra Leone (5.3)
lar/2nd area	Togo (86.6)	Cameroon (37.1)	Angola (31.7)	Niger (3.4)	Burundi (2.5)	Sierra Leone (2.4)
Zipf centers	Togo (-0.30)	Congo Brazzaville (-0.33)	Gabon (-0.37)	Central African Republic (-0.80)	Niger (-0.82)	Sierra Leone (-0.83)
Elas. #cen. pop.	South Sudan (0.43)	Chad (0.41)	Benin (0.40)	Cabo Verde (0.03)	Liberia (0.01)	Guinea-Bissau (0.00)
<p><b>Notes:</b> #centers: Number of centers per city. <i>Sh. 0 cen.</i>: Share of cities without center. <i>Sh. 1 cen.</i>: Share of cities with one center. <i>Sh. 2 cen.</i>: Share of cities with two centers. <i>Sh. 3-5 cen.</i>: Share of cities with three to five centers. <i>Sh. &gt;6 cen.</i>: Share of cities with 6 or more centers. #cen. largest: Number of centers of the largest city. <i>Sh. pop. lar. cen.</i>: City population share of the largest center. <i>Sh. bui. lar. cen.</i>: City built share of the largest center. <i>Sh. area lar. cen.</i>: City area share of the largest center. <i>lar/2nd pop.</i>: Relative population of the second over the largest center. <i>lar/2nd bui.</i>: Relative built-up area of the second over the largest center. <i>lar/2nd area</i>: Relative area of the second over the largest center. <i>Zipf centers</i>: Within city population Zipf's coefficient for city centers. <i>Elas. #cen. pop.</i>: elasticity of the number of centers in the city with respect to its population.</p>						

**Table A4: OLS regression results - relationship between country size and within city concentration characteristics**

Dependent variable	(1)		(2)			(3)			
	Pop.	R2	Density	Area	R2	Density	Area	GDP per cap.	R2
Pop. grad. sh. <0	0.02	0.01	-0.03	0.02	0.12	-0.02	0.03	0.03	0.13
Pop. grad. sh. ns	-0.02	0.02	0.02	-0.02	0.10	0.01	-0.03	-0.03	0.12
Pop. grad. sh. >0	0.00	0.00	0.01**	-0.00	0.21	0.01**	-0.00	-0.00	0.22
Pop. grad. Av.	0.04	0.01	0.16**	0.02	0.13	0.16*	0.02	-0.02	0.14
Pop. grad. P10	-0.24***	0.17	0.02	-0.29***	0.40	0.04	-0.27***	0.09	0.40
Pop. grad. P25	-0.21***	0.16	0.04	-0.26***	0.41	0.07	-0.24***	0.12	0.43
Pop. grad. P50	-0.16**	0.11	0.08	-0.20***	0.40	0.10	-0.19***	0.08	0.41
Pop. grad. P75	-0.04	0.02	0.12*	-0.07	0.24	0.13*	-0.06	0.04	0.24
Pop. grad. P90	-0.02	0.01	0.11**	-0.04	0.25	0.12**	-0.04	0.05	0.26
Built grad. sh. <0	-0.01	0.00	0.01	-0.01	0.03	0.04	0.00	0.09**	0.14
Built grad. sh. ns	0.01	0.00	-0.02	0.01	0.04	-0.04	-0.00	-0.09**	0.14
Built grad. sh. >0	0.00	0.03	0.01**	0.00	0.11	0.00*	0.00	-0.00	0.13
Built grad. Av.	0.01	0.00	0.10*	-0.01	0.11	0.12*	0.01	0.09	0.15
Built grad. P10	-0.00	0.00	0.14**	-0.02	0.28	0.16***	-0.01	0.09	0.31
Built grad. P25	-0.03	0.02	0.08*	-0.05	0.29	0.09**	-0.04	0.07	0.32
Built grad. P50	-0.03	0.03	0.08**	-0.05**	0.42	0.10***	-0.04	0.08**	0.48
Built grad. P75	-0.03	0.04	0.06*	-0.05**	0.37	0.07**	-0.04*	0.05	0.40
Built grad. P90	-0.01	0.00	0.06	-0.02	0.15	0.06	-0.02	0.03	0.15
#centers	0.09	0.04	0.00	0.10	0.08	0.03	0.12	0.12	0.11
Sh. 0 cen.	0.00	0.02	0.01	0.00	0.07	0.01	0.00	-0.01	0.09
Sh. 1 cen.	-0.01	0.01	-0.01	-0.01	0.02	-0.01	-0.01	-0.01	0.02
Sh. 2 cen.	-0.00	0.00	-0.00	-0.00	0.01	-0.00	0.00	0.01	0.05
Sh. 3-5 cen.	0.00	0.01	0.01	0.00	0.06	0.01	0.00	-0.00	0.06
Sh. >6 cen.	0.00	0.03	0.00	0.00	0.05	0.00	0.00	0.00	0.06
#cen. largest	24.06*	0.08	14.42	25.60*	0.10	17.90	27.84**	14.64	0.11
Sh. pop. lar. cen.	-0.00	0.00	-0.03***	0.01	0.40	-0.03***	0.01	0.00	0.40
Sh. bui. lar. cen.	-0.00	0.00	-0.00	0.00	0.00	-0.00	-0.00	-0.01	0.01
Sh. area lar. cen.	-0.02	0.06	0.02	-0.02**	0.34	0.02	-0.02**	0.00	0.34
lar/2nd pop.	-0.57	0.00	-0.25	-0.63	0.00	0.74	0.01	4.16	0.05
lar/2nd bui.	-821.53	0.01	-1910.43	-647.81	0.04	-1597.73	-446.16	1317.72	0.05
lar/2nd area	-0.38	0.00	0.27	-0.49	0.01	0.89	-0.08	2.62	0.03
Zipf centers	-0.01	0.02	-0.00	-0.01	0.03	0.01	-0.00	0.06**	0.17
Elas. #cen. pop.	0.04***	0.24	0.02	0.05***	0.29	0.02	0.05***	0.00	0.29

**Notes:** Estimation results for equations [1] – [3] for different within city concentration measures (listed in the notes of Tables 1 and 3). All regressions based on 43 SSA countries, Comoros, Madagascar, Seychelles and São Tomé Príncipe being excluded for the within-city statistics given their to low number of cities or of centers in these cities.