Habitat International 148 (2024) 103056

Contents lists available at ScienceDirect

Habitat International

journal homepage: www.elsevier.com/locate/habitatint

The unseen population: Do we underestimate slum dwellers in cities of the Global South?

Julius H.P. Breuer [a], John Friesen [a,1,*], Hannes Taubenböck [b,c], Michael Wurm [b], Peter F. Pelz [a]

[a] Chair of Fluid Systems, TU Darmstadt, Otto-Berndt-Straße 2, 64287, Darmstadt, Germany
[b] Deutsches Zentrum für Luft- und Raumfahrt, Earth Observation Center, 82334, Wessling, Germany
[c] Institute for Geography and Geology, Julius-Maximilians-Universität, Würzburg, 97074, Germany

ARTICLE INFO

Keywords:
Slum population
Earth-observation
Validation
Uniform approach

ABSTRACT

The Sustainable development goals (SDG) aim for reducing poverty (SDG 1) and to upgrade all slums (SDG 11). The first indicator in SDG 11 describes the proportion of the urban population residing in slums. However, the currently available data is based on national estimates that follow globally varying methodologies and concepts. In this paper, a uniform approach is implemented to obtain slum population estimates in eight different cities from three continents. The approach relies on earth observation datasets on the spatial extent of the slums and one of the most accepted gridded population dataset: WorldPop. The results shed light on the distribution of population in slums around the world. Nevertheless, the question of the accuracy of these population numbers arises. Therefore, a broad range of literature data containing population counts is gathered for the cities investigated, for varying years and for different spatial scales. The literature data is compared to results obtained by the presented approach. The comparison yields a plausibility assessment for different cities, indicating varying levels of deviation. We find in all cities a clear bias in estimating the slum population - mostly underestimations -, even though some cities reveal a significantly better fit to the data. In conclusion, this study provides a methodology to systematically assess the accuracy of globally available datasets in the context of slums and thereby to highlight the large uncertainties which can empirically be observed.

1. Introduction

Since 2007 worldwide more people live in cities than in rural areas. This trend is projected to increase to 2 in 3 people by 2050 (United Nations, 2018). The rapid urbanization and the emergence of megacities is expected to mainly take place in African and Asian cities. This is where 90 % of the world's population growth is expected to occur (United Nations, 2018). This poses a number of challenges, such as the strain on urban infrastructure. Among the United Nations Sustainable Development Goals (SDGs), especially goal 11 aims to improve life in cities by making them more inclusive, safe, resilient and sustainable. Here, one instrument is upgrading slums (UN General Assembly, 2015).

Therefore, estimating infrastructure necessities is particularly important in these communities, as infrastructure is insufficient here per definition: According to UN-Habitat, slum households are lacking one of the following: adequate water, adequate sanitation, durable housing, security of tenure or less than four people living the same room.

Many authors advocate for nuanced terminology to address the varied dimensions of deprivation (Abascal et al., 2022). Terms like "deprived areas" (Abascal et al., 2022), "informal settlements" (Mahabir et al., 2016), "slums" (Wurm & Taubenböck, 2018), and "arrival cities" (Taubenböck et al., 2018) are commonly used interchangeably. In this paper, we adopt the term "slum" for simplicity, acknowledging the extensive literature on the subject and the fact that slums may have different names in different countries. We are aware that slums are a relative concept and the characteristics and the local understanding of what is meant by this is variable (Gilbert, 2007). By using the term "slum", we refer in this study to a specific morphology (Taubenböck et al., 2018) and the particular urban areas referred to often face challenges with inadequate infrastructure access.

* Corresponding author.
E-mail addresses: julius.breuer@tu-darmstadt.de (J.H.P. Breuer), john.friesen@tu-darmstadt.de, john.friesen@outlook.de (J. Friesen), hannes.taubenboeck@dlr.de (H. Taubenböck), michael.wurm@dlr.de (M. Wurm), peter.pelz@tu-darmstadt.de (P.F. Pelz).

[1] This document is a result of the research project "Uniform detection and modelling of slums to determine infrastructure needs" (Grant ID: LOEWE/5/A002/519/06/00.003(0003)/E15) funded by the LOEWE Program of Hesse State Ministry for Higher Education, Research and the Arts.

https://doi.org/10.1016/j.habitatint.2024.103056
Received 29 September 2023; Received in revised form 2 February 2024; Accepted 4 March 2024

This lack of infrastructure in turn is often the result of increasing population counts and people inhabiting places that were previously neither urbanised nor formalised (Mahabir et al., 2016). Yet an estimated 1 billion people worldwide live in slum households.

However, for many countries, particularly in the Global South, i.e. parts of South America, Africa and Asia, the quality of population data is poor and, where available quickly outdated due to the immense population growth and migration flows (Wardrop et al., 2018). Therefore, the question of where and how people live is fundamental to properly plan infrastructure measures. If known, technical infrastructure such as water and electricity supply and social infrastructure like public transport, medical care facilities and schools, among others, can be designed accordingly.

While the percentage of the global urban population living in slums has decreased from 28 % in 2000 to 24 % in 2018, the absolute numbers have increased (UN-Habitat, 2020). Concretely, the growth in number of slum communities, e.g. in Hyderabad (70 % in 7 years (Kit & Lüdeke, 2013)) or Dhaka (70 % in 9 years (Angeles et al., 2009)) emphasises the urgency for better planning. However, it is difficult to say how accurate these estimates are as there are not many comparable estimates at the global level. In this context, Abbott describes slums as “holes within the urban cadastre” (Abbott, 2001).

This leads to two research questions, we address in this paper:

How many people are living in slums in cities of different world regions? How is the population distributed across the slums in these cities?

Since we answer these questions by conducting a global comparative study, globally uniformly derived datasets are in need: (i) datasets describing the boundaries of slums in different cities and (ii) datasets providing information on the population living in these cities. Therefore, we combine data from two sources, one (i) for slum boundaries and one (ii) for population estimates. Both are accepted by experts for the respective purpose. This in turn raises the third research question:

How plausible is the information gained by combining these two datasets in a uniform approach?

Before presenting the outline of the paper, we give a short introduction on different methodologies of (i) how the location of slums and (ii) how the number of population living there can be identified.

1.1. Where are slums located?

To determine where slums in cities are located, different approaches exist. In some countries like Brazil, Bangladesh and India slums are surveyed and mapped in censuses (Bangladesh Bureau of Statistics, 2014; Instituto Brasileiro de Geografia e Estatistica, 2010; Office of the Registrar General, 2011) but there are also non-governmental sources of slum mapping. For example in (Angeles et al., 2009) slums in six cities of Bangladesh were mapped. Additionally, there are initiatives such as the Know Your City (KYC) campaign by Slum Dwellers International. This is a community-based NGO in which the slum inhabitants create maps of their own settlements. This way, more than 7000 slum settlements in 32 countries were already mapped (Slum/Shack Dwellers International, 2017). Therefore, settlement boundaries are mapped using GPS coordinates around the settlement’s perimeter. Even though the data is not validated through official sources, they provide a good overview of the locations of slum settlements.

In contrast to these time-consuming and possibly inconsistent approaches done by many different people, slum mapping approaches using remote sensing data and techniques have proven immense potential. The morphologic appearance and patterns of slums measured by features such as building size, density or site characteristics are applied to proxy their location in these data (Wurm & Taubenböck, 2018). The data is processed either via visual interpretation by humans (Angeles et al., 2009; Gruebner et al., 2014), machine learning approaches (Wurm et al., 2017, 2019), or a combination of the two. These mappings have accuracies up to 92 %, making it a promising approach that also allows for globally consistent data to be generated (Kuffer et al., 2016). However, slum mapping is by no means simple or unambiguous. Even census maps of slums from government organizations of the same country sometimes contradict each other. For example in Pedro et al. it was shown that for São Mateus in São Paulo even census maps of slums obtained by the municipality do only coincide in about 59 % of slums (Pedro & Queiroz, 2019). In Kuffer et al., different experts were asked to identify the same slums from satellite imagery, and even here, substantial differences in the slum delineations emerged (Kuffer et al., 2018). In addition to the ambiguity on the same target, there is another difficulty. Different organizations, especially government statistical organizations, that are tasked with censuses in their countries have differences in their slum definitions. The minimum number of households forming a slum community already varies strongly between different countries. In Bangladesh, five households can already form a slum community (Bangladesh Bureau of Statistics, 2014), while in Brazil it is 51 households (Pedro & Queiroz, 2019) and in India it is a population of more than 300 inhabitants (Office of the Registrar General, 2011), which corresponds to about 60–70 households. Another factor affecting the definition of slums is the minimum distance between two slums. Dependent on the defined distance between settlements, the number of slums in a city varies and thus slum distributions (e.g. size or population counts) might differ. This factor was observed for the size distribution of slums in Friesen et al., (2019). These ambiguities thrive for a globally accepted definition of slum communities, demanded by Kuffer et al., (2018). The information on slum locations is essential, since only after knowing where slum settlements are located within a city, the question of their population can be answered.

1.2. How many people live in slum settlements?

Population censuses are produced uniformly for each country and have varying degrees of spatial granularity. They tend to be the most comprehensive data on population distribution in a given country. However, the censuses themselves usually do not provide information on slum populations; notable exceptions are Bangladesh, Brazil, and India. Rather, censuses systematically underestimate this particular group of the population. In Thomson et al., 2021 a broad literature review is conducted that shows that population in slums is underestimated on average by about 46 %.

Besides censuses, there are also local studies or reports from NGOs and scientific literature, among others, which give information about the population distribution in slums. For a slum community e.g. in Dhaka Khalequzzaman et al. conducted a complete household survey and thus showed that population estimates based on census data are significantly lower than actual values (Khalequzzaman et al., 2017). Instead of conducting complete counts or samples, projections based on physical properties seem to be an improvement. The KYC campaign mentioned earlier also provides population counts for the slum settlements mapped (Slum/Shack Dwellers International, 2017). These are obtained by counting all front doors and then sampling every nth household to estimate the average household size. The resulting estimate of slum population is then discussed in an open community forum (Thomson, Gaughan, et al., 2021). Similarly, in Taubenböck et al. the average number of inhabitants per household is obtained in a micro-survey. The number of inhabitants is scaled for the entire city of Mumbai, taking into account the number of slum households (Taubenböck et al., 2015). The resulting population counts are compared to literature values yielding a wide range of possible slum population counts in Mumbai’s slums.

However, without further processing micro-surveys might not be representative and too time-consuming for larger areas and thus need to be carefully used for projections to the whole city. Thus, more elaborate methods are needed.

1.2.1. Gridded population datasets

Gridded population datasets generally provide population counts on a uniform, comparably fine grid. To obtain population counts with spatial distribution, two common approaches exist: top-down and bottom-up. Both rely on auxiliary datasets, e.g. obtained by the use of remote sensing (e.g. nighttime lights or classification of settlements) in order to provide information about built-up areas, and thus are proxy attributes for assessing where people may reside. Both approaches differ only in the population input data. So-called bottom-up approaches are based on microcensuses using geo-statistical relationships between population density and auxiliary spatial datasets (e.g. night time lights are thought to provide an estimate where people reside). These relationships are then used for predictions about population in unsampled areas. Top-down approaches on the other hand are more commonly used to obtain spatially distributed population estimates. Therefore, census data is disaggregated using auxiliary spatial (or even spatiotemporal) datasets (Thomson, Rhoda, et al., 2020). One common global gridded population dataset used is the WorldPop-Global with a resolution of approximately 100x100 m. Although a number of other different datasets are known (GHS-Pop, LandScan, GPWv4), analyses have shown that WorldPop has a high accuracy (Archila Bustos et al.). In Section 2.1 the WorldPop dataset used in this paper, is explained in more detail. Whatsoever, we do not perform a detailed analysis of errors appearing in the creation of WorldPop-Global and focus on the plausibility of these global accepted datasets, when it comes to the specific but important urban class of slums. For systematic analyses regarding the WorldPop-Global dataset and its creation, we refer to (Leyk et al., 2019; Archila Bustos et al.; Thomson, Gaughan, et al., 2021).

1.2.2. Using WorldPop-Global to estimate slum population

To estimate the number of people in living in slums in different cities and to analyse they size distribution of the population, the WorldPop-Global dataset is used in this paper together with a dataset of mapped slums. The mapped slums are obtained according to Taubenböck et al. (Taubenböck et al., 2018). To obtain slum population estimates, the mapped slums are used to clip the WorldPop-Global population grid. A similar approach to the one presented in this paper has already been used in (Thomson, Leasure, et al., 2021; Thomson, Leasure, et al., 2021; de Mattos, Agatha, & McArdle, 2020). In particular, the use of the WorldPop-Global dataset to determine the population in slums poses special challenges. It was shown in (Friesen, Taubenböck, et al., 2018) that slums occupy a global mean area of 1 ha. However, this also corresponds to the size of the WorldPop grid cells. This means that about half of the slums in a city still fall short of the granularity of the WorldPop-Global dataset. For this reason, the question arises whether the dataset is at all suitable for use in estimating slum populations. This question was answered in Thomson et al. for certain slums in Nairobi in Kenya as well as for Lagos and for Port Harcourt in Nigeria using the KYC data as reference for validation (Thomson, Leasure, et al., 2021). The study concludes that the WorldPop-Global dataset is inadequate to describe the slums studied, resulting in a systematic underestimation of the slum population. However, (Thomson, Leasure, et al., 2021) emphasises that the statement cannot necessarily be generalised to other types of slums or other cities, let alone other continents. In contrast, Hennigen de Mattos et al. showed that using an older version of the WorldPop dataset for São Paulo, slum population estimates were fairly accurate, with an error of only 6 % (de Mattos, Agatha, & McArdle, 2020). Yet, the ground-truthed data used for comparison was also the input data for the WorldPop-Global dataset and thus potential errors were neglected here.

In another study Thomson et al. use simulated but realistic population counts as input to the WorldPop-Global's approach to identify several errors solely induced by the approach itself (Thomson, Leasure, et al., 2021). The major source of error is found to be the use of average population densities from large administrative units. Thus, high population densities, typically found in slums, are smoothed out.

In addition, the question of slum geometries also plays a role regarding slum population counts. Divergent slum geometries also lead to divergent population numbers. This is impressively shown in (Patel et al., 2019). Here, the slum population in Bangladesh according to the Bangladesh Bureau of Statistics census data in the *Census of Slum Areas and Floating Population* is compared to those resulting from the slum definition according to SDGs. The SDG definition states a 7.5 times higher number of people in slums for Bangladesh.

The question of slum size or slum population distribution can be answered in many different ways. In Friesen et al. the slum sizes, i.e. their area, are plotted as a rank-size distribution, approximated by a log-normal distribution (Friesen, Taubenböck, et al., 2018). The distribution of slum sizes yields insights into similarity of intra-urban interaction across the globe. However up to now, the question of slum population distribution could not be answered on a global level, which would possibly also allow deeper understandings of intra-urban interactions, e.g. as in Pelz et al., (2019).

In this paper, (i) the datasets and method for obtaining slum population estimates for cities around the world are described and (ii) applied to eight different cities. Then (iii) the resulting population distributions and counts are discussed and (iv) a broad comparison with literature values is performed in order to approximate the datasets' and method's plausibility. The organisation of the manuscript is visualised in Fig. 1.

2. Data and methods

In this section, first the used datasets are described and second the method for obtaining different slum population estimates is outlined.

2.1. Gridded population estimates

For obtaining population estimates, one valuable source are gridded population estimates. In this paper, we use the WorldPop-Global datasets that are openly available online under www.worldpop.org (Worldpop, 2018).

The WorldPop-Global datasets are obtained according to (Lloyd et al., 2019) using a so-called "top-down" approach, where population counts by census enumeration areas are disaggregated into grid cells. WorldPop-Global is highly-modelled, using multiple covariate datasets (e.g. intensity of night-light emissions and urban settlement data) that are thought to provide information about where people reside. Census population figures are used together with covariate datasets to provide an estimate of the population per grid cell using the machine learning approach Random Forests (Breiman, 2001). The resulting dataset consists of $3'' \text{ x } 3''$ $(\approx 100x100\,\text{m})$ grid cells for the entire world, where each grid cell contains the number of residents in the respective area.

Furthermore, the datasets are created for the years 2000–2020 annually. This multi-temporality allows projections of slum populations beyond census years and also gives the possibility of analysing population changes over time, which will briefly be discussed in Section 3.3. In addition to the WorldPop-Global dataset outlined as above, a second WorldPop-Global dataset is available where each country's gridded population is projected to match the UN official population. In this paper, the not UN-projected datasets were used as these only rely on census information on a fine level.

Since the datasets are based on census data, they are dependent on the census' quality (Thomson, Gaughan, et al., 2021). This alone, but also the differences in covariates, result in global differences in the quality of the WorldPop-Global dataset. These differences are accounted for in Section 4.

2.2. Mapping morphological slums

In order to know where slums are located in cities, a second dataset is required. Therefore, a globally consistent classification method is used relying on the slum ontology introduced by Kohli et al., (2012). The

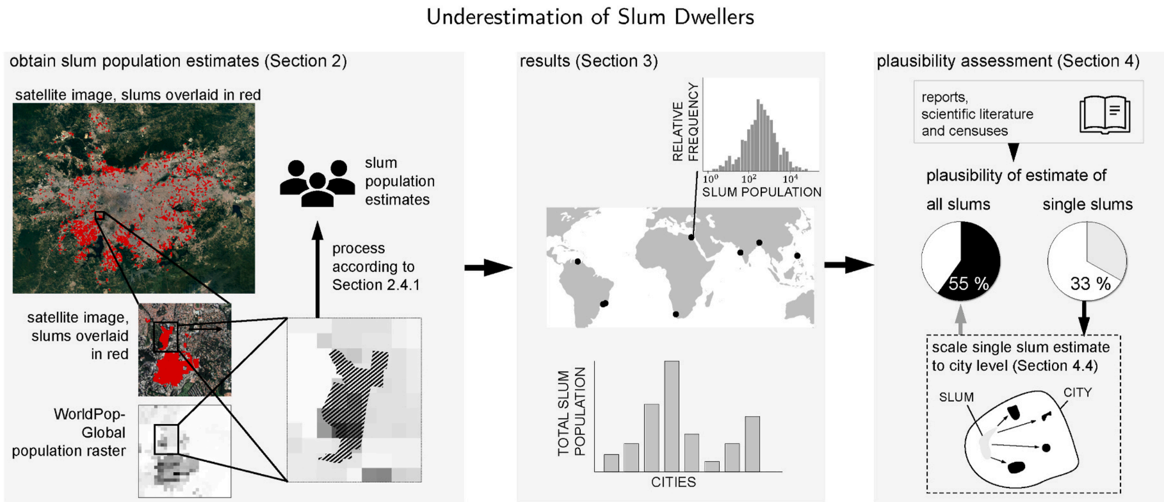


Fig. 1. Organisation of the manuscript.

locations of slums are determined based on the morphological differences in built-up structures to formal settlements according to the method described in Wurm & Taubenböck and Taubenböck et al. (Taubenböck et al., 2018; Wurm & Taubenböck, 2018). In this way, building structures of slums are conceptualized according to these approaches by high building densities, small and low building sizes as well as generally a high complexity of arrangement of buildings. Locally, these structures are evaluated relative to the rest of the built-up structure of the particular city and distinguished from formal structures, which are defined more by geometric arrangements or higher and larger buildings and better roofing materials. To do so, very high-resolution optical satellite data are used in combination with street view imagery and fieldwork. The dataset is produced by a computational approach followed by human visual inspection. In this paper, slum areas resulting from this approach are referred to as morphological slums, accounting for their physical slum-like appearance discernible by Earth Observation methods rather than slum definitions, e.g. by UN-Habitat (UN-Habitat and Slums, 2007). In terms of spatial granularity, slum-mapping was performed on the level of the city block with an average area of 1.6 ha.

As discussed above there are various approaches to spatially map slums. For checking consistency and evaluating the plausibility of approaches, we apply a different dataset of mapped slums in Section 4.5. The Brazilian census, conducted by the *Instituto Brasileiro de Geografia e Estatística* (IBGE) in 2010 provides census administration areas. Here, the smallest administration areas are sectors (portuguese: setor). Slums either form one sector or are divided into multiple sectors. The IBGE provides datasets of all slum sectors, thus a sector level slum map. A detailed comparison of these slum maps from IBGE and the remotely sensed slum morphologies used in this study was performed in Wurm and Taubenböck (Wurm & Taubenböck, 2018).

2.3. Study site selection

In this paper, we investigate slums in eight cities, namely Caracas (Venezuela), Rio de Janeiro (Brazil), São Paulo (Brazil), Cape Town (South Africa), Cairo (Egypt), Mumbai (India), Dhaka (Bangladesh) and Manila (Philippines).

The cities were chosen as a variety of large cities in the Global South. They differ in geographical location, cultural and economic influences as well as geographical topologies. However, Taubenbö et al., (2019) demonstrated that slums display common features independent of their context. Although the classification of 'slum of the slums' may be debatable, the framework used provides an opportunity to create a uniform database for different cities worldwide. This offers the opportunity to compare statistics between these cities. Apart from this, all eight cities contain a variety of slums making them relevant for analysis of a population distribution. Concerning the eight cities, population number and population growth are shown in Table 1.

In Cape Town, a large number of people living in slum-like conditions are residing in townships. These townships deviate from the category of morphologic slums from Taubenböck et al. (Taubenböck et al., 2018) and can therefore not be unambiguously classified with the EO-approach from Section 2.2 (Friesen et al., 2018). Nevertheless, Friesen et al., (2018) showed, that townships seem to be an integral part of areas of urban poverty in Cape Town. To account for this, we performed our analyses with two data sets: one with and one without townships to see to what extent this choice influences our results.

Dhaka is one of the fastest growing cities in the world (UN Department of Economic and Social Affairs, 2018), thus having high population changes, especially in slum communities. According to Angeles et al. the number of slum communities increased by about 70 % from 1996 to 2005 (Angeles et al., 2009). This makes it particularly difficult to obtain reasonable slum population counts for years that are not within the census, as projections in time naturally are more erroneous when population changes are high. As the dataset derived in Angeles et al., (2009) and the morphologic slum dataset used in this study have been produced with a temporal lag of 10 years, the former is not included in the plausibility check. The possible deviation was considered as too high, especially considering the high rate of slum growth in Dhaka. Yet, another dataset on slum population in Dhaka exists. The Census of Slum Areas and Floating Population 2014 (Bangladesh Bureau of Statistics, 2014) provides population counts for all slums. While it is a census dataset, it was not used as input for creating the WorldPop-Global dataset and is thus independent from it. It is used in Section 4.5.

An adjustment is required to check the plausibility of the total

Table 1
Population (in 2015) and population growth (from 2005 to 2015) of the eight cities studied (United Nations and Department of Economic and Social Affairs, 2018).

city	population in mio.	population growth
Caracas	2.9	1 %
Rio de Janeiro	11.8	9 %
São Paulo	18.3	14 %
Cape Town	3.2	29 %
Cairo	15.2	24 %
Mumbai	17.3	12 %
Dhaka	12.3	43 %
Manila	10.8	20 %

population figures of Cairo in section 4.1. The focus of this study is on morphologic slums that fit to the slum definition of UN-Habitat. In Cairo however, all literature sources found on total slum population include a broader definition of slums or informal settlements. Therefore, a plausibility check of only the morphological slum population is not possible due to lack of data. However, since many sources refer to the broader definition of Cairo slums, the total population figure is still an important comparative value. Therefore, in Section 4.1 we extend the definition including all types of slums according to Sims, (2003). A more detailed discussion of the different slum types and implications to the number of slums investigated can be found at Friesen et al., (2019).

2.4. Methods

In this section, first the method for population estimates on single slum level is outlined. Second, necessary statistical measures are defined for the plausibility check in Section 4.

2.4.1. Obtaining slum population estimates

By using the combination of the WorldPop-Global dataset and the morphological slums, that are both created uniformly for the cities studied, it is possible to derive slum population estimates.

The procedure is as follows, the main steps are visualised in Fig. 2:

(i) resampling raster cells

Since the cell size of the WorldPop-Global dataset is about the same size of the average area of a slum (Friesen et al., 2018), the cell granularity must be reduced to be able to map the slum population with the grid. For this purpose, each cell and its population is evenly divided into 256 cells (16 × 16 pixels). This is done under the assumption of equally distributed population throughout each cell. The resizing factor of 256 is the result of a parameter study, see Appendix A.

(ii) crop resampled raster cells to slum shapes
(iii) compute population count

Each slum's raster cells are aggregated to obtain the population of the entire slum. Slum populations of less than one person are excluded.

2.4.2. Statistical measures

For the analysis of results in Section 3 and the plausibility check in Section 4, statistical measures are necessary.

To quantify a cross-scale distribution, the geometric mean (μ) is an appropriate measure (Friesen et al., 2018). It is defined by

$$\mu = \sqrt[N]{\prod_{j=1}^{N} \widehat{p}_j} \tag{1}$$

with $\widehat{p}_j$ being the population estimate of the j-th slum and N being the total number of slums in the city.

For comparing slum population estimates in Section 4, accuracy statistics are introduced. The Median Fraction (MF) indicates the degree of over-/underestimation of population count estimations. It is defined as

$$\mathrm{MF} = \mathrm{median}(f_i) = \mathrm{median}\left(\frac{\widehat{p}_i}{p_i}\right), \tag{2}$$

with f_i being the fraction of slum population estimate $\widehat{p}_i$ and slum population count according to literature values p_i. The MF can be interpreted as a measure of how closely the estimate meets the literature values, providing a meaningful measure of deviation to the literature values.

In the case of both, overestimation and underestimation, for one dataset, the MF value may be misleading because the overestimations and underestimations may cancel each other out. For this reason, a new measure is introduced, the Median Fraction Absolute Difference (MFAD)

$$\mathrm{MFAD} = \mathrm{median}\left(1 - \frac{|\widehat{p}_i - p_i|}{p_i}\right). \tag{3}$$

This measure works unaffected for simultaneous under- and overestimation and gives a measure of error. MF and MFAD are equal if the literature values are either over- or underestimated for all values compared.

Note that statistical measures like Root Squared Mean Error or Mean Absolute Error are not taken into account, as they don't provide sensible estimates on cross-scale data.

2.4.3. Plausibility of slum population estimates

A validation using absolute values of the WorldPop-Global dataset for slums is challenging or almost impossible. In contrast to many countries of the Global North (e.g. Sweden (Archila Bustos et al.)), ground truth information to validate the data is rarely available or outdated. Against this background, we will perform - as other studies have done - a check of plausibility (Taubenböck et al., 2015). Thus, in

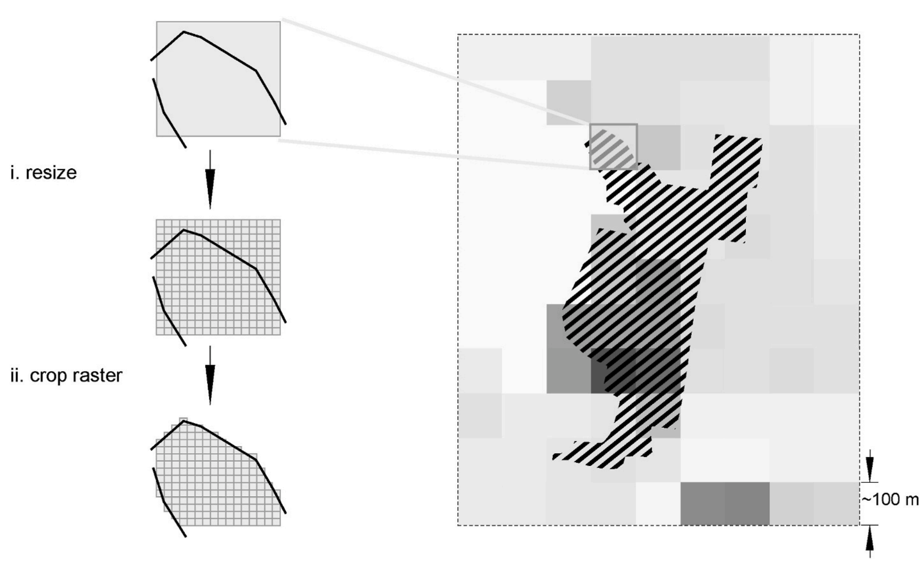


Fig. 2. Overview about the method used to obtain slum population counts and slum areas.

Section 4, the resulting slum population estimates are compared and their plausibility is checked with slum population data from the literature, including census data, academic literature and reports (e.g. from NGOs).

Therefore, we searched for names, locations and the spatial extent of specific slums within the eight cities investigated. We used common data bases, such as PubMed, WebofScience or Google Scholar for publications containing information on the population living in specific slums. Additionally, we categorised the information on population in four classes: (i) estimation with unclear source, (ii) modelled number of population (e.g. using disaggregation techniques), (iii) community-based estimate of population and (iv) survey data. We then searched for the morphological slums in our datasets, representing these slums and obtained the population estimates from the WorldPop dataset to compare both values.

In most cases, if no specific information on the spatial extent of the specific slum is available, the morphological slum areas are used without further adjustment. In some cases, however, a slum area is split, e.g. Rajiv Gandhi Nagar as part of Dharavi in Mumbai, or several slums are combined into one slum region, e.g., for Khayelitsha in Cape Town, or Kamrangirchar in Dhaka. Furthermore, the slum geometries used are not necessarily congruent with the slums found in the literature; rather, for many slums it was not possible to make a determination of the exact location because satellite imagery or geolocations were not given in the literature used. This testifies to the variability of slum classification schemes used by different scientific approaches or by different stakeholders. In these cases, affiliation was determined by the name of the slum in combination with research using Google Maps (NASA) and OpenStreetMap (OpenStreetMap contributors, 2017). One major source of field-references slum data is the KnowYourCity (KYC) campaign (Slum/Shack Dwellers International, 2017). Here 24 slum areas were taken for comparison. However, the process for obtaining KYC data is not uniform and introduces a bias toward rounded numbers, which can be seen simply by looking at the population figures.

As mentioned in previous sections, the morphological slums used were mapped for the year 2015 and therefore do not necessarily exactly match the slum morphology of other years. However, as before the WorldPop-Global dataset is used for the year that corresponds to the literature survey.

The population estimates are compared to literature values using MF and MFAD measures. This allows to check the plausibility of the obtained population estimates on single slum level as well as for the entire city's slum population.

2.4.4. *Scaling the slum population estimates*

The obtained single slum literature values can be used to scale the estimated total slum population. The resulting scaled estimates provide a range of possible total slum populations. The used method is described in Fig. 3. Here, the fraction of single slum population estimates over single slum literature values $f_{s,i}$ is calculated. The fraction is then used to scale the total population.

2.4.5. *Categorizing the errors*

Estimating slum population at the city level using WorldPop-Global and slum areas can lead to different errors. These errors can be divided into four categories:

- errors due to wrong slum geometries, e.g. not all slums are identified
- errors in WorldPop-Global input data, e.g. wrong population counts
- errors occurring in the WorldPop-Global approach, e.g. underfitting of the Random Forests algorithm
- errors due to our approach of resampling and clipping the WorldPop-Global dataset onto slum geometries

In order to isolate some of the four errors, in Section 4.5 we use different datasets available for certain cities and thus deviate from the otherwise uniform approach for all cities.

3. Results

In this section, the total number of slum dwellers, the slum population distribution, as well as the multi-temporal change of slum population are shown. Therefore, we use the outlined method to obtain slum population counts for all slums in the eight cities for the year 2015.

3.1. *Total number of slum dwellers*

First, the slum area locations and the WorldPop-Global datasets for 2015 are processed according to Section 2.4.1. This yields estimates for the population of each slum. Aggregating the number of slum dwellers

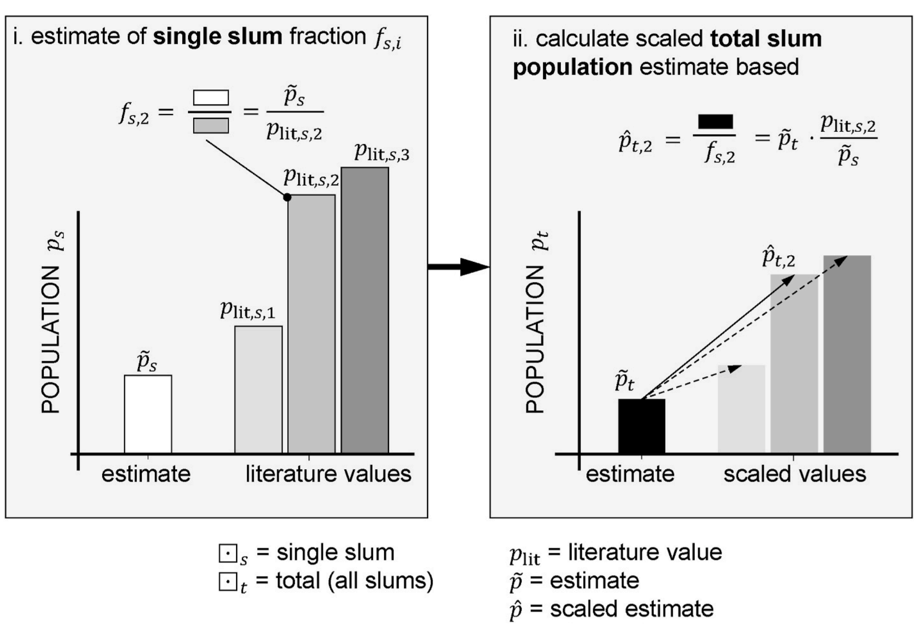


Fig. 3. Method used for scaling the total population estimate. First, the single population estimate is divided by a literature value, yielding the corresponding fraction. Then, the total population estimate is scaled by the fraction yielding a scaled total population estimate.

per city then yields the total slum population count for each city, as shown in Fig. 4. The highest number of slum dwellers, more than 1.5 million, is found in Mumbai, while the lowest number of slum dwellers lives in Cairo.

3.2. Slum population distribution

The resulting population counts on slum level can be used to create histograms showing the relative frequency of slum population counts for each city (see Fig. 5). To ensure that the population distribution of slums does not depend significantly on the defined minimum distance between slums, a systematic test is conducted. The respective analysis can be found in Appendix B and shows that there are no significant dependencies regarding the choice of minimum distance between slums.

All population distributions are cross-scale, which means that slum populations in cities span over multiple magnitudes. E.g. in Rio de Janeiro slums span from as low as 10 inhabitants to larger than 10,000 inhabitants. Because of comparably low number of slums, Caracas and Cairo do not form a smooth distribution.

The population distributions' means differ from 114 inhabitants in Dhaka to 1552 inhabitants in Cairo. In order to derive a global mean population of slums, the mean of all cities' geometric means is calculated, resulting in 558 inhabitants per slum.

Beside the population distribution, the population density distributions for the eight cities is obtained. Refer to Appendix C for results.

3.3. Slum population growth

In addition to estimating slum population counts for one year, our approach allows obtaining estimates for multiple years. Therefore, the WorldPop-Global dataset for different years is combined with the morphologic slum areas of the year 2015. This is based on the assumption that slums do only densify and their areas do not change significantly (Friesen et al., 2023). The resulting growth rates are shown for different cities in Fig. 6, each point corresponds to a slum. It is observable that while the growth in Caracas, Rio de Janeiro, São Paulo, Cape Town and Cairo are similar for all slums, cities like Manila, Dhaka and Mumbai show a wider range of population change with Dhaka averaging at nearly 50 % population growth within 10 years. Among the changes, in the latter three cities some slums experience a decline in population counts.

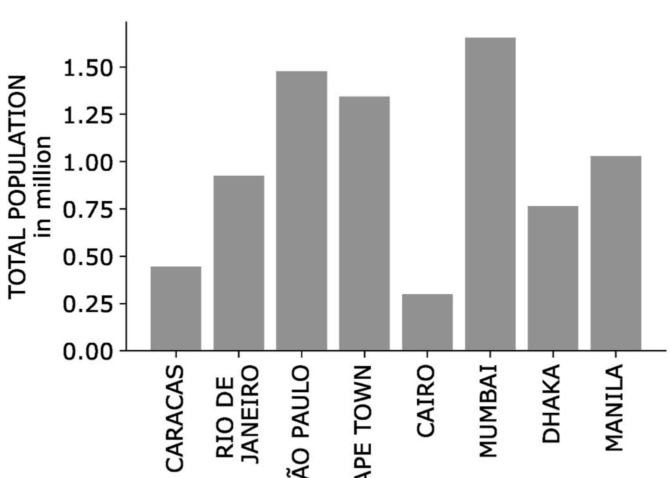


Fig. 4. Total slum population estimates for different cities according to the approach outlined in Section 2.4.1. The values are shown in Appendix F.

4. Plausibility assessment of slum populations

As slum population estimates inherently contain large amounts of uncertainty, the plausibility of the results is assessed. Therefore, first a comparison of total slum population estimates per city is performed, followed by comparing population counts of individual slums for all cities. These two plausibility checks include all four types of errors, even though on single slum level the error of wrong slum geometries is minor. This allows to examine the fitness of use of our approach and in the case of individual slums even of the WorldPop-Global dataset. Lastly, further plausibility checks are performed that allow exclusive consideration of some of the errors.

The censuses used as input to create the WorldPop-Global dataset are not considered at single slum level to avoid circularity.

4.1. City level

To check the plausibility of the slum population at the city level, the approach from Section 3.1 is used. However, to compare the values with the slum population from the literature as accurately as possible, the following datasets are used:

- the morphological slum data for the year 2015 - no other datasets are available here and
- the WorldPop-Global dataset that matches the year of the literature value.

Hence, we make use of the multi-temporal availability of the WorldPop-Global dataset.

The resulting comparisons are shown in Fig. 7. Sources, measures for the accuracy and more details of the literature values used are given in Appendix D.

The results indicate major population underestimations in most of the cities, especially in Caracas, Cairo, Mumbai, Dhaka and Manila. Only for Cape Town the number of slum dwellers estimated by our method is approximately equal to the literature population count.

4.2. Single slum level

In addition to the total slum populations, the population of individual slums are used for plausibility checks.

For each city, a map highlighting the slums used for the plausibility assessment is given. A total of 80 slum populations from the literature were used for the assessment. For Rio de Janeiro, Cape Town and Dhaka, refer to Fig. 8. For the remaining five cities, refer to Appendix E. The comparison of single slum populations is found in Table 2.

These single slum plausibility checks show that almost everywhere the WorldPop-Global dataset disaggregated onto slum areas is underestimating the slum's population. Only in Cape Town several overestimations of slum population occur. In particular, the variations in the $f_i = \widehat{p}_i/p_i$ ratio are substantial across but also within all cities. Since in Cape Town townships do not belong to the group of morphological slums, it makes sense to differentiate between them. Cape Town's townships are therefore referred to as (Twp.). One exception is Khayelitsha, which is largely a township, but where several informal settlements have grown up alongside the formal township settlements. Both types of settlement are referred to here as Khayelitsha. The settlements analysed within Khayelitsha, namely K2, Lindelani Park and AT5, are informal. The MFAD for both types of settlements are rather different, with 68 % in townships and 24 % in informal settlements.

For Ezbet El-Haggana in Cairo, our approach mostly underestimates the population size compared to literature values. However, there is one major outlier. An official estimate for the year 2000 estimates a population of 412. This drastically underestimates the value compared to other population figures reported in the literature for the same slum. For a detailed explanation of this systematic underestimation, see Sabry,

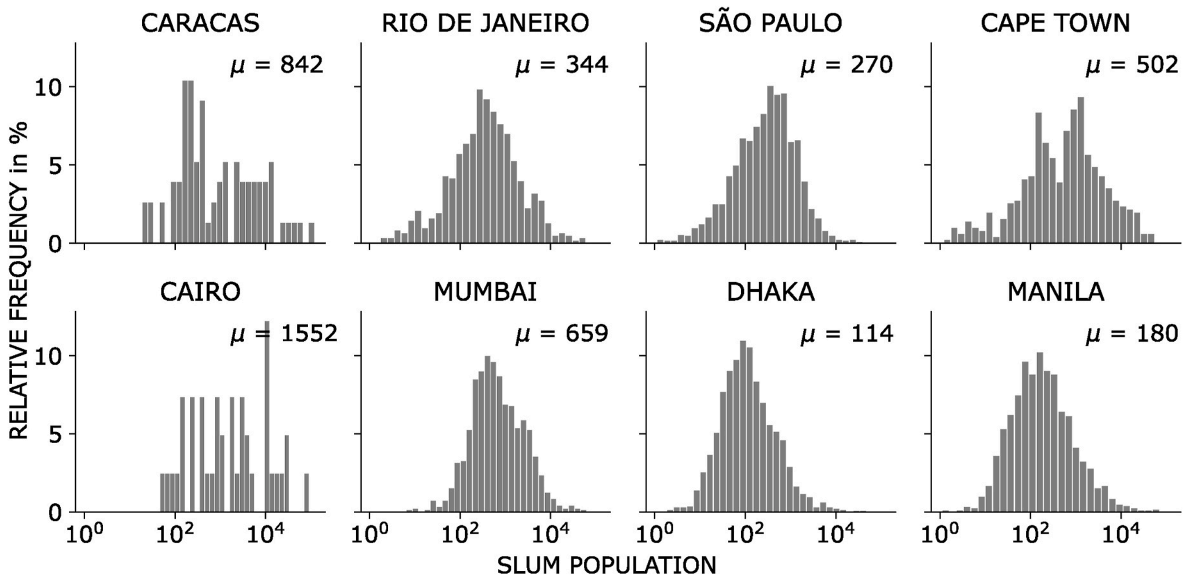


Fig. 5. Relative frequency of slum population for eight different cities. μ refers to the geometric mean of the population estimates.

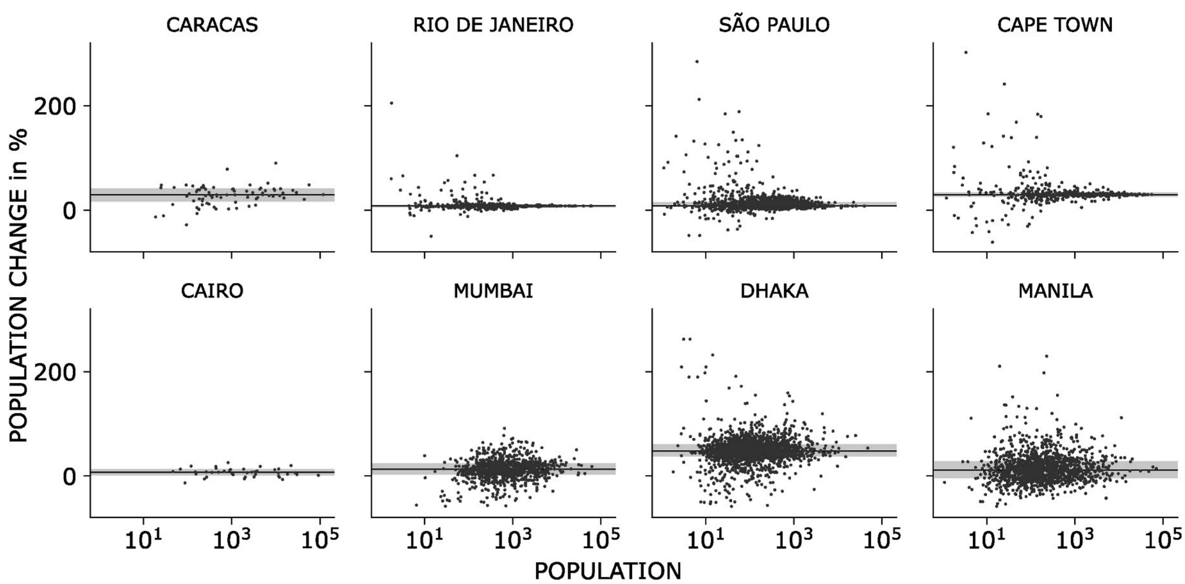


Fig. 6. Scatter plot of slums in eight different cities. Plotted is the population of the slums in 2015 and their respective relative change in slum population from 2005 to 2015 relative to the population in 2005. The black line indicates the median of the population change, while the grey band indicates its quartiles.

2009.

While the population for Dharavi slum in Mumbai is significantly underestimated, it seems accurate for one of the slum's neighborhoods called Rajiv Gandhi Nagar. An explanation for this cannot be given based on the data.

There is no correlation between the quality classes of the literature values and the accuracy of the estimates. For quality class (i) fractions range between 2% and 118%, for class (ii) fractions range between 8% and 3873%. For class (iii) between 3% and 237% and for class (iv) between 6% and 105%. For all quality classes, the MFAD values obtained for all estimates in that class range from 20 % to 27 %.

4.3. Plausibility of the estimates

The plausibility of the total as well as the single slum population estimates can be expressed in MFAD values for each city. Fig. 9 visualises these MFAD values for total and single slum population for the eight different cities.

The results show that Rio de Janeiro and Cape Town have the highest MFAD values with 34–59 % for single slums and 50–90 % for the total slum population, respectively. São Paulo and Manila have poorer MFAD values of 26 % on single slum level. In Mumbai, Dhaka, Cairo and Caracas the MFAD values for single slums are only between 4 and 13 %. Differentiating between Cape Town's township and its informal settlements, yields a MFAD of 68 % for the former and only 24 % for the latter.

Summing up all of the results presented above, a clear bias in the population estimation in slums - mostly a strong underestimation - is observed when WorldPop data is used as source for population figures. The values derived from WorldPop present a reference, which can be adjusted using the single slum estimates.

4.4. Scaling population estimates

The preceding plausibility check of single slums provides

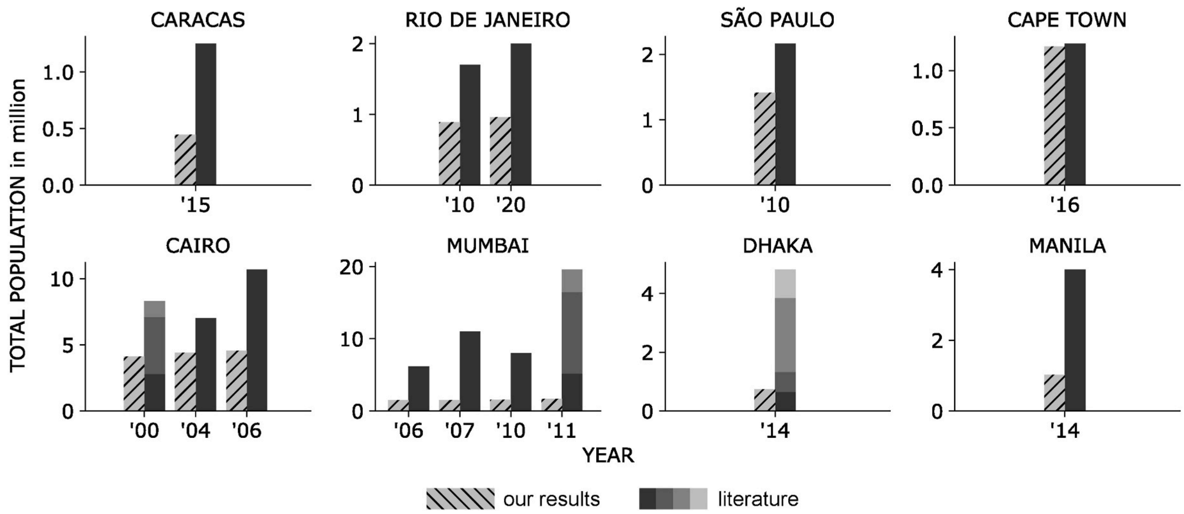
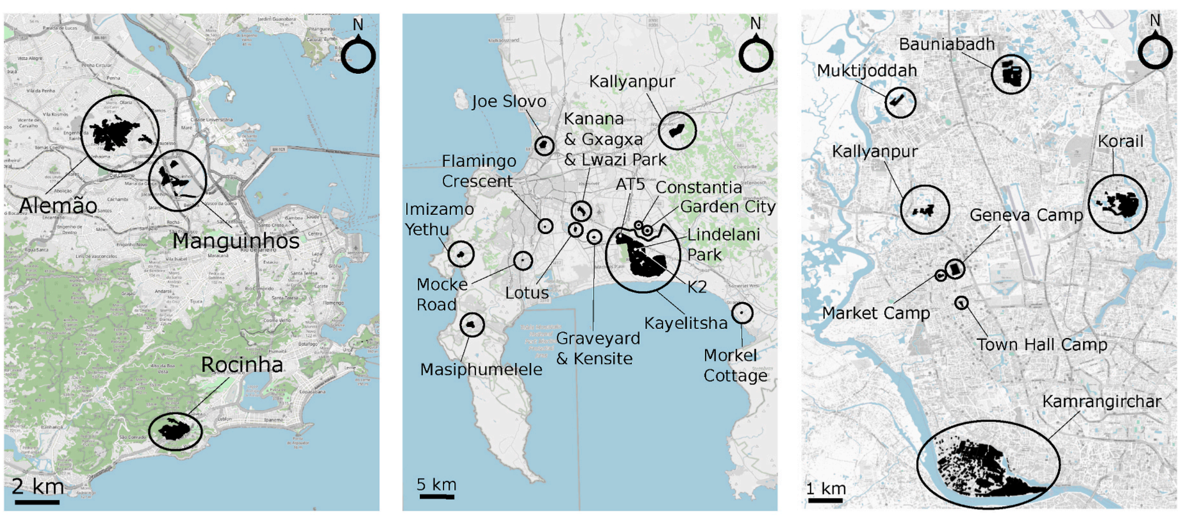


Fig. 7. Total slum population counts of all eight cities obtained by our method and from different literature sources. Different shades of grey define different literature values for the same year. All years refer to the 21st century.



(a) Rio de Janeiro (b) Cape Town with Townships (c) Dhaka

Fig. 8. Maps showing the location of slums in three different cities whose population figures have been checked for plausibility. Water and green landscapes are colored only for contrast. Remaining five cities' slum locations are shown in Appendix E.

information on how much our approach under- or overestimates the population. However, it also allows to take a further step. The single slum results can be used to scale the total population estimates shown in Fig. 4 according to the method described in Section 2.4.4. The resulting scaled estimates in comparison to the reference estimates and the literature values are shown in Fig. 10.

As every city's literature slum population, the scaled estimates present a wide range of total population counts. However, in the latter the ranges are spanning over one magnitude for several cities. The boxplot indicates several outliers for Cape Town, but also for Cairo, Rio de Janeiro and Dhaka. For Caracas, São Paulo and Cape Town the scaled estimates generally tend to be significantly larger than the literature slum population. In contrast, the scaled estimates for the remaining cities are roughly in line with the literature on slum populations. However, scaling increases the range of possible total populations for each city.

4.5. Error analysis

In this section, we deviate from analysing slum populations for all cities, because the data situation in certain cities allows us to isolate some of the four errors mentioned in Section 2.4.3.

The error due to wrong slum geometries cannot be analysed, since a consistent ground truth of slum locations would be needed. However, a comparison with the census geometries can be made, in order to examine the uncertainty between morphological slums and official census slums as an example. Therefore, morphological slum areas are compared to census slum areas for Rio de Janeiro and São Paulo, as these are the only two cities from our sample providing both datasets. Then, the percentage of overlap of the two shapefiles is calculated, resulting in an overlap of only 27 % in São Paulo and even only 18 % in Rio de Janeiro.

The two errors, on the one hand how WorldPop-Global disaggregates the population figures, and on the other hand how we process the WorldPop-Global data with our approach, can be considered isolated from the others by taking data from the Brazilian census. This allows to

Table 2
Results of single slum analysis. The first two columns specify city and name of the slum. The fourth column contains population counts obtained by the method outlined in Section 2.4.1. The fifth column contains population counts found in literature. The sixth column contains the population fraction, a measure for the accuracy, by dividing column three by column four. The seventh column gives the corresponding class of information, according to Section 2.4.3. If individual cell entries are empty, the value above them applies. In Cape Town, if the settlement is a Township, it is denoted by (Twp.).

city	slum name	year	pop our approach	pop literature	fraction in %	class	source
Dhaka	Kallyanpur	2005	1734	15,000	12	(iv)	Islam et al. (2006)
		2015	2307	8129	28	(iv)	Latif et al. (2016)
	Muktijoddah	2005	3537	17,750	20	(iv)	Islam et al. (2006)
	Bauniabadh	2014	2749	34,179	8	(iv)	Khalequzzaman et al. (2017)
	Market Camp	2014	807	1979	41	(iv)	Bangladesh Bureau of Statistics (2014)
	Geneva Camp	2014	5101	18,197	28	(iv)	Bangladesh Bureau of Statistics (2014)
	Town Hall Camp	2017	789	5000	16	(i)	Fattah and Walters (2020)
	Korail	2005	5990	101,000	6	(iv)	Islam et al. (2006)
		2009	7152	86,000	8–4	(iii)	Mridha et al. (2009)
		2009	7152	182,000	4	(iii)	Mridha et al. (2009)
		2011	7798	80,000	10	(i)	(Pramnik et al.)
		2014	8178	36,719	22	(iv)	Bangladesh Bureau of Statistics (2014)
		2017	9291	100,000	9	(i)	Fattah and Walters (2020)
	Kamrangirchar	2005	24,647	265,000	9	(iv)	Islam et al. (2006)
		2010	30,823	400,000	8	(i)	Ali et al. (2013)
São Paulo	Paraisópolis	2020	48,907	100,000	49	(i)	(Savarese et al.)
	Heliópolis	2006	31,254	120,000	85	(i)	(Wade)
		2018	34,239	200,000	17	(i)	Dicionario de Favelas Marielle Franco (2020a)
Rio de Janeiro	Rocinha	2010	58,063	73,410	79	(iv)	Cavalcanti-Ferreira et al. (2016)
				120,000	48	(i)	Dicionario de Favelas Marielle Franco (2020b)
				150,000	39	(iii)	Dicionario de Favelas Marielle Franco (2020b)
				250,000	23	(i)	(Staff)
	Manguinhos	2010	19,979	31,535	63	(iv)	Cavalcanti-Ferreira et al. (2016)
		2020	21,604	36,000	60	(i)	Dicionario de Favelas Marielle Franco (2020c)
	Alemão	2010	52,657	89,912	59	(iv)	Cavalcanti-Ferreira et al. (2016)
Caracas	Catia Sur	2016	64,020	150,000	43	(i)	Falco et al. (2019)
	Corridor 23 de Enero - San Juan	2016	23,314	130,000	18	(i)	Falco et al. (2019)
	Los Erasos	2007	278	5000	6	(iii)	Hulett et al. (2013)
	Petare	2009	42,534	0.6–1 mio.	7–4	(i)	(Badkar)
Mumbai	Rajiv Gandhi Nagar (Dharavi)	2000	5626	5700	99	(ii)	(Karn et al.)
	Dharavi	2003	53,194	350,000	15	(i)	Risbud (2003)
		2005	56,576	0.9–1 mio.	6	(i)	Ahuja and Brosius (2006)
		2011	60,018	300,000	20	(ii)	(Karn et al.)
				509,000	12	(ii)	Taubenbö et al. (2015)
Cape Town	Joe Slovo (Twp.)	2008	12,217	20,000	61	(i)	Jordan (2008)
	Masiphumelele (Twp.)	2004	15,572	15,000	104	(ii)	The Masiphumelele Corporation and Trust (2011)
		2005	15,843	12,703	125	(ii)	The Masiphumelele Corporation and Trust (2011)
		2008	17,323	14,593	119	(i)	The Masiphumelele Corporation and Trust (2011)
		2010	18,198	38,000	48	(i)	The Masiphumelele Corporation and Trust (2011)
	Wallacedene (Twp.)	2018	31,312	36,000	87	(i)	Verzoni (2018)
	Imizamo Yethu (Twp.)	2003	8450	8019	105	(iv)	Huchzermeyer (2010)
	Morkel Cottage (Twp.)	2002	61	357	17	(iv)	Huchzermeyer (2010)
	Mocke Road (Twp.)	2001	20	32	62	(iv)	Huchzermeyer (2010)
	Khayelitsha (mostly Twp.)	2006	261,947	400,000	65	(i)	(Chaffey)
		2016	338,198	2.4 mio.	14	(i)	(Chaffey)
		2017	346,897	400,000	87	(i)	Habitat for Humanity GB (2017)
		2018	355,776	0.5–1.5 mio.	71–24	(i)	(Parliamentary Monitoring Group)
	K2 (Khayelitsha, no Twp.)	2018	629	1855	34	(iii)	Slum/Shack Dwellers International (2017)
	Lindelani Park (Khayelitsha, no Twp.)	2018	159	1760	9	(iii)	Slum/Shack Dwellers International (2017)
	AT5 (Khayelitsha, no Twp.)	2018	267	336	79	(iii)	Slum/Shack Dwellers International (2017)
	Flamingo Crescent	2016	7	400	2	(i)	(Milne)
		2017	7	294	2	(iii)	Slum/Shack Dwellers International (2017)
	Lotus	2018	2552	10,500	24	(iii)	Slum/Shack Dwellers International (2017)
	Kanana	2018	6809	9000	76	(iii)	Slum/Shack Dwellers International (2017)
	Gxagxa	2018	270	900	30	(iii)	Slum/Shack Dwellers International (2017)
	Garden City	2018	1250	2160	58	(iii)	Slum/Shack Dwellers International (2017)
	Grave Yard	2018	949	400	237	(iii)	Slum/Shack Dwellers International (2017)
	Kensite	2018	70	288	24	(iii)	Slum/Shack Dwellers International (2017)
	Constantia	2018	96	1800	5	(iii)	Slum/Shack Dwellers International (2017)
	Lwazi Park	2017	109	320	34	(iii)	Slum/Shack Dwellers International (2017)
Manila	Baseco	2014	12,512	51,060	25	(iv)	Su and Mlčák (2014)
		2015	12,666	60,000	21	(iv)	Philippine Statistics Authority (2015)
	Gulayan P. HOA	2017	1561	2640	59	(iii)	Slum/Shack Dwellers International (2017)
				6000	26	(iii)	Slum/Shack Dwellers International (2017)
				9000	17	(iii)	Slum/Shack Dwellers International (2017)

(*continued on next page*)

Table 2 (*continued*)

city	slum name	year	pop our approach	pop literature	fraction in %	class	source
	Damata Kadama	2017	3630	8400	43	(iii)	Slum/Shack Dwellers International (2017)
				12,500	29	(iii)	Slum/Shack Dwellers International (2017)
				13,000	28	(iii)	Slum/Shack Dwellers International (2017)
	LDS Pioneer	2017	130	112	116	(iii)	Slum/Shack Dwellers International (2017)
	Letre Urban	2017	334	1760	19	(iii)	Slum/Shack Dwellers International (2017)
	Flordeliz Village	2017	1332	6825	20	(iii)	Slum/Shack Dwellers International (2017)
	Litre	2017	623	1820	34	(iii)	Slum/Shack Dwellers International (2017)
	Angela	2017	13	186	7	(iii)	Slum/Shack Dwellers International (2017)
Cairo	Ezbet El-Haggana	2000	15,957	412	3873	(ii)	Sabry (2009)
				400,000	4	(i)	Sabry (2009)
				1 mio.	2	(i)	Sabry (2009)
		2006	16,606	39,433	42	(iv)	Sabry (2009)
		2007	16,872	212,575	8	(ii)	Sabry (2009)

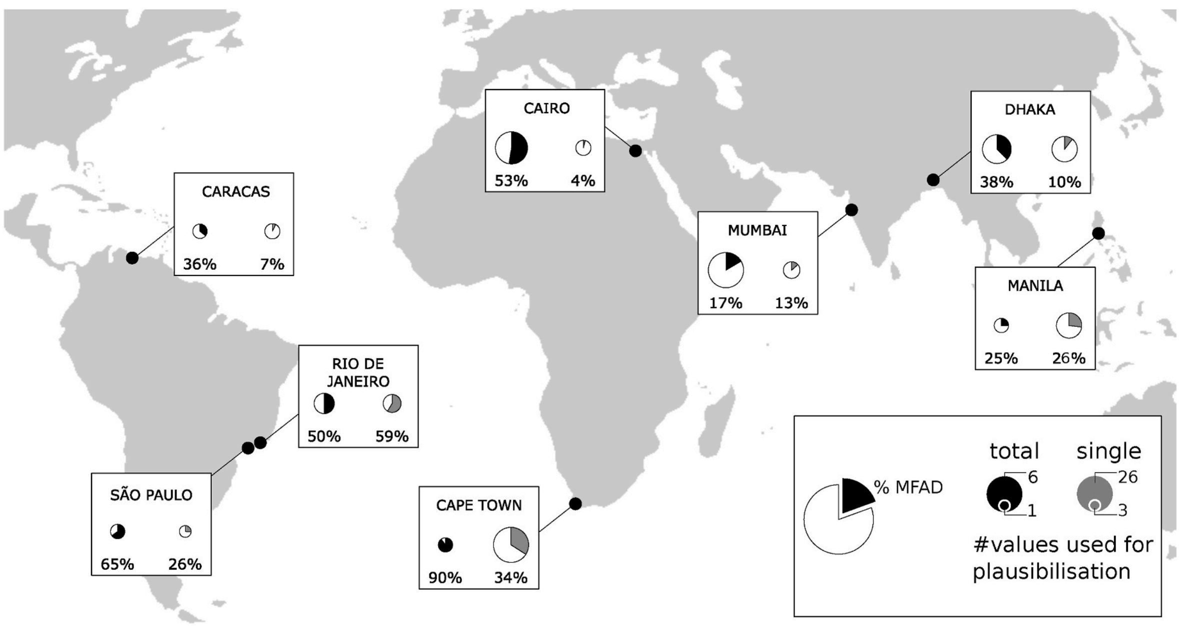


Fig. 9. Resulting MFAD values for the eight cities, each for the total city slum population (total, left pie chart) and for all single slums (single, right pie chart) evaluated. The size of the pie chart depends on the number of values used for the plausibility check, see legend in the lower right for size comparisons. Larger pie charts therefore mean more values used for the plausibility check. The total slum population of Cairo also includes non-morphological types of slums.

look at the two types of errors detached from potentially divergent slum geolocations or errorneous input to the WorldPop-Global dataset. The results are shown in Fig. 11. The MF and MFAD values rank between 85 % and 90 %, and 82 % - 77 %, respectively. This means that the two types of errors mentioned above together are responsible for a mean deviation of about 20 %. It is important to note that this analysis does not require any statement about the correctness of the input data, since it is purely an assessment of the two errors mentioned.

In Dhaka, for comparing the slum population on city and region level, the Census of Slum Areas and Floating Population 2014 (Bangladesh Bureau of Statistics, 2014) is used. The dataset does not provide explicit information about the slum locations. To overcome this, the slum's populations are aggregated by city districts allowing for a district-wide comparison with the data obtained in Section 2.4.1. Therefore, district names from the Global Administrative Areas (GADM) project (ADministrative Areas, 2021) are compared to those from the census and matching names are counted as the same regions. The result is an overestimation of the population with a MF of 149 % and a MFAD of 40 %. Hence, the results fit to that obtained for Dhaka in the previous sections.

Even without information on the location of the census slums, it's possible to generate a population distribution. Calculating the geometric means, the census slums have $\mu = 62$ inh., while the morphologic slums have $\mu = 110$ inh. With regard to the constant underestimations of the morphologic slums for the other plausibility checks, this result emphasises the underestimation of slum population by census authorities shown in Patel et al., 2019.

5. Discussion

It is remarkable to note that in a globally accepted dataset - when disaggregated to a particular structural type, slums, in our cities - a clear bias in the estimation of the population can be observed. In our study, we presented an approach to obtain slum population estimates that is similar to those used in other studies. Yet the major difference is that we were able to use the combination of two worldwide uniformly generated datasets, which allows for obtaining comparable population estimates for slums in eight cities of the Global South, located on three different continents. This approach proves that slum dwellers worldwide are extremely underrepresented in global datasets. We thereby confirm findings in small slum-specific case studies showing an underestimation of slum dwellers in global datasets (Thomson, Gaughan, et al., 2021).

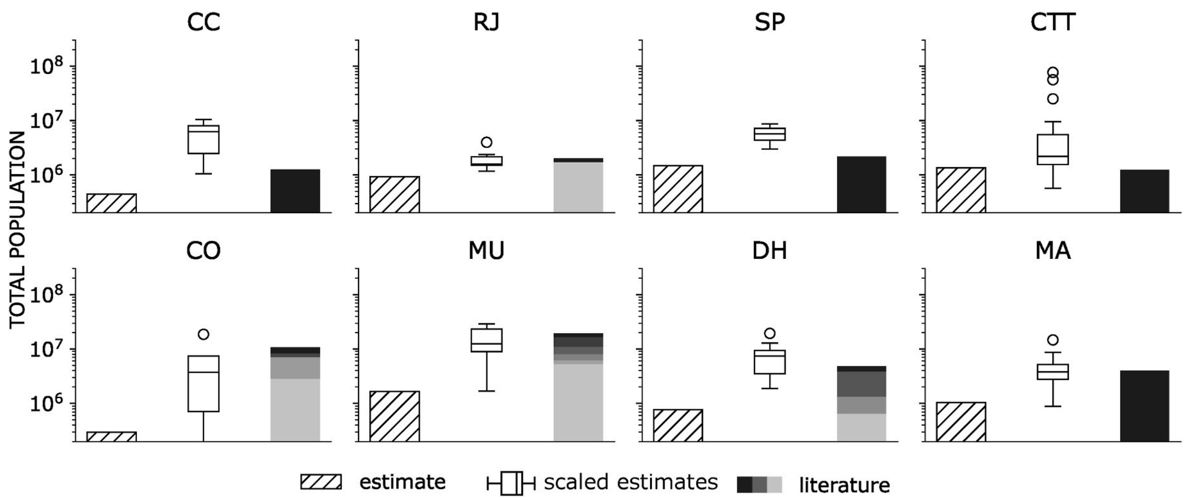
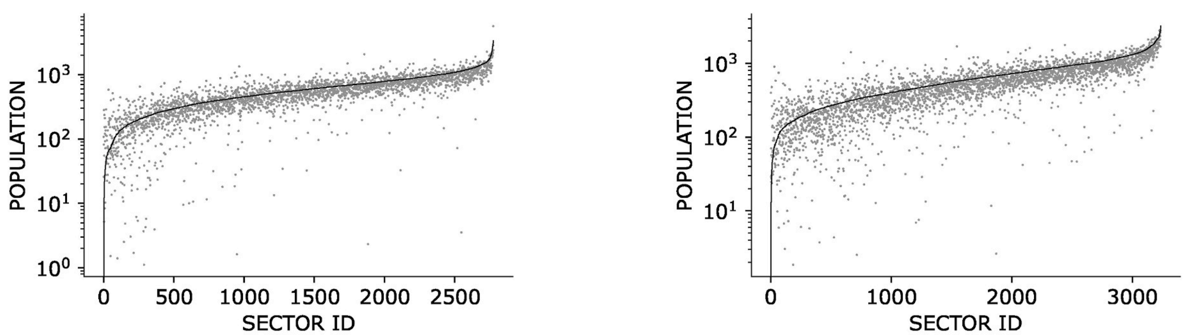


Fig. 10. Comparison of total slum population counts for each of the eight cities. The population estimate refers to the reference total population count obtained using the approach described in Section 2.4.1 for the year 2015, while the boxplot shows the corresponding range of scaled estimates. These scaled estimates are derived according to the method shown in Section 2.4.4 incorporating the fractions from Table 2. The literature values are the same as in Fig. 7, with the different years shown together for clarity. The values corresponding to the boxplot are shown in Appendix F.



(a) Rio de Janeiro. MF = 85%, MFAD = 82%

(b) São Paulo. MF = 90%, MFAD = 77%

Fig. 11. Sector level slum population comparison of all census slums in the according city. The sectors are sorted in ascending order by population according to census. The black line represents census population counts, grey scatters are population counts obtained by our approach.

According to our estimate, the mean number of slum dwellers differs globally by more than one magnitude. In contrast to the similar size of slums (Friesen, Taubenböck, et al., 2018), there seems to be no uniform population distribution around the world. We further obtained population changes of all slums, indicating the level of slum growths. However, the detailed analysis of slum change, despite its relevance for understanding temporal evolution, is outside the scope of this paper; rather, our purpose here is to provide a basis for discussion. For an analysis of the spatial and temporal changes of slum population, we therefore refer to (Breuer & Friesen, 2023) and (Friesen et al., 2023).

Accurate population counts and their spatial distributions are of utmost importance for informed decision-making. It is crucial to avoid incorrect population figures - especially underestimations - as they can lead to undue pressure on infrastructure and informal development within the city. The availability of accurate data is important for the efficient allocation of resources, including healthcare, education, and sanitation, as well as urban planning and development strategies. Population figures are particularly crucial in deprived areas to ensure adequate emergency response planning and representation in governance.

The importance of correct population figures makes the assessment of plausibility crucial in order to verify the suitability of our approach. Thus, a broad range of data is obtained for different cities, varying years and on different scales. This could theoretically reduce errors that appear when relying on one population dataset only as is the case for existing gridded population evaluations (Thomson, Kuffer, et al., 2020; Thomson, Gaughan, et al., 2021; Thomson, Leasure, et al., 2021). The global plausibility check of a gridded population dataset for a specific urban class done in this study is the first of its kind thus a valuable contribution to the question of applicability, again indicating that the research community has very limited information about population living in slums.

The performed plausibility check provides insights on how good the derived population estimates are. Therefore, they are compared with literature values on two different scales, resulting in two different MFAD values (refer to Fig. 9). The plausibility check reveals that the slum population is underestimated, sometimes drastically, on both scales for almost all datasets.

In Cape Town, where most single slum population comparisons were conducted, it appears that our method may not effectively capture the slum population. The MFAD for all slums in Cape Town is around 34 %. However, closer examination of the slum populations in this city shows significant variations. Differentiating between townships and informal settlements yields their distinct characteristics. As detailed in Sec. 4.3, township population estimates are considerably more reliable, with an accuracy of about 68 %, in contrast to just 24 % for informal settlements. This disparity is likely due to the different census methods used; these methods are more successful in counting populations in townships but less so in informal settlements, which present more survey challenges due to their unstructured layout. Although it's valid to compare the townships and informal settlements of Cape Town, research by Friesen et al. (Friesen, Taubenböck, et al., 2018) indicates that the development of townships significantly influenced the expansion and spatial distribution of slums in the city. Therefore, an analysis that excludes the townships and focuses only on informal settlements fails to provide a complete understanding of the situation.

The best fit to literature values is provided for Rio de Janeiro - and Cape Town's townships -, yet not yielding sensible estimates. However, as most of the slum population estimates are an underestimation, the results describe a reference slum population. This reference is used to scale the total slum population estimates. A similar scaling of individual slum population counts to obtain a total slum population estimate was performed before in (Taubenböck et al., 2015). Yet, in this study we use a broader set of single slum literature values. The outcome, however, is similar. The resulting boxplots in Fig. 10 emphasise that the uncertainty in the number of slum dwellers around the globe is excessive. Here, for most cities the range of possible slum population spans more than a magnitude. This uncertainty casts serious doubt on the official estimate of 1 billion slum dwellers worldwide. Based on our findings, it is likely that hundreds of millions of slum dwellers are not included in the numbers at all. For most cities this finding would increase the entire cities population numbers (see Table 1) – for some tremendously, e.g. in Caracas, where the median scaled population more than doubles the cities total population.

In Section 4.2 different quality classes for the literature slum population estimates are used. Here, no correlation between the quality class and the accuracy of the estimates is observed. Thus, the bias in estimating the slum population is independent of the quality of the literature values.

An analysis of the results shown in this work leads to the conclusion, that validation data to test the accuracy of gridded population data in slums is in urgent need. One possible way is to use mobile data like Steele et al. (Steele et al.) showed for Dhaka, Bangladesh. Another way would be systematically documented surveys, as Taubenböck and Kraff (Taubenböck et al., 2015) conducted for Mumbai or Abascal et al. (Abascal et al.) proposed on a bigger scale for several cities.

Nevertheless, since especially the latter approaches would require huge personal and financial resources even for a single city, a validation on a global level is not possible up to now.

5.1. Discussion of errors

To depict reasons for the severe underestimations, the errors are divided into four categories and some of them are analysed in isolation:

- errors due to wrong slum geometries, e.g. not all slums are identified
- errors in WorldPop-Global input data, e.g. wrong population counts
- errors occurring in the WorldPop-Global approach, e.g. underfitting of the Random Forests algorithm
- errors due to our approach of resampling and clipping the WorldPop-Global dataset onto slum geometries

We contributed to the question of wrong slum geometries. To quantify ambiguity for the morphologic slums used in this paper, we are able to compare slum geometries in São Paulo and Rio de Janeiro. Showing that the census slums compared with morphologic slums resulted in only 27 % and 18 % of overlap, respectively. So, in agreement with previous sources, there is large ambiguity in slum mapping. For uniform population estimates of slums around the world, which is important in order to reach SDG 11, Kuffer et al. outline: "first we have to agree on a clearer conceptualization of such [slum] areas […]" (Kuffer et al., 2018). However, in the study of Taubenböck et al. on morphologic appearances of poor urban areas across the globe, it has been shown that morphological variabilities for slums exist in different cultural parts of the world and that a global approach without considering the local context for delineation from "non-slums" is conceptually elusive (Taubenböck et al., 2018). Thus, a global slum ontology needs to be as uniform as possible but still require local adaption (Kuffer et al., 2016). Surely, the slum areas used in this paper are a notable application of such a slum ontology.

Another reason for the different performance of population estimates driven by wrong slum geometries is that slums in São Paulo, for example, are morphologically very distinct from their surroundings, while e.g. in Manila or Dhaka there is often a smooth transition between slums and non-slums (Kuffer et al., 2016). This clearer delineation may also contribute to a more concrete identification of the slums.

Generally, the reasons for wrong slum geometries or populations are manifold. Governments often want the population numbers to be small, because otherwise the pressure to provide socio-technical infrastructure is getting higher (Lucci et al., 2018). This could be a reason for different slum definitions across countries and global organizations, e.g. UN-Habitat. For Dhaka and several cities in India it was shown that using a slum definition accepted by the international development community rather than the country's official slum definition yields tremendous increase in slum population (Lucci et al., 2018; Patel et al., 2019) (see Appendix D for details on Dhaka). Another reason is that official censuses often only represent formal population counts, thus not counting informal or floating population (Agarwal, 2011; Lucci et al., 2018). In these cases, the WorldPop-Global dataset does not even consider slum dwellers as input population. Yet, another error arises from the fact that most gridded population datasets are only available on a yearly basis. Slums might have high population dynamics that thus can not be accounted for, for example by seasonal circular dynamics (Beguy et al., 2010).

It has also been mentioned that one major drawback of the WorldPop dataset is the dasymmetrical mapping of the population since water bodies are the only constraint. Unlike other datasets that only disaggregate their data up to built-up areas, the WorldPop dataset also disaggregates beyond them.

The estimation of population numbers incurs an error when resampling the data, as uniform density is assumed within WorldPop pixels. For instance, if one assumes that the values of a pixel represent the average of an area of both high and low densities, the low resolution of WorldPop implies that the population in that cell only represents the average value. As a result, resampling the data in this case leads to an underestimation of the actual population size in the respective slum.

Concerning the errors occurring within the WorldPop-Global procedure as well as our resampling approach, it was possible to quantify the error for São Paulo and Rio de Janeiro. In Hennigen de Mattos et al. the error of estimating slum population using a similar WorldPop dataset was only 6 % (de Mattos, Agatha, & McArdle, 2020). In accordance with our data (while the MF value for São Paulo is 90 %, the mean fraction is 89 %), the error of the combination of the two approaches is small in comparison to the overall error. While this is true for Rio and São Paulo, it probably cannot be applied to other countries. However, based on this research, large inaccuracies at single slum level in São Paulo are likely to be due to large discrepancies between census and literature values. Overall, in Brazil the IBGE census is comparably fine, resulting in very fine resolution training data. This matches the conjecture of Thomson et al. that fine resolution population training data could be one major reason for a good fit of input and output data (Thomson, Gaughan, et al., 2021). Comparing plausibility measures with census resolution for different cities could thus be interesting for future works shedding light on the underlying relation.

5.2. Limitations

One limitation of this study is the inconsistency in literature population counts used for plausibility assessment. Given the poor data availability in slums, this result is hardly surprising. As pointed out, the inconsistency of concepts and approaches for mapping slums and their population to date limits our knowledge of slums.

As for literature data of single slum populations, geometries are not necessarily given nor identical to the morphologic slum geometries. We only take account of slums that were assignable by their name, which limits the number of slums. The limited number of slums poses a problem, especially considering the distribution of the population across scales. For all eight cities, mostly population data of large slum communities is available. The reason for this might be that the majority of scientific studies preferably investigates large slums. For a plausibility check, population counts of slums at different scales would be important to determine possible differences induced by different scales.

Furthermore, investigations are performed based on morphologic slums observed for the year 2015. This limits the assessment of plausibility, especially for cities experiencing large growth in number of slum communities. Estimating the error introduced by incorrect slum geometries due to time inconsistencies is outside the scope of this study. Time-varying slum areas are only available for a few cities worldwide (examples for single cities include (Gruebner et al., 2014), (Kit & Lüdeke, 2013)). Up to now, the lack of globally consistent multi-temporal slum area datasets does not allow for an assessment.

Considering the growth rates obtained for the slums in all cities, it is undeniable that they are flawed by the fact that the same slum locations are used for different years without knowing exactly how the slums there have changed. In particular, the emergence of new slums and slum growth, as well as the respective opposite, cannot be captured with our method. Again, time-varying slum areas would be needed.

6. Conclusion and outlook

In the presented study, slum population estimates for eight different cities are obtained using a uniform approach. These estimates define a starting point for global comparison of total populations, population distributions and even population changes in slums.

However, there is a great deal of ambiguity and lack of knowledge about slum areas and slum populations. Although the datasets used in this study are far from perfect, there is currently no better data available on a global scale. According to our results, we find that the only globally available and widely accepted dataset with a yearly population estimation mostly underestimate the population living in slums when compared to other datasets. This clearly shows that in order to meet international goals, data needs to be greatly improved. Based on this, the current estimate of 1 billion global slum residents could ignore many if not hundreds of millions of people.

Different levels of underestimations were found for the respective cities and possibly underlying reasons discussed. These differences could be used in future studies to further identify the sources of errors in the approach. Yet, in order to make valid statements considering the accuracy, the number of single slum population data from the literature are too little for many cities. In the future, more micro-surveys are needed that aim at counting slum populations. Here, the KYC campaign by Slum Dwellers International could be a reasonable starting point (Slum/Shack Dwellers International, 2017). As showcased in Thomson et al., with greater and more reliable data availability, the WorldPop-Global dataset could be trained on such data to better identify slum populations (Thomson et al., 2021).

The scope of this study only scratches the surface of slum population changes or slum mobility investigations by using time-independent slum areas in dynamic environments. For proper planning of infrastructure, multi-temporal slum maps could be beneficial for insights into the dynamics of slums and thus help to create scenarios of where population growth will probably be the strongest (Friesen et al., 2018). This step would allow infrastructure to be planned for the future rather than for the cities of today.

In addition to these steps to improve slums worldwide, as described by UN Sustainable Development Goal 11.1, it is crucial to systematically analyse the current knowledge and statistics regarding slums and their populations, and to accurately reveal and quantify differences in estimates based on different data sources. We followed a standardised methodology to analyse the situation in the eight cities studied.

We are aware that counting slum populations does not improve just a single life situation. However, a realistic assessment of the situation in quantity and spatial distribution might raise awareness and provide decision-makers with better information for policies and intervention strategies.

CRediT authorship contribution statement

Julius H.P. Breuer: Conceptualization, Formal analysis, Methodology, Software, Visualization. **John Friesen:** Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing. **Hannes Taubenböck:** Conceptualization, Writing – review & editing. **Michael Wurm:** Conceptualization, Writing – review & editing. **Peter F. Pelz:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

A Resample Study

As outlined in Section 2.4.1, a resample study was performed to ensure that the resampled and cropped raster cells provide good geographical fits to the slum's geometry. The study was performed according to the method outlined in Section 2.4.1 but with different resample factors. Here, a resample factor of 4 resamples every raster cell by 2x2. The original cell's population is used as each resampled cell's population.

The study was performed for the slums of São Paulo using the morphologic slums. The results are shown in Fig. 12. A threshold of 95 % overlap was chosen, resulting in a resize factor of 256. This way a good compromise between computation time and resulting population accuracy is obtained.

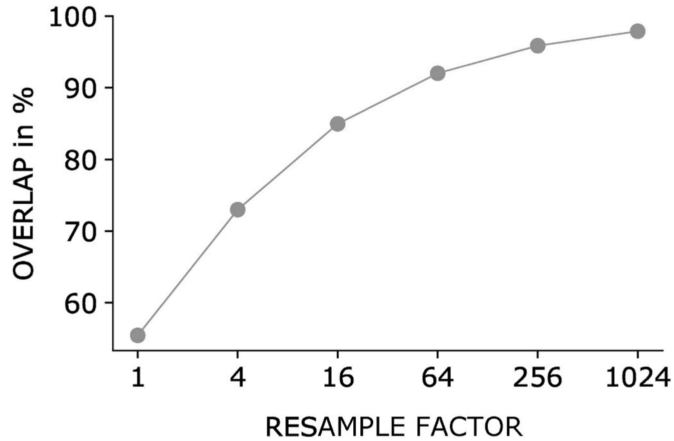


Fig. 12. Resample study performed with São Paulo slums. Each point shows the percentage overlap of a specific resample factor, lines are for visualisation only.

B Buffer Study

As outlined in the introduction, dependent on the defined distance between slums in a city, their number varies and thus the slum population distributions might differ. Thus, in order to investigate the impact of the defined distance between slums a study is performed. Here, "buffer" refers to a radial increase of the slum area. Therefore, each city's slums are buffered before step (ii) of the method outlined in Section 2.4.1.

The geometric mean of the resulting slum populations is calculated. The results are shown in Fig. 13, each graph represents one city. The influence of buffering does only have a significant impact on the geometric mean population, for buffer sizes as high as 20 or 50 m. Therefore, the slum population counts shown in Section 3 are independent of buffer size.

While mostly a positive correlation between buffer size and geometric mean population is found, sometimes an increase in buffer size resulted in a lower geometric mean population. This seems counter-intuitive, however the reason is that buffering slums results in less slums. If large slums are merged while small slums are not, e.g. because they are more isolated, the total number of slums decreases. In this case, the buffering leads to a decrease in geometric mean slum population.

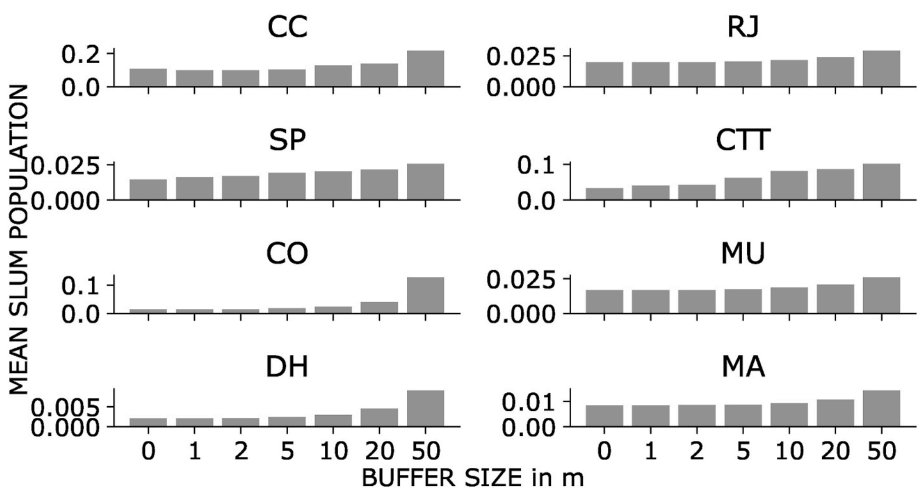


Fig. 13. buffer study.

C Slum Population Density

The method also allows to calculate population density distributions for the slums in the eight cities.

In order to derive the population density estiamtes, a pre-step to the method outlined in Section 2.4.1 is necessary. The WorldPop-Global dataset contains no-value cells that account for no population at all assigned to this region. This is the case e.g. for lakes or other structures that are excluded before the disaggregation of census counts. Calculating population density estimates makes it necessary to clip those novalue cells from the slum shapefiles in order to provide a proper area calculation of the shapefiles. The pre-step is achieved by polygonizing the raster, excluding all novalue cells and then intersecting the resulting shapefile with the slum shapefiles.

The resulting slum population density distributions are shown in Fig. 14. Compared to the slum population distribution, there are some major differences.

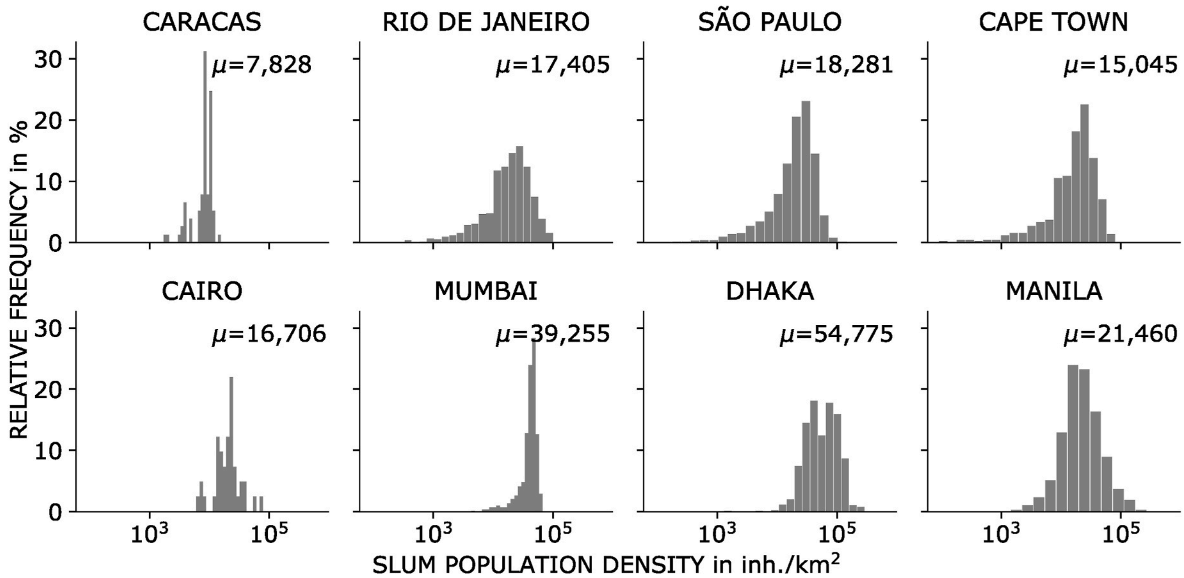


Fig. 14. relative Frequency of Slum Population Density. μ refers to the geometric mean of the population density estimates.

D Literature Sources for Total Slum Population Counts

Sources of the literature values used in Fig. 7 are shown in Table 3.

For the total number of slum population of Dhaka, the 2014 census was used on the one hand and on the other hand an estimation according to (Patel et al., 2019) was made. In (Patel et al., 2019), a systematic underestimation of the slum population in the 2014 census is calculated and based on that, the slum population of Bangladesh is estimated to be more in line with international standards (e.g. SDGs). These estimates were used here and projected to Dhaka by taking the proportion of Dhaka's slum population to Bangladesh's slum population as a ratio and multiplying it by Bangladesh's slum populations calculated in (Patel et al., 2019).

Table 3
results of total slum analysis. The first columns specifies the city. The third column contains population counts obtained by the method outlined in Section 2.4.1. The fourth column contains population counts found in literature. The fifth column contains the population fraction, a measure for the deviation, by dividing column three by column four. The seventh column gives the corresponding class of information, according to Section 2.4.3. If individual cell entries are empty, the value above them applies.

city	year	pop our approach in million	pop literature in million	fraction in %	class	source
Caracas	2015	0.44	1.25	36	(i)	Falco et al. (2019)
Rio de Janeiro	2010	0.89	1.70	52	(iv)	Instituto Brasileiro de Geografia e Estatistica (2010)
	2020	0.96	2	48	(i)	D+C and Covid (2021)
São Paulo	2010	1.41	2.16	65	(iv)	Instituto Brasileiro de Geografia e Estatistica (2010)
Cape Town	2016	1.38	1.23	112	(i)	(Milne)
Cairo	2000	4.13	2.8	147	(iv)	Sabry (2009)
			7.1	58	(iv)	Sabry (2009)
			8.3	50	(ii)	Sabry (2009)
	2004	4.40	7	63	(iv)	Sabry (2009)
	2006	4.55	10.7	43	(i)	Sabry (2009)
Mumbai	2006	1.47	6.1	24	(i)	Asha (2006)
	2007	1.50	11	14	(i)	Davis (2007)
	2010	1.56	8	19	(ii)	Taubenbö et al. (2015)
	2011	1.65	5.2	32	(iv)	Office of the Registrar General (2011)
			16.5	10	(ii)	Taubenbö et al. (2015)
			19.6	8	(ii)	Taubenbö et al. (2015)
Dhaka	2014	0.74	0.64	115	(iv)	Bangladesh Bureau of Statistics (2014)
			1.32	56	(ii)	Patel et al. (2019)
			3.84	19	(ii)	Patel et al. (2019)
			4.81	15	(ii)	Patel et al. (2019)
Manila	2014	1.02	4	25	(i)	(Roy)

E Slum Maps

In Fig. 15 the slum maps for the remaining five cities are shown.

(a) Caracas (b) São Paulo (c) Cairo

(d) Mumbai (e) Manila

Fig. 15. maps showing the location of slums in five different cities whose population figures have been checked for plausibility. Water and green landscapes are colored only for contrast.

F. Scaled Total Population Estimates

Table 4
results of scaled total population estimates. The third column contains the total slum population counts for the year 2015 (as shown in Fig. 4). The fourth column contains population counts obtained by scaling the values from the third column by the fractions according to Table 2. These scaled population counts are the values corresponding to the boxplots of Fig. 10.

city	pop our estimate in million	pop scaled estimate in million	pop scaled estimate's median in million
Caracas	0.45	1.04, 2.48, 6.28, 8.00, 10.46	6.28
Rio de Janeiro	0.92	1.17, 1.46, 1.54, 1.58, 1.91, 2.39, 3.98	1.58
São Paulo	1.48	3.02, 5.67, 8.62	5.67
Cape Town	1.34	1.08, 1.13, 1.27, 1.29, 1.54, 1.55, 1.89, 2.05, 2.15, 2.20, 2.80, 5.66, 7.85, 9.52, 56.36, 76.68	2.10
Cairo	0.30	0.01, 0.71, 3.75, 7.47, 18.66	3.75
Mumbai	1.65	1.68, 8.27, 10.88, 14.03, 26.31, 29.23	12.45
Dhaka	0.76	1.87, 2.69, 2.73, 3.43, 3.83, 4.84, 6.61, 8.22, 8.22, 9.19, 9.50, 9.92, 12.88, 19.44	7.41
Manila	1.03	0.89, 1.74, 2.38, 3.54, 3.69, 3.96, 4.20, 4.87	3.61

References

Abbott, J. (2001). Use of spatial data to support the integration of informal settlements into the formal city. *International Journal of Applied Earth Observation and Geoinformation, 3*(3), 267–277. https://doi.org/10.1016/S0303-2434(01)85033-9

Globel Administrative Areas. (2021). Global administrative areas: [digital geospatial data] (29.09. URL http://www.gadm.org/country.

Agarwal, S. (2011). The state of urban health in India; comparing the poorest quartile to the rest of the urban population in selected states and cities. *Environment and Urbanization, 23*(1), 13–28. https://doi.org/10.1177/0956247811398589

Ahuja, R., & Brosius, C. (2006). *Mumbai, Delhi, Kolkata: Annäherungen an die Megastädte Indiens*. Heidelberg: Draupadi-Verl.. URL http://deposit.dnb.de/cgi-bin/dokserv?id=2776503&prov=M&dok_var=1&dok_ext=htm.

Ali, E., Zachariah, R., Shams, Z., Vernaeve, L., Alders, P., Salio, F., Manzi, M., Allaouna, M., Draguez, B., Delchevalerie, P., & Harries, A. D. (2013). Is mid-upper arm circumference alone sufficient for deciding admission to a nutritional programme for childhood severe acute malnutrition in Bangladesh? *Transactions of the Royal Society of Tropical Medicine and Hygiene, 107*(5), 319–323. https://doi.org/10.1093/trstmh/trt018

Angeles, G., Lance, P., Barden-O'Fallon, J., Islam, N., Mahbub, A. Q. M., & Nazem, N. I. (2009). The 2005 census and mapping of slums in Bangladesh: Design, select results and application. *International Journal of Health Geographics, 8*, 32. https://doi.org/10.1186/1476-072X-8-32

Abascal, A., Rothwell, N., Shonowo, A., Thomson, D. R., Elias, P., Elsey, H., Yeboah, G., & Kuffer, M. (2022). "domains of deprivation framework" for mapping slums, informal settlements, and other deprived areas in lmics to improve urban planning and policy:

A scoping review. *Environment and Urban Systems*. https://doi.org/10.20944/preprints202102.0242.v1

M. F. Archila Bustos, O. Hall, T. Niedomysl, U. Ernstson, A pixel level evaluation of five multitemporal global gridded population datasets: A case study in Sweden, 1990–2015, Population and Environment 42 (2). doi:10.1007/s11111-020-00360-8.

Asha, K. (2006). *Urban slums in India - the myths and the reality: Asha-seattle's quarterly newsletter*.

Badkar, M.. A shocking satellite tour of the world's biggest slums. Business Insider. URL https://www.businessinsider.com/worlds-biggest-slums-2011-2?op=1#petare-venezuela-a-slum-in-caracas-with-approximately-600000-1-million-people-8.

Bangladesh Bureau of Statistics. (2014). *Census of slum areas and floating population 2014*.

Beguy, D., Bocquier, P., & Zulu, E. M. (2010). Circular migration patterns and determinants in nairobi slum settlements. *Demographic Research, 23*, 549–586. https://doi.org/10.4054/demres.2010.23.20

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breuer, J. H., & Friesen, J. (2023). Methods to assess spatio-temporal changes of slum populations. *Cities, 143*, Article 104582. https://doi.org/10.1016/j.cities.2023.104582

Cavalcanti-Ferreira, P., Monge-Naranjo, A., & Pereira, T.de M. (2016). Working Papers. *Luciene, of cities and slums* (Vol. 22). St. Louis: Federal Reserve Bank of. https://doi.org/10.20955/wp.2016.022.

B. Chaffey, Life inside one of the world's largest slums, The Borgen Project. URL https://borgenproject.org/life-inside-worlds-largest-slums/.

D+C, Covid-19. (2021). *strikes hard in rio de janeiro's many informal settlements*. URL https://www.dandc.eu/en/article/covid-19-strikes-hard-rio-de-janeiros-many-informal-settlements.

Davis, M. (2007). *Planet of slums*. London and New York: Verso.

Dicionario de Favelas Marielle Franco. (2020a). Heliópolis. URL https://wikifavelas.com.br/index.php/Heli%C3%B3polis.

Dicionario de Favelas Marielle Franco. (2020b). URL *Rocinha* https://wikifavelas.com.br/index.php?title=Rocinha.

Dicionario de Favelas Marielle Franco. (2020c). Favela de manguinhos. URL https://wikifavelas.com.br/index.php/Favela_de_Manguinhos.

Falco, E., Zambrano-Verratti, J., & Kleinhans, R. (2019). Web-based participatory mapping in informal settlements: The slums of caracas, Venezuela. *Habitat International, 94*, Article 102038. https://doi.org/10.1016/j.habitatint.2019.102038

Fattah, K. N., & Walters, P. (2020). "a good place for the poor!" counternarratives to territorial stigmatisation from two informal settlements in dhaka. *Social Inclusion, 8* (1), 55–65. https://doi.org/10.17645/si.v8i1.2318

Friesen, J., Knoche, C., Hartig, J., Pelz, P. F., Taubenböck, H., & Wurm, M. (2019). *Sensitivity of slum size distributions as a function of spatial parameters for slum classification* (pp. 1–4). IEEE. https://doi.org/10.1109/JURSE.2019.8808944, 52019.

Friesen, J., Kraff, N. J., & Taubenböck, H. (2023). Spatiotemporal dynamics of slum populations in caracas, Venezuela. In *2023 joint urban remote sensing event (JURSE)* (pp. 1–4). https://doi.org/10.1109/JURSE57346.2023.10144182

Friesen, J., Rausch, L., Pelz, P., & Fürnkranz, J. (2018). Determining factors for slum growth with predictive data mining methods. *Urban Science, 2*(3), 81. https://doi.org/10.3390/urbansci2030081. URL https://www.mdpi.com/2413-8851/2/3/81.

Friesen, J., Taubenböck, H., Wurm, M., & Pelz, P. F. (2018). The similar size of slums. *Habitat International, 73*, 79–88. https://doi.org/10.1016/j.habitatint.2018.02.002

Friesen, J., Taubenböck, H., Wurm, M., & Pelz, P. F. (2019). Size distributions of slums across the globe using different data and classification methods. *European Journal of Remote Sensing, 52*(sup2), 99–111. https://doi.org/10.1080/22797254.2019.1579617

Gilbert, A. (2007). The return of the slum: Does language matter? *International Journal of Urban and Regional Research, 31*(4), 697–713.

Gruebner, O., Sachs, J., Nockert, A., Frings, M., Khan, M. M. H., Lakes, T., & Hostert, P. (2014). Mapping the slums of dhaka from 2006 to 2010. *Dataset Papers in Science 2014*, 1–7. https://doi.org/10.1155/2014/172182

Habitat for Humanity GB, the world's largest slums: Dharavi, kibera, khayelitsha & neza. URL https://www.habitatforhumanity.org.uk/blog/2017/12/the-worlds-largest-slums-dharavi-kibera-khayelitsha-neza/, (2017).

Huchzermeyer, M. (2010). *Informal settlements: A perpetual challenge?* (reprint Edition). Lansdowne: UCT Press.

Hulett, A. C., Yibirin, M. G., Brandt, R. B., García, A., Hurtado, D., & Puigbó, A. P. (2013). Home/social environment and asthma profiles in a vulnerable community from caracas: Lessons for urban Venezuela? *Journal of Asthma: Official Journal of the Association for the Care of Asthma, 50*(1), 14–24. https://doi.org/10.3109/02770903.2012.747205

Instituto Brasileiro de Geografia e Estatistica. (2010). *Censo demografico 2010: Resultados do universo, agregados por setores censitarios*.

Islam, N., Mahbub, A., Nazem, N., Angeles, G., & Lance, P. (2006). Slums of urban Bangladesh: Mapping and census, 2005. URL https://www.measureevaluation.org/resources/publications/tr-06-35.html.

Jordan, B. (2008). Gateway housing project in a shambles. URL https://web.archive.org/web/20090215003448/.

de Mattos, H., Agatha, C., & McArdle, G. (2020). *Assessing the quality of gridded population data for quantifying the population living in deprived communities* (preprint). doi: https://doi.org/10.48550/arXiv.2011.12923.

S. K. Karn, S. Shikura, H. Harada, Living environment and health of urban poor: A study in Mumbai, Economic and Political Weekly.

Khalequzzaman, M., Chiang, C., Hoque, B. A., Choudhury, S. R., Nizam, S., Yatsuya, H., Matsuyama, A., Hirakawa, Y., Islam, S. S., Iso, H., & Aoyama, A. (2017). Population profile and residential environment of an urban poor community in dhaka, Bangladesh. *Environmental Health and Preventive Medicine, 22*(1), 1. https://doi.org/10.1186/s12199-017-0610-2

Kit, O., & Lüdeke, M. (2013). Automated detection of slum area change in hyderabad, India using multitemporal satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 83*, 130–137. https://doi.org/10.1016/j.isprsjprs.2013.06.009

Kohli, D., Sliuzas, R., Kerle, N., & Stein, A. (2012). An ontology of slums for image-based classification. *Computers, Environment and Urban Systems, 36*(2), 154–163. https://doi.org/10.1016/j.compenvurbsys.2011.11.001

Kuffer, M., Pfeffer, K., & Sliuzas, R. (2016). Slums from space—15 years of slum mapping using remote sensing. *Remote Sensing, 8*(6), 455. https://doi.org/10.3390/rs8060455

Kuffer, M., Wang, J., Nagenborg, M., Pfeffer, K., Kohli, D., Sliuzas, R., & Persello, C. (2018). The scope of earth-observation to improve the consistency of the sdg slum indicator. *ISPRS International Journal of Geo-Information, 7*(11), 428. https://doi.org/10.3390/ijgi7110428

Latif, M. B., Irin, A., & Ferdaus, J. (2016). Socio-economic and health status of slum dwellers of the kalyanpur slum in dhaka city, Bangladesh. *Journal of Scientific Research, 29*(1), 73–83. https://doi.org/10.3329/bjsr.v29i1.29760

Leyk, S., Gaughan, A. E., Adamo, S. B., de Sherbinin, A., Balk, D., Freire, S., Rose, A., Stevens, F. R., Blankespoor, B., Frye, C., Comenetz, J., Sorichetta, A., MacManus, K., Pistolesi, L., Levy, M., Tatem, A. J., & Pesaresi, M. (2019). The spatial allocation of population: A review of large-scale gridded population data products and their fitness for use. *Earth System Science Data, 11*(3), 1385–1409. https://doi.org/10.5194/essd-11-1385-2019

Lloyd, C. T., Chamberlain, H., Kerr, D., Yetman, G., Pistolesi, L., Stevens, F. R., Gaughan, A. E., Nieves, J. J., Hornby, G., MacManus, K., Sinha, P., Bondarenko, M., Sorichetta, A., & Tatem, A. J. (2019). Global spatio-temporally harmonised datasets for producing high-resolution gridded population distribution datasets. *Big Earth Data, 3*(2), 108–139. https://doi.org/10.1080/20964471.2019.1625151

Lucci, P., Bhatkal, T., & Khan, A. (2018). Are we underestimating urban poverty? *World Development, 103*, 297–310. https://doi.org/10.1016/j.worlddev.2017.10.022. URL https://www.sciencedirect.com/science/article/pii/S0305750X17303467.

Mahabir, R., Crooks, A., Croitoru, A., & Agouris, P. (2016). The study of slums as social and physical constructs: Challenges and emerging research opportunities, regional studies. *Regional Science, 3*(1), 399–419. https://doi.org/10.1080/21681376.2016.1229130

The masiphumelele corporation and trust, masiphumelele demographics. URL https://web.archive.org/web/20110918172234/, (2011).

N. Milne, The tale of two slums in south africa as residents seek to upgrade lives, Reuters Media. URL https://www.reuters.com/article/us-safrica-slums-upgrading-idUSKBN1431PO.

Mridha, M. K., Hossain, A., Alam, B., Sarker, B. K., Wahed, T., Khan, R., & Roy, S. (2009). *Manoshi - working paper: The perceptions of community groups to improve mnch in urban slums: An exploratory case study of korailslum in dhaka*.

NASA. TerraMetrics, map data, Google maps: Satellite image of the world (n.d.). URL http://mt0.google.com/vt/lyrs=s&hl=en&x=x&y=y&z=z.

Office of the registrar general, census commissioner, 2011 census data. URL https://censusindia.gov.in/2011-Common/CensusData2011.Html, (2011).

OpenStreetMap contributors. (2017). *Planet dump*. retrieved from. URL https://www.openstreetmap.org.

Parliamentary Monitoring Group. Khayelitsha hospital challenges and successes (25.09.2018). URL https://pmg.org.za/committee-meeting/27112/.

Patel, A., Joseph, G., Shrestha, A., & Foint, Y. (2019). Measuring deprivations in the slums of Bangladesh: Implications for achieving sustainable development goals. *Housing and Society, 46*(2), 81–109. https://doi.org/10.1080/08882746.2019.1622346

Pedro, A. A., & Queiroz, A. P. (2019). Slum: Comparing municipal and census basemaps. *Habitat International, 83*, 30–40. https://doi.org/10.1016/j.habitatint.2018.11.001

Pelz, P. F., Friesen, J., & Hartig, J. (2019). Similar size of slums caused by a turing instability of migration behavior. *Physical review, E 99*(2–1), Article 022302. https://doi.org/10.1103/PhysRevE.99.022302

Philippine Statistics Authority. (2015). Census of population 2015. URL http://psada.psa.gov.ph/index.php/catalog/173.

B. Pramnik, D. C. Sarker, R. C. Sarker, Assessment of water supply and sanitation facilities for korail slum in dhaka city, International Journal of Civil & Environmental Engineering IJCEE-IJENS 11 (11).

Risbud, N. (2003). The case of Mumbai, India: Understanding slums: Case studies for the global report on human settlements. URL. In *UN-HABITAT* (p. 6). UN-Habitat https://www.ucl.ac.uk/dpu-projects/Global_Report/cities/mumbai.htm.

P. Roy, In the slums of manila, inequality is so bad that the worst off have no chance to protest, New Statesman. URL https://www.newstatesman.com/world/2014/09/slums-manila-inequality-so-bad-worst-have-no-chance-protest.

Sabry, S. (2009). *Poverty lines in greater Cairo: underestimating and misrepresenting poverty, IIED*.

Savarese, M., Jeantet, D., & Pollastri, T.. Favela centennial shows Brazil communities' endurance. Associated Press. URL https://apnews.com/article/brazil-sao-paulo-9b49da65d993e32b7b85f0393c1bc9d4.

Sims, D. (2003). *Understanding urban slums*. URL. Cairo, Egypt: GTZ https://www.ucl.ac.uk/dpu-projects/global_report/pdfs/cairo.pdf.

Slum/Shack Dwellers International. (2017). *Know your city*. URL https://sdinet.org/explore-our-data/.

Staff, I.. An intro to rio's favelas, the Santa Barbara Independent. URL https://www.independent.com/2010/06/04/intro-rios-favelas/.

J. E. Steele, P. R. Sundsøy, C. Pezzulo, V. A. Alegana, T. J. Bird, J. Blumenstock, J. Bjelland, K. Engø-Monsen, Y.-A. de Montjoye, A. M. Iqbal, K. N. Hadiuzzaman, X. Lu, E. Wetter, A. J. Tatem, L. Bengtsson, Mapping poverty using mobile phone and

satellite data, Journal of The Royal Society Interface doi:http://doi.org/10.1098/rsif.2016.0690.
Su, M.-C., & Mlčák, D. (2014). *Survey of drinking water access and utilization in baseco slum, manila, Philippines –influence factors, drinking water type preferences, and spatial distribution*. https://doi.org/10.13140/RG.2.2.28863.94880. Ph.D. thesis, Unpublished.
Taubenböck, H., Kraff, N. J., & Wurm, M. (2018). The morphology of the arrival city - a global categorization based on literature surveys and remotely sensed data. *Applied Geography, 92*, 150–167. https://doi.org/10.1016/j.apgeog.2018.02.002
Taubenböck, H., Weigand, M., Esch, T., Staab, J., Wurm, M., Mast, J., & Dech, S. (2019). A new ranking of the world's largest cities—do administrative units obscure morphological realities? *Remote Sensing of Environment, 232*, Article 111353. https://doi.org/10.1016/j.rse.2019.111353
Taubenböck, H., Wurm, M., Esch, T., & Dech, S. (2015). *Globale urbanisierung*. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-44841-0
Thomson, D. R., Gaughan, A. E., Stevens, F. R., Yetman, G., Elias, P., & Chen, R. (2021). Evaluating the accuracy of gridded population estimates in slums: A case study in Nigeria and Kenya. *Urban Science, 5*(2), 48. https://doi.org/10.3390/urbansci57020048
Thomson, D. R., Kuffer, M., Boo, G., Hati, B., Grippa, T., Elsey, H., Linard, C., Mahabir, R., Kyobutungi, C., Maviti, J., Mwaniki, D., Ndugwa, R., Makau, J., Sliuzas, R., Cheruiyot, S., Nyambuga, K., Mboga, N., Kimani, N. W., de Albuquerque, J. P., & Kabaria, C. (2020). Need for an integrated deprived area "slum" mapping system (ideamaps) in low- and middle-income countries (lmics). *Social Sciences, 9*(5), 80. https://doi.org/10.3390/socsci9050080
Thomson, D. R., Leasure, D. R., Bird, T., Tzavidis, N., & Tatem, A. J. (2021). How accurate are WorldPop-global gridded population data at the cell-level?: A simulation analysis in urban Namibia. *Preprints*. https://doi.org/10.20944/preprints202102.0492.v2
Thomson, D. R., Rhoda, D. A., Tatem, A. J., & Castro, M. C. (2020). Gridded population survey sampling: A systematic scoping review of the field and strategic research agenda. *International Journal of Health Geographics, 19*(1), 34. https://doi.org/10.1186/s12942-020-00230-4. URL https://ij-healthgeographics.biomedcentral.com/articles/10.1186/s12942-020-00230-4.
Thomson, D. R., Stevens, F. R., Chen, R., Yetman, G., Sorichetta, A., & Gaughan, A. E.. Improving the accuracy of gridded population estimates in cities and slums to monitor sdg 11: Evidence from a simulation study in Namibia. https://www.preprints.org/manuscript/202107.0510.
UN Department of Economic and Social Affairs. (2018). World urbanization prospects: The 2018 revision. URL https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf.
UN General Assembly. (2015). Transforming our world : The 2030 agenda for sustainable development. URL https://sdgs.un.org/goals.
UN-Habitat. (2020). *The value of sustainable urbanization, Vol. 2020 of World cities report.* Nairobi, Kenya: UN-Habitat.
UN-habitat, slums: Some definitions. URL https://mirror.unhabitat.org/documents/media_centre/sowcr2006/SOWCR%205.pdf, (2007).
United Nations. (2018). World urbanization prospects. URL https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf.
United Nations, Department of Economic and Social Affairs. (2018). *Population Division, World urbanization prospects: The 2018 revision*. online edition https://population.un.org/wup/ [Online; accessed <today's date>].
Verzoni, A. (2018). Shantytowns, nov dec 2018. URL https://www.nfpa.org/News-and-Research/Publications-and-media/NFPA-Journal/2018/November-December-2018/Features/Shantytowns.
Wade, T.. Favelas no more? Some brazilian slums mature. Reuters Media. URL https://www.reuters.com/article/us-cities-brazil/favelas-no-more-some-brazilian-slums-mature-idUSN2844848720061211.
Wardrop, N. A., Jochem, W. C., Bird, T. J., Chamberlain, H. R., Clarke, D., Kerr, D., Bengtsson, L., Juran, S., Seaman, V., & Tatem, A. J. (2018). Spatially disaggregated population estimates in the absence of national population and housing census data. *Proceedings of the National Academy of Sciences, 115*(14), 3529–3537. https://doi.org/10.1073/pnas.1715305115
Worldpop. (2018). *Global 100m population*. https://doi.org/10.5258/SOTON/WP00645
Wurm, M., Stark, T., Zhu, X. X., Weigand, M., & Taubenböck, H. (2019). Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing, 150*, 59–69. https://doi.org/10.1016/j.isprsjprs.2019.02.006
Wurm, M., & Taubenböck, H. (2018). Detecting social groups from space – assessment of remote sensing-based mapped morphological slums using income data. *Remote Sensing Letters, 9*(1), 41–50.
Wurm, M., Taubenböck, H., Weigand, M., & Schmitt, A. (2017). Slum mapping in polarimetric sar data using spatial features. *Remote Sensing of Environment, 194*, 190–204. https://doi.org/10.1016/j.rse.2017.03.030