

# Chapter 14

## Making Urban Slum Population Visible: Citizens and Satellites to Reinforce Slum Censuses



Angela Abascal, Stefanos Georganos , Monika Kuffer ,  
Sabine Vanhuysse, Dana Thomson, Jon Wang, Lawrence Manyasi,  
Daniel Manyasi Otunga, Brighton Ochieng, Treva Ochieng, Jorge Klinnert,  
and Eléonore Wolff

**Abstract** In response to the “Leave No One Behind” principle (the central promise of the 2030 Agenda for Sustainable Development), reliable estimate of the total number of citizens living in slums is urgently needed but not available for some of the most vulnerable communities. Not having a reliable estimate of the number of poor urban dwellers limits evidence-based decision-making for proper resource

---

A. Abascal (✉)

Navarra Centre for International Development, University of Navarra, Pamplona, Spain

Faculty of Geo-information Science and Earth Observation (ITC), University of Twente,  
Enschede, The Netherlands

e-mail: [maabascal@unav.es](mailto:maabascal@unav.es)

S. Georganos

Geomatics, Department of Environmental and Life Sciences, Karlstad University,  
Karlstad, Sweden

e-mail: [stefanos.georganos@kau.se](mailto:stefanos.georganos@kau.se)

M. Kuffer · D. Thomson · J. Wang

Faculty of Geo-information Science and Earth Observation (ITC), University of Twente,  
Enschede, The Netherlands

e-mail: [m.kuffer@utwente.nl](mailto:m.kuffer@utwente.nl); [d.r.thomson@utwente.nl](mailto:d.r.thomson@utwente.nl); [j.wang@utwente.nl](mailto:j.wang@utwente.nl)

S. Vanhuysse · E. Wolff

Department of Geosciences, Environment and Society, Université libre de Bruxelles (ULB),  
Brussels, Belgium

e-mail: [sabine.vanhuysse@ulb.be](mailto:sabine.vanhuysse@ulb.be); [eleonore.wolff@ulb.be](mailto:eleonore.wolff@ulb.be)

L. Manyasi · B. Ochieng · T. Ochieng

Community Researcher Non-affiliated, Nairobi, Kenya

D. M. Otunga

Community Researcher Single Mother Association of Kenya, Nairobi, Kenya

J. Klinnert

Department of Economics, University of Maryland, College Park, MD, USA

e-mail: [klinnert@umd.edu](mailto:klinnert@umd.edu)

© The Author(s) 2024

M. Kuffer, S. Georganos (eds.), *Urban Inequalities from Space*, Remote Sensing and  
Digital Image Processing 26, [https://doi.org/10.1007/978-3-031-49183-2\\_14](https://doi.org/10.1007/978-3-031-49183-2_14)

allocation in the fight against urban inequalities. From a geographical perspective, urban population distribution maps in many low- and middle-income cities are most often derived from outdated or unreliable census data disaggregated by coarse administrative units. Moreover, slum populations are presented as aggregated within bigger administrative areas, leading to a large diffuse in the estimates. Existing *global and open population databases* provide homogeneously disaggregated information (i.e. in a spatial grid), but they mostly rely on census data to generate their estimates, so they do not provide additional information on the slum population. While a few studies have focused on bottom-up geospatial models for slum population mapping using survey data, geospatial covariates, and earth observation imagery, there is still a significant gap in methodological approaches for producing precise estimates within slums. To address this issue, we designed a pilot experiment to explore new avenues. We conducted this study in the slums of Nairobi, where we collected in situ data together with slum dwellers using a novel data collection protocol. Our results show that the combination of satellite imagery with in situ data collected by citizen science paves the way for generalisable, gridded estimates of slum populations. Furthermore, we find that the urban physiognomy of slums and population distribution patterns are related, which allows for highlighting the diversity of such patterns using earth observation within and between slums of the same city.

**Keywords** Urban population · Slums · Citizen science · Data collection · Earth observation · Machine learning · Census

## 14.1 Introduction

Global sustainable development, adoption of the United Nations (UN) Sustainable Development Goals (SDGs) (UN-Habitat 2015), and international development agreements have triggered a much-needed data revolution, with countries and institutions around the world recognising the critical role of geospatial data for evidence-based policymaking. Increasingly, high-quality geospatial datasets are becoming an essential source of information to guide social, economic, and environmental policies at global, regional, national, and subnational scales (Nilsson et al. 2016).

Among the wide variety of geospatial datasets needed to create credible measures of sustainable development (e.g. data on land use, land cover, risks, or climate indicators), some of the most spatially necessary datasets are those describing the spatial distribution of the urban form (Benza et al. 2016) and the human population. Global population datasets must ensure to include all populations, with particular attention to urban slums, which exhibit distinct patterns from the rest of the city and are often not accounted for in census data in most low- and middle-income countries (LMICs). While open population databases are widely used, including censuses, surveys, and gridded datasets (e.g. WorldPop), they systematically and dramatically underrepresent slum populations in LMICs (Thomson et al. 2020). Global

population models come with high uncertainties in LMIC cities, e.g. the underestimation of the urban population living in slums (Kuffer et al. 2022).

Slums are characterised by sub-standard services, lack of open space, and high building density. Deprivation is exacerbated by high urban built-up density and overcrowding of households that are much higher than in urban formal areas. In addition, slum dwellers commonly lack land tenure security, so they are often not recognised by local authorities and are not included in official population counts (i.e. in national censuses). This means that population data collected through traditional methods may not fully reflect the number of people living in slums. This problem is compounded by the fact that slum populations are often highly mobile (people move in and out of neighbourhoods frequently) and with diverse population patterns according to social strata, which can result in multiple families residing within the same household. The presence of such diversity in slum populations makes it challenging to estimate their numbers accurately based on limited observations or extrapolation methods.

It should be noted, as shown in Fig. 14.1, that not only is there a morphological distinction between slums and formal areas, but there is variation in size and physiognomy between and within slums (Georganos et al. 2021). Therefore, the extrapolation of slum data needs to be carefully considered. To address this problem, more targeted and innovative approaches to data collection in LMIC, especially in slums, are needed. This may involve the use of new technologies, such as satellite imagery and geospatial data, to identify slum populations. It may also require closer collaboration between local authorities, researchers, and community organisations to ensure that slum populations are adequately represented in population data.

Choosing the appropriate *spatial unit* for modelling urban slum population depends on several factors, including the scale and geographic extent of the study, the level of detail required for the analysis, and the available input data and the protection of privacy. To accurately model the population of slums within global population layers, the spatial unit chosen needs to match the intra-slum level patterns of the population density. This will enable a more accurate representation of the population distribution and help to address any potential biases in the mapping process. Therefore, administrative units may not be appropriate as they often contain very diverse areas and potentially include one or several slums. The choice of the spatial unit should also be driven by the availability and quality of data. For example, census tracts may be ideal for modelling urban populations at larger scales, in case of high-quality census data availability. However, in the case of LMIC cities, census data are not reliable data sources for population estimation in slums (Carr-Hill 2013) because they are often outdated regarding the high pace of their population growth, and their smallest administrative unit is too small. Therefore, considering the fine level of detail required in this study, fine grid units would be optimal (Fig. 14.2). It should be noted that the selection of appropriate spatial units involves adapting to the diversity within each grid to accurately represent the population distribution at a finer resolution.

The process of gridded population mapping requires understanding the integration approaches between diverse geo-datasets, from source units into the target units



**Fig. 14.1** Physiognomic diversity between and within slums in Nairobi city

(Leyk et al. 2019). Previous research in different fields has shown that *spatial aggregation* (in top-down approaches) influences urban modelling results (Zhang and Kukadia 2005; Duque et al. 2018; Weigand et al. 2019). This common challenge is called the modifiable areal unit problem (*MAUP*) (Gehlke and Biehi 1934; Openshaw 1984; Wong 2009) and describes how results change as the spatial aggregation of data changes. A good approach to evaluating the influence of the aggregation level is to conduct a sensitivity analysis, testing different spatial units to see which one provides the best results and insights for the research question at hand.

As shown in Fig. 14.3, several global open-gridded population datasets are available, and the modelling process involved in their creation is variable (for further information, visit the POPGRID website (2020)). The first distinction is whether they are *top-down or bottom-up approaches*. Top-down approaches to distributing spatial population patterns have been shown to grossly underestimate the slum population, as they are based on the disaggregation of census data into smaller units (Thomson et al. 2020). On the other hand, bottom-up approaches are typically



Fig. 14.2 Spatial units diagram: from administrative to more physiognomic

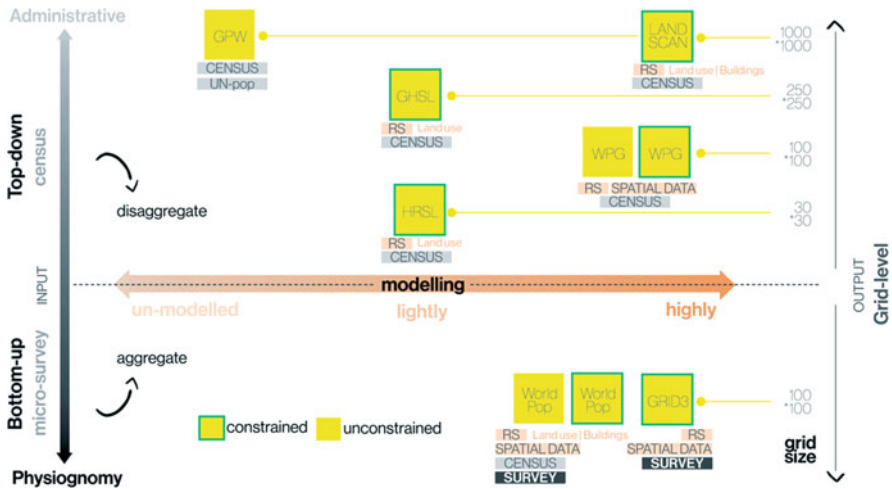


Fig. 14.3 Global population datasets diagram

more accurate, employing commonly in situ survey data with a set of covariates in statistical and machine learning approaches (Boo et al. 2020). A distinction is also made between approaches that do and do not constrain population redistribution using settlement extents. A constrained approach produces improved population distributions (Linard et al. 2011). Finally, the degree of modelling, from

**Table 14.1** Disaggregation methods

Method type	Method name	Weighting	Authors	
Disaggregation	Statistical analysis	Means of population	Birkin and Clarke (1988)	
	Areal weighting	Uniform distribution	Goodchild and Lam (1980)	
	Dasymetric mapping	Binary weight		Eicher and Brewer (2001)
		Empirical weight		Mennis and Hultgren (2006)
		Statistical weight		Reibel and Bufalino (2005)
Hybrid dasymetric	Statistical analysis and dasymetric mapping		Stevens et al. (2015), Sorichetta et al. (2015), Reed et al. (2018)	

un-modelled to highly modelled, indicates the size of the covariate set used to predict the population.

When creating spatial population data, it is necessary to *disaggregate* or *aggregate* data into a common spatial unit. *Disaggregation* is typically employed when the source data is census or administrative data, and the target grid cell is a smaller spatial unit than the source units. It has been shown that the incorporation of ancillary urban data, named constrained, results in better accuracies than unconstrained modelling when disaggregating from census data. As shown in Table 14.1, there are several disaggregation methods to allocate population distribution into smaller units. One common method is the areal weighting method, an unconstrained approach, which assumes that the population is uniformly redistributed from the source units to the target cells that overlap with the source units. Although this assumption is an oversimplification due to non-uniform population distributions, the method is computationally efficient and can create spatially explicit and globally consistent population estimates. An example of this approach is the gridded population of the world (GPW) product. When using ancillary data, the population redistribution is achieved through areal interpolation, referred to as the dasymetric mapping method. All dasymetric mapping techniques rely on the relationships between population (provided by the input census data) and ancillary data (such as land cover) that can be utilised to reallocate the population to finer spatial units with greater accuracy. The disaggregation method applied in traditional dasymetric approaches varies, ranging from binary dasymetric refinement to more complex weighting schemes or hybrid methods. As Leyk et al. (2019) stressed, they differ in the way relationships between population and ancillary data are derived (e.g. presence/absence based, empirically derived, or optimised) to determine weights for different locations to guide the disaggregation of population totals.

Efforts have been made to develop models that are not dependent on census data and rely on bottom-up approaches. However, these attempts have not been applied on a global scale. For example, the GRID3 project has developed population estimates for LMICs that are independent of census data (Grid3 2023). Recently, a method combining household survey data with building footprints has been

suggested as a possible census-independent approach (Thomson et al. 2020; Boo et al. 2022), the most promising method for creating benchmark models for slum populations. It is noted that attempts to develop bottom-up population estimations use a random sampling scheme of survey locations, and while this methodology suggests that sample surveys can provide unbiased estimates of the total population size of an aggregated population (Cottam et al. 1957), the precision of these estimates will depend on the level of aggregation, the distribution of population within the aggregate units, and the extent to which the sample is representative of the different urban physiognomies. As stated by Georganos et al. (2021), the latter differ vastly within and between slums.

In this study, we aim to propose a novel method to produce more accurate gridded population estimates that can be used as a benchmark to evaluate other open population datasets and validate and/or highlight existing uncertainties. Data was collected in situ in collaboration with slum dwellers through a novel data collection, as all the grid size was surveyed, using citizen science, i.e. slum dwellers. The authors hypothesise that leveraging bottom-up earth observation (EO) methods can improve slum population estimates compared to existing censuses and open-gridded population datasets.

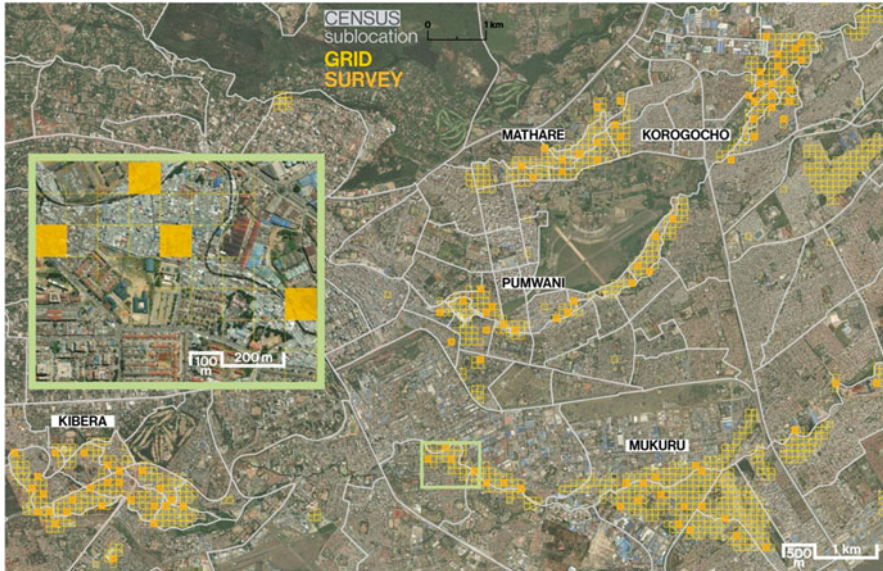
## 14.2 Data

### 14.2.1 *Onekana Population Database*

The study covers six slum areas in Nairobi city, namely Kibera, Mukuru, Waruku, Pumwani, Korogocho, and Mathare. The study area was divided into 2000 uniformly sized grid cells, each measuring 1 hectare (100 m × 100 m). Out of these, 117 cells were randomly selected and comprehensively surveyed, resulting in a total of 10,550 surveys conducted (Fig.14.4). Throughout the data collection process, certain grid cells had to be excluded from the survey due to safety concerns reported by citizens. These cells were predominantly situated close to rivers, areas with known drug or alcohol consumption, or fenced-off areas with restricted access.

### 14.2.2 *Open Geo-Datasets*

Interpretable urban variables in the form of open geo-data were collected from two sources, namely satellite imagery and cooperative citizen databases such as OpenStreetMap. Variables derived from satellite imagery were produced in prior research studies utilising very high-resolution WorldView-3 images (0.30 cm) of 2019 (Abascal et al. 2022; Georganos et al. 2021; Wang et al. 2023). They consist of proportions of land cover classes and morphological metrics aggregated at the grid



**Fig. 14.4** Nairobi case study. Census sub-locations, total grids corresponding to slums, and surveyed grid cells are depicted

cell level. Additional morphological metrics were computed from the Google Open Building layer, which uses satellite imagery of 0.50 cm resolution (Sirko et al. 2021). Finally, road and river data were collected from OpenStreetMap (OSM) contributed by community members in 2021.

### 14.2.3 *Open Population Datasets*

Table 14.2 presents the population datasets selected for comparison, which were chosen based on two criteria: spatial resolution finer than or equal to 100 m by 100 m and data availability for Nairobi city.

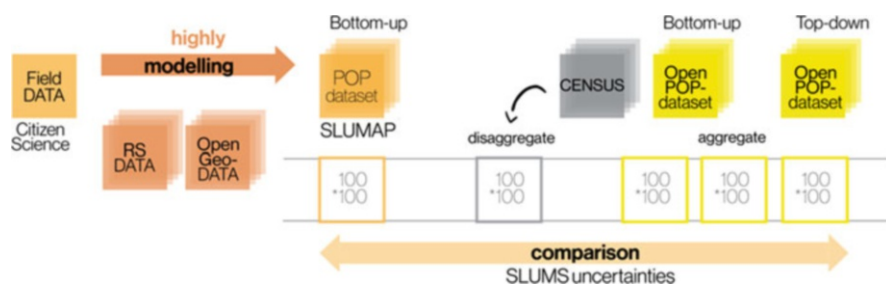
## 14.3 Methodology

Figure 14.5 illustrates the general workflow of the study. The first step involves designing and collecting field data, which are then subjected to cleaning and imputation to estimate the population of non-responding households in the grid. Subsequently, the data obtained from surveying 117 grids are used to train and validate the model using advanced spatial aggregation techniques. This results in the creation of the Onekana population dataset, which provides predicted population



**Table 14.2** Population geospatial datasets for Nairobi with spatial resolution finer than or equal to  $100\text{ m} \times 100\text{ m}$  used in this research

Name and source	Institution	Authors	Geographic coverage	Spatial resolution	Data	Date
Census	KNBS	KNBS (2019)	Kenya	Adm. Sub-location	Census	2019
High Resolution Settlement Layer (HRSL) – constrained	Columbia University, Facebook	CIESIN (2016) Meta (2023)	Africa, Asia, Latin America	Grid (30 m)	Census, RS	2020
WorldPop – constrained	University Southampton	WorldPop (2023)	Global	Grid (100 m)	Census, Survey, Geospatial data, and RS	2020
GRID 3 – constrained	Columbia University, WorldPop	Gadiaga et al. (2021)	Sub-Saharan Africa (11 countries)	Grid (100 m)	Building patterns (Dooley et al. 2020), Geospatial data (Lloyd et al. 2017)	2023

**Fig. 14.5** Overall workflow

data for the remaining slum grids. Finally, this dataset is compared with more recent and disaggregated open population datasets described in Table 14.1.

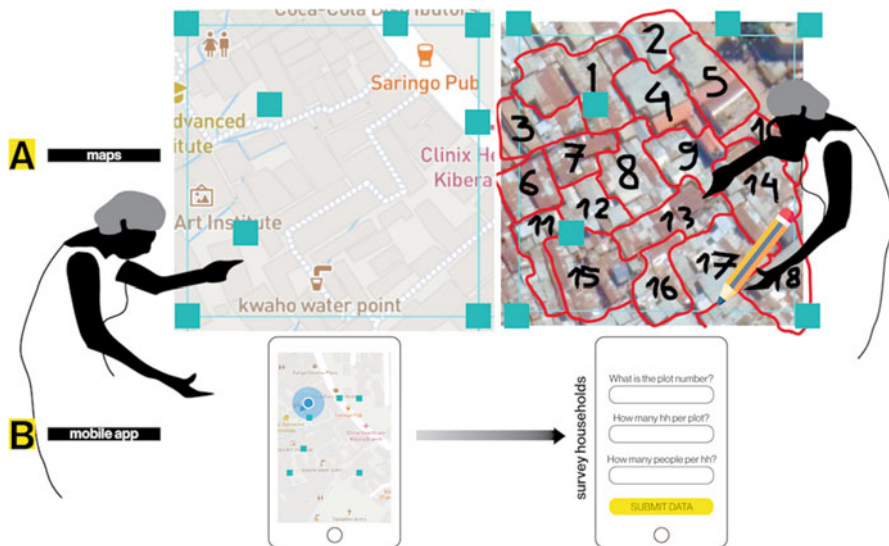
### 14.3.1 ONEKANA Population Database

**Field Data Collection** To model population, we assume households living in similar physical housing and environmental conditions have similar demographic characteristics, and therefore similar population numbers (Pearce et al. 2010). To gather population data on the ground, slum dwellers participated in a citizen science

process. The survey and its objectives were explained to them, and they were involved in designing the work plan and determining their remuneration. They also helped to convey the purpose of the surveys to the respondents to build trust and achieve a high response rate. However, reaching some households was not possible due to the unstable employment situation and unpredictable work schedules of some residents.

The data collection process involves two main steps: (A) map exploration and plot delineation and (B) a field survey conducted using a mobile app. Due to the complicated physical layout of the slums, it was not feasible to survey each dwelling individually as originally planned. Instead, maps were created from very-high-resolution (VHR) satellite image chips and OSM data to identify landmarks and serve as orientation guides within the slums. Surveyors then delineated plots, which are fenced compounds of grouped dwellings with a single entrance. The field survey also consists of two phases: First, counting the number of households in a plot and second, interviewing the inhabitants. Geo-data surveys were conducted using a mobile app created on the Firebase platform for this research (Fig. 14.6).

**Aggregation Methods** To evaluate the sensitivity of slum population patterns and assess the aggregation error, the missing data were filled in using two different aggregation methods. It should be noted that the total number of dwellings per plot and per grid cell was known, as they were counted in the field. The two methods used were (1) *grid aggregation method*: Population numbers were aggregated at the grid cell level, and the missing population number for a household was assumed to be



**Fig. 14.6** Field data collection process. (a) Locating the assigned grid cell and drawing plot boundaries. (b) Carrying out household surveys and filling out the form on the mobile app (grid id, plot id, household number, and population number)

equal to the mean household population number in the grid cell (2) *Plot to grid aggregation method*: Population numbers were aggregated at the plot level, and the missing population number for a household was assumed to be equal to the mean household population number in the plot. A first R-squared ( $R^2$ ) test measurement was performed to evaluate the proportion of variance and strength of the relationship between variables.

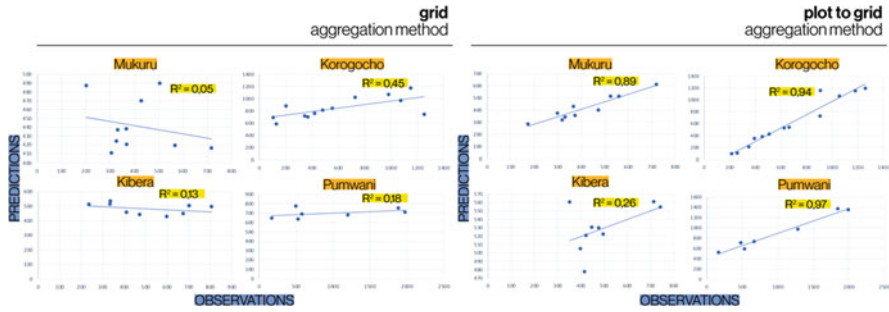
**Population Modelling** Slum population density is predicted for the six biggest slums within the city of Nairobi. As a modelling approach, we employed the random forest (RF) algorithm, where population numbers were predicted from geospatial covariates, such as land cover and building footprints and morphological metrics, for each slum separately, at the grid level. Additionally, to create a more predictive and generalisable slum population model, we used a spatial variant of RF, geographical random forest (Georganos et al. 2019) in which data from all slums were used simultaneously. We evaluate the results both by visual inspection and by validating our predictions against a set of surveys left out of the training stage by computing performance indicators such as root mean square error (RMSE) and mean absolute error (MAE).

**Comparison of Slum Population Estimates** To assess the cell-level accuracy of the four existing gridded population datasets considered (Table 14.2), each cell-level estimate (e.g. WorldPop) was compared to the “ground true” population count obtained from the exhaustive household surveys. The reference grid used in the comparison was the Onekana grid, 100 by 100 m. The *HRSL dataset* is aggregated using the *areal weighting method*, as we know the population distribution within each grid cell could be diverse (see Fig. 14.1). The areal-weighted interpolation is a more sophisticated method compared to simple spatial aggregation as it considers the spatial distribution of the population within each grid cell. The Census has been disaggregated into the reference grid by the *dasymetric mapping method*. This method considers the spatial distribution of the census data and uses ancillary data, such as land use or land cover, to redistribute the census data within each grid cell based on the spatial characteristics of the ancillary data. The other 100 by 100 m open population dataset (i.e. WorldPop and GRID3), as there was a spatial mismatch between the cells being compared, were converted to the reference grid using the *inverse distance weighting spatial interpolation method*. This method estimates the values of unsampled locations based on the values of nearby sampled locations. This method ensured all compared cells were on the same grid and minimised errors due to spatial mismatches between data sets.

## 14.4 Results

### 14.4.1 ONEKANA Population Database

This study employed two distinct aggregation methods to evaluate the impact on modelling outcomes: the grid aggregation method and the plot-to-grid aggregation



**Fig. 14.7** Population modelling per slum (with a distinct RF model for each slum), with two different aggregation methods

**Table 14.3** Performance indicators for the ONEKANA population estimate

Database	RMSE	R-squared	MAE
ONEKANA	307,28	0,48	199,13

method. Figure 14.7 shows the predicted versus observed population for four slums, when modelling the population for each slum separately, revealing that the plot-to-grid aggregation method yielded a remarkable outcome, with an  $R^2$  value approaching 0.90 for the major slums except for Kibera. However, the results were less satisfactory ( $R^2$  ranging from 0.20 to 0.25) when all the slums were modelled together. The finding that the model demonstrated high accuracy when modelling each slum individually, but not when modelling all slums together, provides support for the hypothesis that population patterns vary across slums.

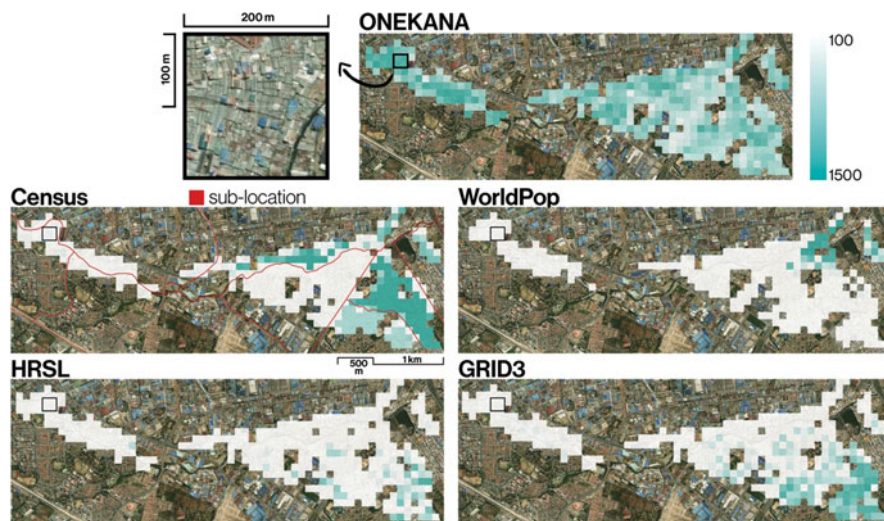
After the aggregation method, the final modelling was determined through a rigorous testing process, and the model was subsequently refined by incorporating a geographical component. The utilisation of the geographical random forest (GRF) method produced the best model and yielded enhanced results, as evidenced in Table 14.3.

### 14.4.2 Slum Population Comparison

Compared to existing population estimates, our results show a mean estimate per grid of almost 600 people, whereas others are less than half of that. As shown in Table 14.4, among the existing open population layers, the best estimate is GRID3 and the worst are both Census and the WorldPop database. This could be explained as when adding survey data (e.g. GRID3) or disaggregating the spatial area and model with exhaustive Remote Sensing data (e.g. HRSL) results improve quantitatively. Despite the improvement, it is clear, e.g. in Fig. 14.8, how the existing databases respond to a spatial distribution of the population such as the census data, as it has been the main data to train the models.

**Table 14.4** Population estimate comparison in Nairobi slums

Population estimate	Statistics				Accuracy metrics		
	Mean	Median	Min	Max	RMSE	R-squared	MAE
ONEKANA	596,83	581,85	315,69	1500,34	307,28	0,48	199,13
Census	350,12	264,16	0,1	1472,10	191,30	0,02	191,19
HRSL	266,22	239,43	0	1218,14	189,30	0,04	189,23
WorldPop	447	326	9	3238	193,80	0,02	193,64
GRID 3	374	361	0	1231	187,80	0,06	187,75

**Fig. 14.8** Comparison of population estimates in Mukuru Slum, with ONEKANA (population counts from HH surveys) on top

ONEKANA population account for more than one million people, whereas the census estimates less than half a million. These results highlight the need to produce reliable population datasets to have a more reliable and complete understanding of the urban population, accounting for the slum population and “Leave No One Behind”.

## 14.5 Conclusion

To reduce urban poverty, including upgrading and planning slum areas and providing slum dwellers with services, it is necessary to improve statistics on the urban population. For this purpose, the extent, nature, and location of slums are needed. Making the population of urban slums visible will help design suitable urban policies, such as the provision of local services. According to our findings, the use

of satellite imagery combined with in situ data collected by citizen science allows us to create generalisable, gridded estimates of slum populations. The  $R^2$  of the overall model (i.e. when modelling all slums together) is 0.48, although it is promising that when isolating the modelling of individual slums, in some cases the  $R^2$  reaches 0.9. This indicates that there is a relationship between the urban characteristics of slums and population distribution patterns. Our work provides insights as to how urban population models should tackle slum areas, as there is currently a lack of ad hoc approaches. The knowledge gained will contribute to a better understanding of the evolution of sub-Saharan African cities, enhancing evidence-based policymaking and ensuring sustainable urban growth.

**Acknowledgements** The research pertaining to these results received financial aid from the Belgian Federal Science Policy (BELSPO) according to the agreements of subsidy no. SR/11/380 (SLUMAP), SR/11/405 (ONEKANA), and from NWO grant number VI. Veni. 194.025.

## References

- Abascal A, Rodríguez-Carreño I, Vanhuyse S, Georganos S, Sliuzas R, Wolff E, Kuffer M (2022) Identifying degrees of deprivation from space using deep learning and morphological spatial analysis of deprived urban areas. *Comput Environ Urban Syst* 95:101820
- Benza M, Weeks JR, Stow DA, Lopez-Carr D, Clarke K (2016) A pattern-based definition of urban context using remote sensing and GIS. *Remote Sens Environ* 183:250–264
- Birkin M, Clarke M (1988) Synthesis – a synthetic spatial information system for urban and regional analysis: methods and examples. *Environ Plan A* 20(12):1645–1671
- Boo G, Darin E, Thomson DR, Tatem AJ (2020) A grid-based sample design framework for household surveys. *Gates Open Res* 4
- Boo G, Leasure DR, Dooley CA, Chamberlain HR, Lázár AN et al (2022) High-resolution population estimation using household survey data and building footprints. *Nat Commun* 13(1):1–10
- Carr-Hill R (2013) Missing millions and measuring development progress. *World Dev* 46:30–44
- CIESIN (2016) Facebook Connectivity Lab and Center for International Earth Science Information Network – High-Resolution Settlement Layer (HRSL). <https://ciesin.columbia.edu/data/hrsl/>. Accessed April 2023
- Cottam G, Curtis JT, Catana AJ (1957) Some sampling characteristics of a series of aggregated populations. *Ecology* 38(4):610–622
- Dooley CA, Tatem AJ, Bondarenko M (2020) Gridded maps of building patterns throughout sub-Saharan Africa, version 1.0. University of Southampton: Southampton, UK. Source of building Footprints “Ecopia Vector Maps Powered by Maxar Satellite Imagery”
- Duque JC, Laniado H, Polo A (2018) S-maup: statistical test to measure the sensitivity to the modifiable areal unit problem. *PLoS One* 13(11)
- Eicher CL, Brewer CA (2001) Dasymetric mapping and areal interpolation: implementation and evaluation. *Cartogr Geogr Inf Sci* 28(2):125–138
- Gadiaga AN, Abbott TJ, Chamberlain H, Lloyd CT, Lazar AN, Darin E, Tatem AJ (2021) Census-disaggregated gridded population estimates for Kenya, version 1.0. University of Southampton. <https://doi.org/10.5258/SOTON/WP00747>
- Gehlke CE, Biehl K (1934) Certain effects of grouping upon the size of the correlation coefficient in census tract material. *J Am Stat Assoc* 29(185A):169–170

- Georganos S, Grippa T, Niang Gadiaga S, Linard C, Lennert M, Vanhuyse S, Mboga N, Wolff E, Kalogirou S (2019) Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto Int* 36(2):121–136
- Georganos S, Abascal A, Kuffer M, Wang J, Owusu M, Wolff E, Vanhuyse S (2021) Is it all the same? Mapping and characterizing deprived urban areas using WorldView-3 superspectral imagery. A case study in Nairobi, Kenya. *Remote Sens* 13(24):4986
- Goodchild MF, Lam NSN (1980) Areal interpolation: a variant of the traditional spatial problem. *Geo-processing* 1(3):297–312
- GRID3 (2023) National population estimates v1.2. Kenya. <https://grid3.org/resources/data>. Accessed April 2023
- KNBS (2019) 2019 Kenya population and housing census results. <https://www.knbs.or.ke/2019-kenya-population-and-housing-census-results/>. Accessed April 2023
- Kuffer M, Owusu M, Oliveira L, Sliuzas R, Van Rijn F (2022) The missing millions in maps: exploring causes of uncertainties in global gridded population datasets. *ISPRS Int J Geo Inf* 11(7):403
- Leyk S, Gaughan AE, Adamo SB, de Sherbinin A, Balk D, Freire S, Rose A, Stevens FR, Blankespoor B et al (2019) The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use. *Earth Syst Sci Data* 11:1385–1409
- Linard C, Gilbert M, Tatem AJ (2011) Assessing the use of global land cover data for guiding large area population distribution modelling. *GeoJournal* 76:525–538
- Lloyd CT, Sorichetta A, Tatem AJ (2017) High-resolution gridded population dataset data for use in population studies. *Nat Sci Data* 4:170001
- Mennis J, Hultgren T (2006) Intelligent Dasyymmetric mapping and its application to areal interpolation. *Cartogr Geogr Inf Sci* 33(3):179–194
- Meta (2023) Data for good. <https://dataforgood.facebook.com/>. Accessed April 2023
- Nilsson M, Griggs D, Visbeck M (2016) Policy: map the interactions between sustainable development goals. *Nature* 534:320–322
- Openshaw S (1984) The modifiable areal unit problem. *Norwick: geo books. mapping population distribution from high resolution remotely sensed imagery in a data poor setting. Remote Sens* 10(9)
- Pearce JR, Richardson EA, Mitchell RJ, Shortt NK (2010) Environmental justice and health: the implications of the socio-spatial distribution of multiple environmental deprivation for health inequalities in the United Kingdom. *Trans Inst Br Geogr* 35:522–539
- POPGRID (2020) Global population grids: summary characteristics. [www.popgrid.org/data-docs-table1](http://www.popgrid.org/data-docs-table1). Accessed April 2023
- Reed FJ, Gaughan AE, Stevens FR, Yetman G, Sorichetta A, Tatem AJ (2018) Gridded population maps informed by different built settlement products. *Data* 3(3):33
- Reibel M, Bufalino ME (2005) Street-weighted interpolation techniques for demographic count estimation in incompatible zone systems. *Environ Plan A* 37(1):127–139
- Sirko W, Kashubin S, Ritter M, Annkah A, Bouchareb YSE, Dauphin Y, Keyzers D, Neumann M, Cisse M, Quin JA (2021) Continental-scale building detection from high resolution satellite imagery. *ArXiv:2107.1283.s*
- Sorichetta A, Hornby GM, Stevens FR, Gaughan AE, Linard C, Tatem AJ (2015) High-resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015, and 2020. *Sci Data* 2(1):1–12
- Stevens FR, Gaughan AE, Linard C, Tatem AJ (2015) Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS One* 10(2):e0107042
- Thomson D, Rhoda D, Tatem A, Castro M (2020) Gridded population survey sampling: a systematic scoping review of the field and strategic research agenda. *Int J Health Geogr* 19:34

- UN-Habitat (2015) UN General Assembly, transforming our world: the 2030 agenda for sustainable development, A/RES/70/1. <https://www.refworld.org/docid/57b6e3e44.html>. Accessed April 2023
- Wang J, Fleischmann M, Venerandi A, Romice O, Kuffer M, Porta S (2023) EO + Morphometrics: understanding cities through urban morphology at large scale. *Landscape Urban Planning* 233:104691. <https://doi.org/10.1016/j.landurbplan.2023.104691>
- Weigand M, Wurm M, Dech S, Taubenböck H (2019) Remote sensing in environmental justice research – a review. *ISPRS Int J Geo Inf* 8(1)
- Wong DW (2009) Modifiable areal unit problem. In: Kitchin R, Thrift N (eds) *International encyclopedia of human geography*. Elsevier, Oxford, pp 169–174. <https://doi.org/10.1016/B978-008044910-4.00475-2>
- WorldPop (2023) Population counts. <https://hub.worldpop.org/project/categories?id=3>. Accessed April 2023
- Zhang M, Kukadia N (2005) Metrics of urban form and the modifiable areal unit problem. *Transportation Research Record* 1902:71–79

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

