Response to Editor and Reviewers' Comments on Earth System Science Data manuscript ESSD-2025-260

Title: Geospatial micro-estimates of slum populations in 129 Global South countries using machine learning and public data

Reviewers Comments:

Anonymous Referee #3

This study proposes a standardized and scalable bottom-up approach for estimating the high-resolution spatial distribution of slum populations in Global South countries. This approach integrates household survey data, satellite imagery, and gridded population data, and employs machine learning and transfer learning techniques (e.g., ResNet-34 and XGBoost). It generates slum population distribution maps with a spatial resolution of approximately 6.72 km (at the equator) across 129 Global South countries. However, the current manuscript is insufficient in its discussion of the accuracy and reusability of this geospatial data product, and there are the following issues that need clarification:

Response: We sincerely appreciate your constructive comments regarding the accuracy and reusability of our data product. In response, we have expanded the discussion to address the spatial accuracy, visualization validation and application scenarios. Please see our point-by-point responses below.

1. What standards are used to define slums? When conducting mapping for Global South countries, how is the consistency of slum feature extraction ensured across different regions and data sources to guarantee the accuracy of the product?

Response: In this study, we define the slum population following UN-Habitat's operational definition of a slum household (UN-Habitat, 2021): a household lacking one or more of the following: (i) access to improved water sources, (ii) access to improved sanitation facilities, (iii) sufficient living area, (iv) durable housing, or (v) security of tenure. Because the last criterion is particularly difficult to assess (UN-Habitat, 2003), we exclude security of tenure from our final definition. These details are clarified in Methods (Section 3.1) and the indicator framework is summarized in Table 2.

We recognize that slum characteristics vary across countries and regions. To ensure consistent feature extraction across the Global South, we adopt the following measures:

1) Standardized definition and indicators: a widely recognized, operational

framework from UN-Habitat.

2) Transfer learning with enriched inputs: ResNet pretraining on >100 million ImageNet images to capture broadly representative visual features, combined with expanding inputs from 3 RGB bands to 7 bands (including near-infrared, etc.) to enhance spatial–spectral richness.
3) Regional stratification for generalization: models stratified by geographic region and national income group to improve transferability and reduce domain shift.
4) Robust data foundations: publicly available Landsat imagery and comprehensive, representative DHS household data for ground truth.

Together, these choices promote consistent and accurate extraction of slum-related features across diverse settings and highlight the scalability advantages of our framework for large-area applications. We have added a dedicated discussion on consistency in slum feature extraction in the revised version as follows.

Discussions

…

Slum characteristics vary across countries and regions (da Fonseca Feitosa et al., 2021; do Nascimento et al., 2025), making consistent feature extraction at large spatial scales challenging. We address this systematically across the workflow, from data selection to model design and feature engineering. At the core we adopt transfer learning with Resnet-34 pre-trained on over 100 million ImageNet samples and adapt it to 7-channel multispectral patterns associated with deprived housing conditions. While the original RGB-pretrained ResNet-34 is optimized for general features (e.g., edges and textures), fine-tuning enables early and intermediate filters to learn context-specific cues, such as roof material, settlement density, and vegetation coverage, while leveraging the added information from near-infrared and other bands. The combination of broadly learned representations and increased spectral richness provides a robust foundation for identifying spatial and morphological signatures of slum environments. The approach outperforms RGB-only baseline, highlighting the value of multispectral fine-tuning and high-dimensional feature extraction for modeling housing-related deprivation.

During the model development phase, we constructed regionally stratified models based on geographic location and national income level, rather than separate country-specific models, to improve generalizability and ensure consistent feature extraction across heterogeneous contexts. In addition, the use of publicly available Landsat imagery and DHS ground-truth data provides a uniform, reproducible basis for cross-country analysis. Collectively, these design choices enable consistent and accurate slum feature extraction at scale and underpin the production of a harmonized slum population product.

…

References:

Da Fonseca Feitosa, F., Vasconcelos, V. V., de Pinho, C. M. D., da Silva, G. F. G., da Silva Gonçalves, G., Danna, L. C. C., & Lisboa, F. S. (2021). IMMerSe: An integrated methodology for mapping and classifying precarious settlements. Applied geography, 133, 102494.

Do Nascimento, G.A., Giannotti, M., Regueira, T.A. and Tomasiello, D.B., 2025. Identifying slum areas: A multidimensional analysis leveraging with explanatory machine learning techniques. Sustainable Cities and Society, 131, 106645.

UN-habitat. Urban indicators database. https://data.unhabitat.org/pages/housing-slums-and-informal-settlements, 2021

UN-Habitat. The challenge of slums: global report on human settlements 2003. Management of Environmental Quality: An International Journal 15, 337-338, https://doi.org/10.1108/meq.2004.15.3.337.3 , 2004.

2. Visualization is an important basis for verifying product accuracy. Can the authors provide visualized comparison results of selected regional cases? By comparing with known data or real-world conditions, further verification of the product's accuracy could be achieved.

Response: We appreciate your valuable suggestion to include visual comparisons. Following your advice, we have added comparisons between our slum population estimates and observed conditions for four representative Global South cities, including (a) Nairobi, Kenya; (b) Cape Town, South Africa; (c) Medellín, Colombia; and (d) Dhaka, Bangladesh. We overlay the gridded slum population estimates on high-resolution satellite imagery and Google Earth imagery. The resulting maps show consistent spatial correspondence between areas with high estimated slum populations and known locations of dense slums areas or informal settlements, supporting the reliability of our mapping results. These visualizations are included in the revised main text as Figure 6 (pasted below for the reviewing convenience).

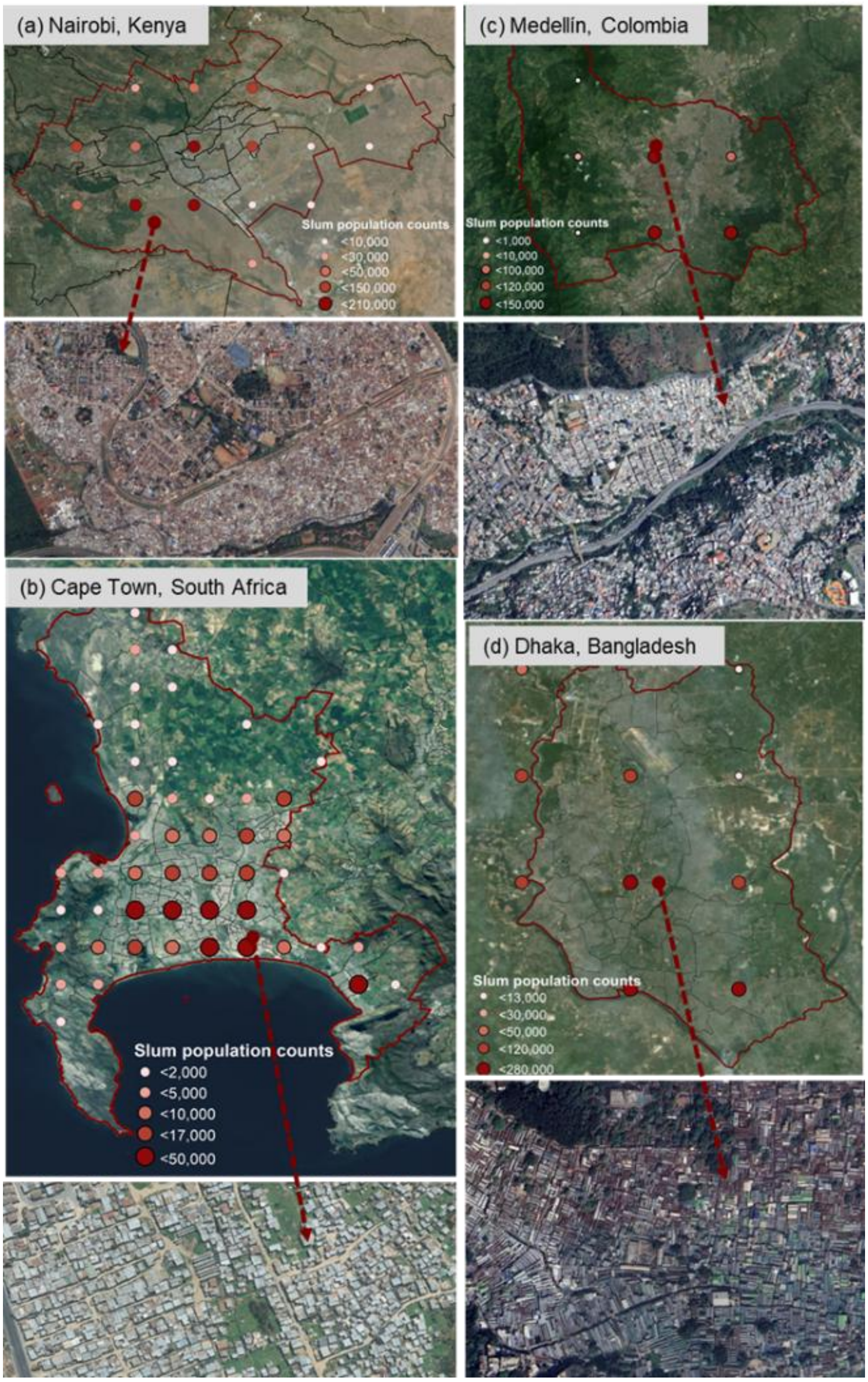

Figure 6. Showcases of visual comparisons between estimated slum population distribution and conditions observed in Google Earth imagery for four representative Global South Cities: (a) Nairobi, Kenya, (b) Cape Town, South Africa, (c) Medellín, Colombia, and (d) Dhaka, Bangladesh. Estimated population counts are shown as circular markers overlaid on Landsat basemaps, with each marker representing a 6.72 km × 6.72 km grid cell. Red arrows indicate locations where high estimated slum populations align with visible informal settlements in Google Earth, providing visual validation of the product.

3. Compared with traditional slum population estimation methods and existing relevant data, what advantages does the current product have in terms of spatial continuity? How to demonstrate these advantages from both theoretical and practical application perspectives?

Response: Thanks for your valuable comments. We agree that the advantage of spatial continuity better highlights the strengths of our product. Traditional approaches to estimating slum populations typically rely on (1) census or household survey aggregation or (2) the overlap of morphological slum boundaries with gridded population maps. While census and survey data provide official, well-validated estimates, they are often available only at coarse administrative levels or as sample-based estimates, limiting spatial detail and coverage. Likewise, combining mapped slum boundaries with population grids can pinpoint settlement locations but is constrained by the discrete nature of those boundaries and the uneven availability of high-quality maps—largely limited to a few city hotspots (e.g., Cape Town, Mumbai, Dhaka).

By contrast, our framework produces spatial continuous estimates, meaning that slum population is estimated for every 6.72 km grid cell covering inhabited areas. This does not imply continuously designated slum locations; rather, it yields a wall-to-wall grid that (i) captures local variation and spatial gradients, overcoming the aggregation effects of administrative units and survey areas; (ii) does not depend on the existence of pre-mapped slum boundaries, allowing coverage in regions lacking such data; and (iii) can be dynamically updated as new imagery and ancillary data become available. In practice, this enables flexible aggregation across multiple spatial scales (neighborhood, city, national), supports integration with research on urban inequality, sustainability, environmental exposure, and spatial justice, and provides a scalable, data-driven foundation for human-centered urban planning, targeted resource allocation, and evidence-based policy design.

In line with response to your comment #7, we have added discussion of these theoretical and practical benefits and outlined specific application scenarios in Discussion (pasted below for the reviewing convenience).

Discussions

The product advances a human-centered mapping framework, prioritizing populations living in inadequate housing rather than focusing solely on the physical delineation of slum settlements. Theoretically, it exposes gaps in progress toward the Sustainable Development Goals by providing spatially continuous, gridded estimates that directly quantify the number of people affected by deficits in water, sanitation, and electricity. These wall-to-wall estimates reveal local variation and spatial gradients that are often obscured by administrative boundaries or survey aggregates. Because the approach does not depend on pre-defined morphological slum boundaries, it supports population estimation in regions lacking boundary data and can be dynamically updated as new datasets become available.

Practically, this continuous representation enables flexible aggregation across multiple spatial scales (e.g., neighborhood, city, national) and provides a robust, data-driven foundation for human-centered urban planning, targeted resource allocation, and evidence-based policy design. For example, planners can optimize the spatial distribution of schools, clinics, and parks to improve accessibility and equity for vulnerable populations, thereby supporting inclusive urban development. These estimates also inform resource allocation and emergency response planning aligned with actual population needs. When combined with spatial information on environmental hazards, such as floods, heat islands, or air pollution, the gridded product becomes a critical tool for resilience planning and adaptive governance, enabling social protection and welfare programs tailored to those at greatest risk.

4. This study mentions the issue of underestimation of the actual slum population size. What strategies does the current product adopt to solve this problem? To what extent can it improve the underestimation? Can both quantitative verification (e.g., the proportion of error reduction, the degree of proximity between estimated values and actual values) and qualitative verification (e.g., the rationality of the method, the completeness of data coverage) be provided?

Response: Thank you for raising this concern about strategies to address the underestimation. Many prior approaches estimate slum populations by overlapping mapped (morphological) slum boundaries with population grids at the settlement level. Recent studies show these methods can underestimate slum populations relative to literature or official statistics, largely due to inaccuracies in slum geometries and uncertainties in the underlying population datasets.

1) What strategies does the current product adopt to solve this problem?

Response: We take a direct estimation approach by predicting the proportion of slum households within each enumeration area, thereby reducing errors introduced by post-hoc scaling. Our strategy has three components: (i) Modelling design: we define slums from a household perspective rather than relying on binary morphological maps (which are unavailable at scale). This household-based framework provides a more direct basis for estimating slum populations. (ii) Ground-truth labeling: we use DHS surveys for household labels. DHS employs standardized instruments and rigorous quality controls, providing reliable and valid measures of slum households. (iii) Population dataset selection: we adopt GHS-POP, which refines population estimates along the urban gradient using human settlement information and share a production framework with other datasets used in this study, thereby reducing uncertainties from heterogeneous sources. These points are elaborated in the revised Discussion to clarify how our method mitigates underestimation, as follows:

Discussions

…

Previous studies have addressed slum population underestimation by scaling single-slum estimates to match literature-reported values and then applying these ratios to adjust aggregated population totals (Breuer et al.,2024; Thomson et al., 2021). In contrast, we adopt a direct estimation approach that predicts the proportion of slum households within each enumeration area. First, we operationalize slums from a household perspective by constructing a household-level indicator, expressed as the share of slum households among all households in an enumeration area, and use deep learning to link this indicator directly to features extracted from remote sensing imagery. This design avoids reliance on dichotomous morphological slum maps (which are rarely unavailable at scale) and provides a more flexible, generalizable framework. Second, we use high-quality DHS survey data as labels; their standardized instruments and rigorous quality-control protocols yield valid household-level indicators essential for accurate prediction. Finally, for population distribution, we employ GHS-POP, which refines population estimates along the urban gradient and share a production framework with the GHSL datasets used in this study, thereby reducing uncertainties arising from heterogeneous sources.

2) To what extent can it improve the underestimation? Can both quantitative verification (e.g., the proportion of error reduction, the degree of proximity between estimated values and actual values) and qualitative verification (e.g., the rationality of the method, the completeness of data coverage) be provided?

Response: Our results include both quantitative and qualitative verification.

(A) quantitative verification: We validate the predicted slum household indicators against DHS-derived ground-truth indicators using RMSE and R^2. We also compare our slum population estimates with: (i) scaled-up city-level estimates derived from multiple individual slum samples in the literature, (ii) official city statistics, (iii) country-level statistics reported by World Bank, and (iv) regional-level statistics reported by UN-Habitat. Across these benchmarks, our estimates fall within the range of literature-based scaled-up values and show strong alignment at city, country, and regional scales

(Figure 3, Table 6 and Figure 7 in the revised version, which are presented below for review convenience).

(B) qualitative verification. We discuss the methodological soundness and data coverage of our framework, including the rationale for key dataset choices—DHS surveys (for standardized, quality-controlled household labels), Landsat imagery (for globally consistent, multispectral inputs), and GHS-POP (for human-settlement–informed population distribution). We also provide robustness analyses that explain how these selections support consistent performance and broad geographic applicability.

Results

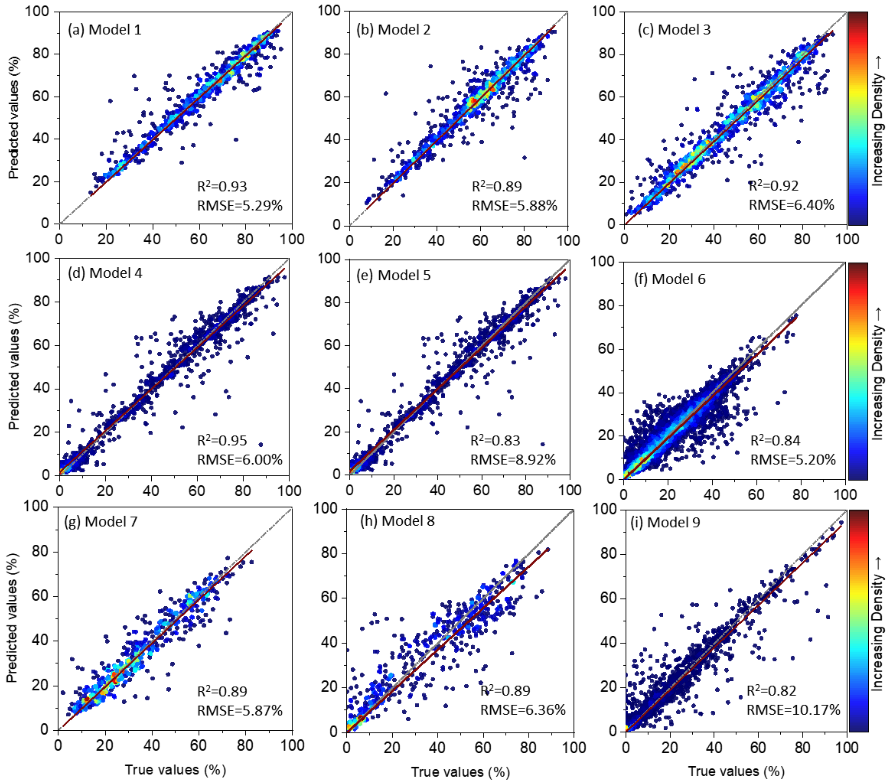

Figure 3 Model performances across different geographical regions and income groups. The true values represent the slum indicator calculated from DHS data using our slum framework, while the predicted values are produced by our machine learning-based models. For each model, a linear regression line is displayed alongside a reference line with a slope of one for comparison.

Table 6 Comparison of model-derived slum population estimates with scaled-up figures from sampled slums and published city-level statistics, circa 2015.

City	Our estimate (million)	Scaled-up estimates (million, from sampled slums in other literature)	City-level statistics in other literature (million)	References
Rio de Janeiro	3.29	0.96, 1.17, 1.4, 1.46, 1.54, 1.58, 1.91, 2.39, 3.98	2.00	Breuer et al. (2024); Trindade et al. (2021);
São Paulo	2.09	1.3, 1.41, 3.02, 5.67, 8.62	2.16	Breuer et al. (2024); Trindade et al. (2021)
Cape Town	1.45	1.08, 1.13, 1.27, 1.29, 1.38, 1.54, 1.55, 1.89, 2.05, 2.15, 2.2, 2.8	1.23	Breuer et al. (2024)
Cairo	2.56	0.01, 0.3, 0.71, 3.75, 4.55, 7.47	/	Breuer et al. (2024); Sabry (2009);
Mumbai	2.70	1.65, 1.68, 5.2	/	Breuer et al. (2024); Taubenböck and Wurm (2015)
Dhaka	1.85	0.64, 0.76, 1.87, 2.69, 2.73, 3.43, 3.83, 3.84, 4.81, 4.84, 6.61	1.32	Breuer et al. (2024); Patel et al. (2019)
Johannesburg	0.80	/	0.6	Trindade et al. (2021)
Hyderabad	2.28	/	2.3	Trindade et al. (2021)

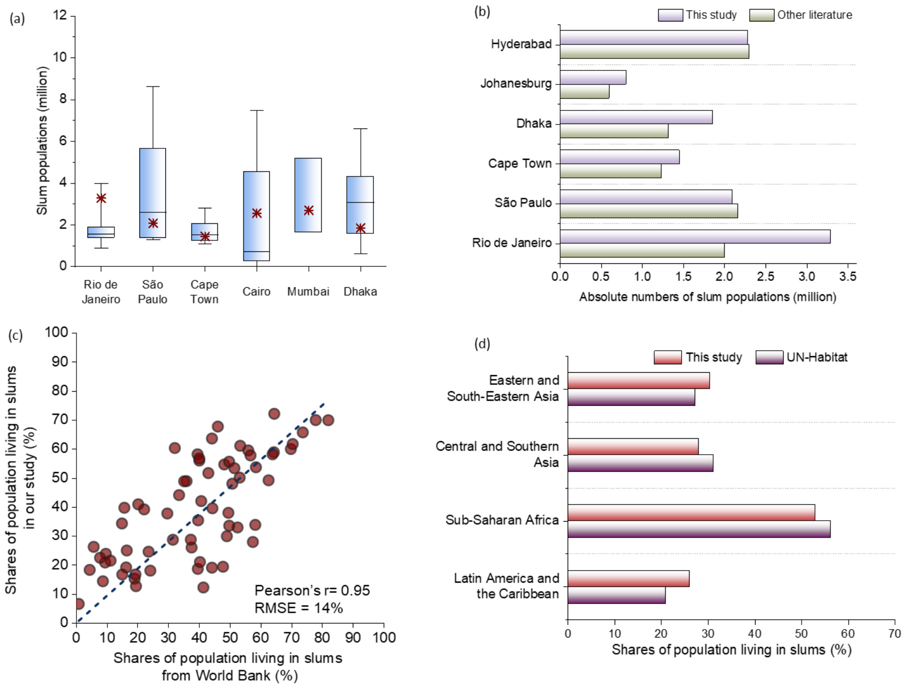

Figure 7 Comparison of our slum population estimates with data from previous literature and official statistics, presented as both absolute numbers and relative shares. (a) Comparison with scaled-up values from previous studies for six cities: Rio de Janeiro (n=9), São Paulo (n=5), Cape Town (n=12), Cairo (n=6), Mumbai (n=3), and Dhaka (n=11). Scaled-up values are derived by extrapolating slum population counts from multiple sampled slums to the city level. Asterisks indicate estimates from this study. Box plots display

the median (central line), interquartile range (box), and minimum-maximum range (whiskers). (b) Comparison with official city-level statistics for the same six cities. (c) Comparison of country-level slum population shares between our estimates and World Bank data; each dot represents a Global South country. (d) Comparison of regional-level slum population shares between our estimates and UN-Habitat data.

Discussions

5.1 Robustness and external validation

Our approach builds on well-established deep learning techniques widely used for mapping poverty, infrastructure access, and progress toward the Sustainable Development Goals (Jean et al., 2016; Chi et al.,2022). We fine-tune the ResNet-34 model on 7-band satellite imagery to better capture the complex spatial-spectral patterns associated with deprived housing conditions…. The original RGB-pretrained ResNet-34 is optimized for general features (e.g., edges and textures), fine-tuning enables early and intermediate filters to learn context-specific cues, such as roof material, settlement density, and vegetation coverage ...

5.2 Spatial resolution and dataset selection

… The performance of any machine learning or deep-learning model relies heavily on the quality and suitability of its trained data (Meena et al., 2023). The DHS datasets provide accurate, representative, household-based survey data on demographic and socioeconomic conditions; our samples cover over 1 million households across 67,204 clusters in 53 countries of the Global South.

… To enrich the input feature space, we prioritize images with more spectral bands than standard RGB. Landsat imagery offers consistent temporal coverage, global reach, and multiple spectral bands at 30-meter resolution, making it well suited to our task (Wulder et al., 2022)

… Grided population data are also critical for measuring and mapping slum populations. Widely used global products, such as GPWv4.11 (CIESIN et al., 2018), GHS-POP (Pesaresi et al., 2016), and World Pop (Stevens et al., 2015), as well as regional/national products like HRSL (Smith et al., 2019), differ markedly in inputs, ancillary datasets, and dasymetric methods for redistributing population to grid cells. While regional products such as HRSL can achieve high accuracy at finer resolutions, their limited spatial coverage impedes large-scale, cross-country comparisons, especially in the Global South. Following the 'fitness-for-use' principle (Juran et al., 1979), we select GHS-POP for its distinctive advantages aligned with our objectives. Because it is grounded in human-settlement information, GHS-POP effectively refines population estimates along the urban gradient, which is critical for slum population modeling. Moreover, the Global Human Settlement Layer (GHSL), which leverages Landsat imagery, aligns with the satellite images used in this study, reducing uncertainties introduced by heterogeneous sources.

Reference:

Breuer, J.H., Friesen, J., Taubenböck, H., Wurm, M. and Pelz, P.F. (2024). The unseen population: Do we underestimate slum dwellers in cities of the Global South? Habitat International, 148, p.103056.

Thomson, D.R., Gaughan, A.E., Stevens, F.R., Yetman, G., Elias, P. and Chen, R. (2021). Evaluating the accuracy of gridded population estimates in slums: A case study in Nigeria and Kenya. Urban Science, 5(2), p.48.

Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B., Ermon, S. (2016) Combining satellite imagery and machine learning to predict poverty. Science 353, 790-794.

Chi, G., Fang, H., Chatterjee, S., Blumenstock, J.E. (2022). Microestimates of wealth for all low-and middle-income countries. Proceedings of the National Academy of Sciences 119, e2113658119.

5. Does the data used by the authors to verify the product have the underestimation issue as mentioned? If it does, how to ensure the validity of accuracy verification under such circumstances? Are there any other verification methods or supplementary data to enhance the reliability of the product?

Response: Thanks for your valuable comments. We agree that the reliability of validation data is critical. To assess our product as comprehensively as possible, we validate against multiple sources and spatial scales—from micro-level household surveys to macro-level regional and national statistics. Specifically, we draw on DHS household survey data, literature-reported estimates, official statistics, and datasets from international organizations (e.g., the World Bank, UN-Habitat), enabling comparisons at city, national, and regional levels. By reducing dependence on any single dataset, we strengthen the credibility of our accuracy assessment.

In Methods, we describe our external comparisons. We have also added a discussion on the quality and limitations of the validation datasets, as follows:

Discussions

…

Another important consideration for accuracy verification is that external validation datasets may themselves contain uncertainties and limitations, stemming from inconsistent reporting standards, heterogeneous data collection methods, or time lags in statistical updates. To address this, we minimize reliance on any single source by employing a diverse suite of validation data across multiple spatial scales. At the micro level, DHS household surveys provide standardized, representative indicators of slum households based on rigorous sampling. At the meso level, literature-reported estimates and city-level statistics, including both scaled-up values from individual slum areas and directly reported city totals, offer informative reference ranges. At the macro level, national and regional statistics from international organizations such as the World Bank and UN-Habitat provide cross-country benchmarks. Validating against multiple sources and scales strengthens robustness and reduces the risk of bias associated with any single dataset.

6. Currently, the product only includes data from 2018. Does the current method have transferability, and can it be applied to data processing in different years in the future to support temporal analysis of slum populations? If transferable, what conditions need to be met or what adjustments need to be made? This requires in-depth discussion.

Response: Thank you for highlighting the transferability of our approach. Our framework is transferable and supports dynamic, temporal analyses of slum populations. Although the current product focuses on 2018, this year was chosen because it offers the most complete, high-quality combination of ground-truth surveys and satellite datasets, providing a robust benchmark for methodological demonstration. The framework is particularly valuable in data-scarce regions, where satellite imagery can be used to generate spatial predictions of slum populations. Temporally, the model can be extended to future years by leveraging patterns learned from historical data in combination with updated satellite and nighttime light imagery. Nonetheless, we recommend validation against available ground-truth or survey data, as rapid urban development can significantly alter slum population distributions. We have added the following discussion in the revised manuscript, as follows:

Discussions

Our framework is transferable and supports dynamic, temporal analyses of slum populations. It enables forward-looking predictions by leveraging patterns learned from existing datasets in combination with updated Landsat imagery and nighttime lights data. Through data-driven learning, the framework extracts informative features from satellite imagery and captures the relationships between spectral characteristics and slum household proxies, facilitating generalization to new spatial and temporal contexts. To ensure reliable predictions, input satellite datasets must be consistent in resolution and spatial extent, with preprocessing steps such as cloud masking carefully applied. Recalibration using newly released ground-truth survey data is recommended, as these labels provide an empirical benchmark to evaluate model performance and correct biases, particularly in rapidly urbanizing areas where slum populations may change significantly. By integrating spatial and temporal information, the framework provides a flexible and robust tool for rapid monitoring and spatiotemporal prediction in data-scarce and fast-changing contexts.

7. What are the specific application scenarios of the current product? For example, in urban planning, resource allocation, and social policy formulation, how can the data and results provided by this product be used to guide practical work?

Response: We agree that specifying potential application scenarios is critical for downstream use of this slum population product. Our product advances a human-cantered framework that prioritizes population living in inadequate housing rather than focusing solely on the physical delineation of slum settlements. Gridded slum population estimates provide a direct measure of how many people are affected by deficits in housing and infrastructure services, better reflecting the scale of human vulnerability and directing interventions toward densely populated areas where needs and risks are greatest. In line with your suggestions in comments 3 and 7, we have added a discussion of implications for urban planning, resource allocation, and social policy in Discussion, as follows:

Discussions

… Practically, this continuous representation enables flexible aggregation across multiple spatial scales (e.g., neighborhood, city, national) and provides a robust, data-driven foundation for human-centered urban planning, targeted resource allocation, and evidence-based policy design. For example, planners can optimize the spatial distribution of schools, clinics, and parks to improve accessibility and equity for vulnerable populations, thereby supporting inclusive urban development. These estimates also inform resource allocation and emergency response planning aligned with actual population needs. When combined with spatial information on environmental hazards, such as floods, heat islands, or air pollution, the gridded product becomes a critical tool for resilience planning and adaptive governance, enabling social protection and welfare programs tailored to those at greatest risk.