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

Overall Response: We would like to thank the editor and reviewer #4 for their detailed comments and suggestions, which are very helpful for improving the quality of the manuscript. Please see our point-by-point response to all the comments and suggestions from the reviewers. Text marked in red denotes newly added or amended content.

Editor's comments

The paper makes a valuable and timely contribution by providing a large-scale, spatially continuous slum population dataset with strong validation results. However, the current presentation is not sufficiently accessible to interdisciplinary readers and policy users, and several methodological choices require clearer justification and more prominent communication. I therefore invite a minor revision focused on improving clarity and transparency.

Overall Response: Thank you for summarizing these key revision points. We have revised the manuscript accordingly to improve clarity, accessibility, and interpretability for interdisciplinary readers. We have also strengthened the clarification and justification including methodological choices and communication. Please find our point-by-point responses below.

Please address the following in a revision:

1. Add brief plain-language explanations for key technical terms/datasets at first mention (e.g., ResNet-34, XGBoost, DHS clusters, GHS-POP, SDG 11.1).

Response: Thank you for the helpful comment. In response to your comment and Reviewer #4's Major comment #1, we have added plain-language explanations at the first mention of key terms and datasets, including ResNet-34, XGBoost, DHS clusters, GHS-POP, and the SDG 11.1 framework. The revisions are as follows (Lines116-117, 325-326, 201-202, 214-215, 245-247):

- *Using this data, we construct a slum indicator framework and fine-tune the ResNet-34, a widely used deep learning model for image recognition, to extract features from Landsat and nighttime light imagery.*
- *Subsequently, we used eXtreme Gradient Boosting tree (XGBoost), a popular and flexible supervised-learning algorithm, to predict the occurrence of slum households in each cluster from 512 features.*

- *The input images used for model training were centered on the geographic coordinates of each DHS cluster, a small geographic survey unit based on a census enumeration area…*
- *We utilized and processed the Global Human Settlement Population dataset (GHS-POP), a global population dataset that distributes census counts into 1-km grid cells developed by the EU Joint Research Center, to estimate the slum population (Schiavina et al., 2022).*
- *We used the SDG 11.1 framework, referring to the United Nations indicator framework for monitoring access to adequate, safe, and affordable housing and basic services, as a proxy to estimate the occurrence of households living in slums or slum-like conditions within a cluster.*

2. Clarify and more prominently acknowledge the origin and implications of the 6.72 km grid resolution, including which inputs impose it and why higher-3. resolution alternatives were not adopted.

Response: Thank you for the helpful comment. In response to your comment and Reviewer #4's Major comment #2, we have clarified that the 6.72 km resolution is derived from the Landsat-based 224 × 224 image-tile structure and explicitly stated that the dataset is intended for broad-scale comparative analysis rather than site-specific mapping. We have also explained why higher-resolution alternatives such as Sentinel-2 and PlanetScope were not used as primary inputs. The revisions are as follows (Lines26-32, 37-41, 206-210, 181-188):

Abstract

…*Here, we develop a standardized bottom-up approach to estimate slum populations at the grid-cell level (~6.72 km resolution at the equator) for 129 Global South countries in 2018. Leveraging the Sustainable Development Goal 11.1 framework and machine learning, our estimation integrates publicly accessible household-based surveys, satellite imagery, and gridded population data. The resolution is thus jointly determined by the model input patch size and the spatial resolution of Landsat imagery. …To our knowledge, this is the first comprehensive geospatial inventory of slum populations across Global South countries. Although not designed for site-specific operational applications, the dataset provides valuable insights for national and regional urban sustainability assessments and further research on vulnerable populations.*

Section 2.2 Satellite imagery

…The Landsat image patches were set to 224 × 224 pixels to match the input size required by our convolutional neural network architecture and to cover the spatial displacement of DHS cluster coordinates. Consequently, our mapping resolution is determined to be 6.72 kilometers at the equator (30m-pixel size × 224 pixels = 6.72 km).

…Higher-resolution alternatives, including Sentinel-2 and PlanetScope, were considered but

not used as primary inputs. For example, Sentinel-2 offers 10-20 m multispectral imagery, but its shorter historical record limits temporal consistency with many DHS survey years. PlanetScope provides approximately 3m imagery, but its commercial licensing, cost, and data-access constraints limit reproducibility for global-scale analysis. We therefore relied on Landsat imagery because it provides publicly available, long-term, and globally consistent observations suitable for cross-country analysis.

3. Make key assumptions and uncertainty handling easier to find (notably DHS geolocation displacement and household-size assumptions; include a short sensitivity summary in the Results).

Response: Thank you for the helpful comment. In response to your comment and Reviewer #4's Minor comment #6 and #11, we have added text explaining how DHS coordinate displacement is mitigated by the 6.72 km grid-cell resolution and moved a short household-size sensitivity summary into the Results section. The revisions are as follows (Lines 348-352, 635-641):

Section 3.3

…Since DHS cluster coordinates are randomly displaced for confidentiality purposes, with a maximum displacement of 2 km for urban clusters and 5 km for rural clusters, the approximately 6.72 km grid-cell size helps absorb part of this displacement and reduce the influence of positional uncertainty on model training and prediction.

Section 4.4

…To quantify the potential influence of the equal-household-size assumption, we conducted a sensitivity analysis using alternative household-size scenarios. Specifically, we assumed that slum households were, on average, either 10% larger or 10% smaller than non-slum households and recalculated the resulting national slum-population estimates. The results show that assuming a ±10% difference in slum household size changes the national slum population estimates by about 8%. Detailed results are shown in Figure S5.

4. Improve figure readability and interpretability (legends/thresholds stated in the legend, clearer color meaning, and an accessibility note for color-dependent maps).

Response: Thank you for the helpful comment. We have revised the relevant figure captions and legends to define color meanings, thresholds, and accessibility limitations for color-dependent maps. Please see the point-by-point responses to Reviewer #4' Minor comments.

5. Implement the remaining minor textual clarifications suggested by the reviewer (definitions such as "Global South," brief explanations of referenced initiatives/rights language, and rationale for indicator component choices).

Response: Thank you for the helpful comment. We have added or revised short explanations for terms and concepts such as “Global South,” “adequate housing as a human right,” the Know Your City Campaigns, and the rationale for including access to electricity while excluding security of tenure. Please see the point-by-point responses to Reviewer #4’ Minor comments.

Reviewer #4's Comments:

The authors present a valuable and large-scale slum population dataset for the Global South. However, the manuscript assumes a high level of expertise in remote sensing, machine learning, and slum studies, which may hinder interdisciplinary readers and policy makers. The following revisions are recommended to improve clarity and accessibility:

- Add brief plain-language explanations of key technical terms (e.g., ResNet-34, XGBoost, DHS cluster, GHS-POP) at their first appearance.
- Clarify the coarseness of the 6.72 km resolution, explicitly state which input data impose this limit, and discuss whether higher-resolution alternatives were considered.
- Include a simple reader-guide figure linking the workflow to the final outputs.
- Address several minor clarifications (as listed in the review) regarding definitions, assumptions (e.g., equal household size), and figure legends.

Overall Response: Thank you for these valuable suggestions. We agree that improving the clarity and accessibility of the manuscript will benefit interdisciplinary readers and policymakers. In response, we have revised the manuscript as follows:

(1) We have added concise explanations at the first occurrence of these key technical terms, including ResNet-34, XGBoost, DHS cluster, GHS-POP. Please see our response to Major comment #1 for details;

(2) We have clarified the meaning and source of the 6.72 km spatial resolution and we have discussed the reasons why other higher-resolution alternatives are not used. Please see our response to Major comment #2 for details;

(3) We have added a simple reader-guide diagram linking the main workflow steps to the final dataset outputs, please see the revised Figure 1;

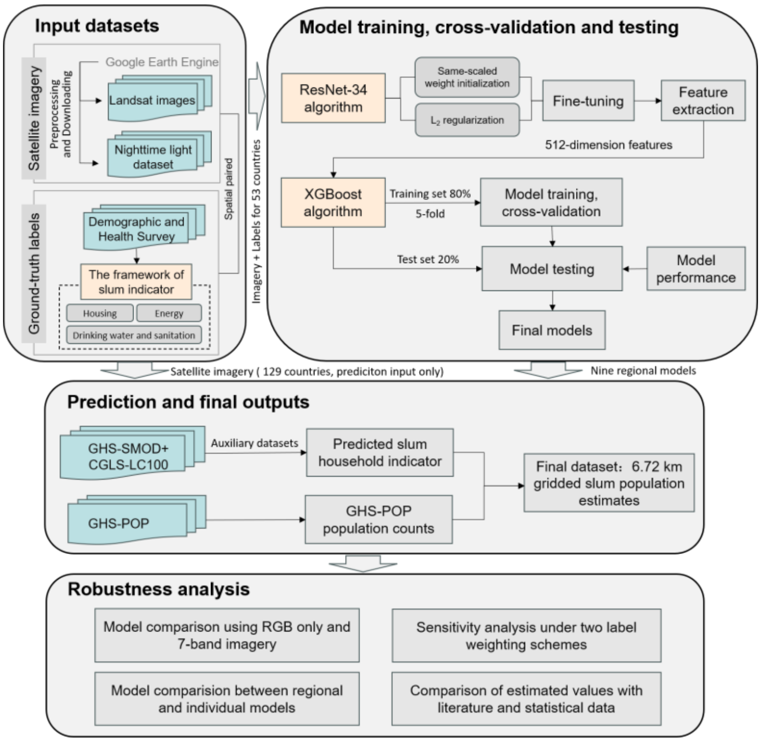


Figure 1 Workflow of this study. *Blue backgrounds represent datasets, yellow backgrounds indicate the indicator framework or model architecture, and gray backgrounds denote specific processes or procedural steps.*

(4) We have addressed the detailed clarification points raised in the review, including definitions, assumptions such as equal household size, and figure legend descriptions. Please see our point-by-point response to Minor comments below.

Summary

This paper presents a standardized methodology that combines machine learning, satellite imagery, and household survey data to estimate slum populations at the grid-cell level (~6.72 km) across 129 Global South countries. The authors construct a slum indicator using DHS survey data, extract features from Landsat and nighttime light imagery with a ResNet-34 network, and then apply XGBoost for prediction, ultimately producing a spatial distribution map of slum populations for the year 2018. The model achieves a median R^2 of 0.89 for within-country predictions and 0.85 for country-level holdout validation. To the best of the authors' knowledge, this is the first large-scale, spatially continuous inventory of slum populations covering the entire Global South.

Response: Thank you for the thoughtful summary of our manuscript.

General Comments

This study involves a large volume of data and a complex methodology, and it carries significant social and policy implications. However, as a reviewer not specialised in slum research, I encountered considerable difficulties in understanding the manuscript. The paper is written under the assumption that readers possess deep background knowledge in remote sensing, machine learning, and slum studies, which is not friendly to interdisciplinary readers or policy makers. The comments below are intended to help the authors improve the readability and transparency of the paper, so that a wider audience (including researchers from other fields and students) can understand and use this valuable data product.

Response: Thank you for the thoughtful and constructive comments. We agree that improving the readability and transparency of the paper is crucial to making the dataset more friendly to interdisciplinary researchers, students, and policymakers, especially for readers without deep background knowledge in remote sensing, machine learning, and slump studies. To this end, we have revised the manuscript to reduce assumed background knowledge and to make the it easier to follow. These revisions are described in detail in the point-by-point responses below.

Major Comments

1. Add in-text explanations of key technical terms at first appearance for non-specialist readers. Many terms (e.g., "ResNet-34", "XGBoost", "DHS cluster", "GHS-POP", "SDG 11.1 framework") are introduced without any intuitive explanation. Readers from other will struggle to follow the logic. I suggest that for each important term, the authors add a short plain-language sentence the first time it appears – for example: "ResNet-34, a widely used deep learning model for image recognition" or "GHS-POP, a global population dataset that distributes census counts into 1-km grid cells."

Response: Thank you for this helpful comment. In the revised version, we have added brief plain-language explanations when key terms first appear in the manuscript, including "ResNet-34", "XGBoost", "DHS cluster", "GHS-POP", "SDG 11.1 framework. The revisions are as follows (Lines116-117, 325-326, 201-202, 214-215, 245-247):

- *Using this data, we construct a slum indicator framework and fine-tune the ResNet-34, a widely used deep learning model for image recognition, to extract features from Landsat and nighttime light imagery.*
- *Subsequently, we used eXtreme Gradient Boosting tree (XGBoost), a popular and flexible supervised-learning algorithm, to predict the occurrence of slum households in each cluster from 512 features.*
- *The input images used for model training were centered on the geographic coordinates of each DHS cluster, a small geographic survey unit based on a census enumeration area…*
- *We utilized and processed the Global Human Settlement Population dataset (GHS-POP), a global population dataset that distributes census counts into 1-km grid cells developed by the EU Joint Research Center, to estimate the slum population (Schiavina et al., 2022).*
- *We used the SDG 11.1 framework, referring to the United Nations indicator framework for monitoring access to adequate, safe, and affordable housing and basic services, as a proxy to estimate the occurrence of households living in slums or slum-like conditions*

within a cluster.

2. Address the coarseness of the 6.72 km resolution and clarify where this resolution comes from. The mapping resolution (6.72 km at the equator) is derived from Landsat 30 m pixels × 224 pixels. However, a 6.72 km grid cell (≈45 km^2 at the equator) is quite coarse for studying slums, which are often small clusters of a few hundred meters. The authors mention in Section 5.2 that this resolution is "not intended for site-specific operational applications", but I suggest they strengthen this limitation earlier in the paper (e.g., in the abstract or results) and explicitly state which input data force this resolution. Also, please clarify whether any alternative higher-resolution data (e.g., Sentinel-2, Planet) were considered and why they were not used. Readers will question whether such coarse cells can meaningfully capture slum populations, especially in heterogeneous urban areas.

Response: Thank you for this helpful comment. We agree that an approximately 6.72 km grid cell is relatively coarse for studying slum. We have therefore revised the Abstract to acknowledge this limitation earlier in the paper, explicitly explain the source of the 6.72 km resolution, and clarify the appropriate scope of application of the dataset. Moreover, in Section 2.2, we have provided a detailed explanation of how the 6.72 km mapping resolution was derived. In addition, we have expanded the Method section to clarify why higher-resolution alternatives, such as Sentinel-2 and Planet imagery, were not used. The revisions are as follows (Lines26-32, 37-41, 206-210, 181-188):

Abstract

…*Here, we develop a standardized bottom-up approach to estimate slum populations at the grid-cell level (~6.72 km resolution at the equator) for 129 Global South countries in 2018. Leveraging the Sustainable Development Goal 11.1 framework and machine learning, our estimation integrates publicly accessible household-based surveys, satellite imagery, and gridded population data. The resolution is thus jointly determined by the model input patch size and the spatial resolution of Landsat imagery. …To our knowledge, this is the first comprehensive geospatial inventory of slum populations across Global South countries. Although not designed for site-specific operational applications, the dataset provides valuable insights for national and regional urban sustainability assessments and further research on vulnerable populations.*

Section 2.2 Satellite imagery

…The Landsat image patches were set to 224 × 224 pixels to match the input size required by our convolutional neural network architecture and to cover the spatial displacement of DHS cluster coordinates. Consequently, our mapping resolution is determined to be 6.72 kilometers at the equator (30m-pixel size × 224 pixels = 6.72 km).

...Higher-resolution alternatives, including Sentinel-2 and PlanetScope, were considered but not used as primary inputs. For example, Sentinel-2 offers 10-20 m multispectral imagery, but its shorter historical record limits temporal consistency with many DHS survey years. PlanetScope provides approximately 3m imagery, but its commercial licensing, cost, and data-access constraints limit reproducibility for global-scale analysis. We therefore relied on Landsat imagery because it provides publicly available, long-term, and globally consistent observations suitable for cross-country analysis.

Minor Comments

1. Line 20 ("Slums are a visible manifestation...") – The term "Global South" is used throughout but never defined in the abstract or main text. Please add a brief definition (e.g., "developing regions as defined by UNCTAD") on first use.

Response: Thank you for this helpful comment. We have added the definition of Global South at the first occurrence in the main text. The revisions are as follows (Lines 110-112):

The Global South refers to the countries and territories in developing regions listed by the United Nations Conference on Trade and Development (UNCTAD, 2025).

2. Line 42–46 – The opening paragraph cites several human-rights references. For a non-legal reader, a one-sentence explanation of why "adequate housing" is a human right would be helpful. Consider a short footnote.

Response: Thank you for this helpful comment. Considering the ESSD manuscript formatting guidance that "Footnotes should be avoided in the text, as they tend to disrupt the flow of the text", we have added a short explanatory sentence directly in the main text. The revisions are as follows (Lines 47-50):

Adequate housing means more than physical shelter; it includes secure tenure, protection from forced eviction, cultural suitability, and access to essential services, schools, and employment.

3. Line 74–76 ("Know Your City Campaigns") – Most readers will not know what this is. Add a short explanation (e.g., "a participatory slum mapping initiative by Slum Dwellers International").

Response: Thank you for this helpful comment. Following your comment, we have added a short explanation of Know Your City Campaigns. The revisions are as follows (Lines 81-82):

National statistics and sparse survey data, such as those from censuses or Know Your City Campaigns, a participatory slum mapping initiative by Slum Dwellers International, fail to provide a spatially continuous view of slum population (Angeles et al., 2009; Pedro and Queiroz, 2019; Persello and Kuffer, 2020).

4. Line 101–102 – "Grid-cell level (6.72 km × 6.72 km at the equator)" – Add a parenthetical example of how many grid cells a typical city (e.g., Nairobi or Mumbai) would contain. This gives a concrete sense of scale.

Response: Thank you for this helpful comment. We have added parenthetical examples of how many grid cells in Nairobi City County. The revisions are as follows (Lines 109):

…we integrate ground-surveys, public satellite imagery, and advanced machine learning to map slum population at the grid-cell level (6.72km × 6.72km at the equator; approximately 16 grid cells for Nairobi City County) across Global South countries.

5. Line 135–139 – "A cluster refers to a DHS enumeration unit, approximately the size of a neighbourhood or village." – Please add approximate population or household range (e.g., "typically 50–300 households") so readers can understand the granularity of the training data.

Response: Thank you for this helpful comment. We have added the approximate household range. The revisions are as follows (Lines 146):

Here, a cluster refers to a DHS enumeration unit, approximately the size of a neighbourhood or village, typically containing 50–300 households.

6. Line 150–152 – Mention of random displacement (2 km urban, 5 km rural) is critical. Please add a short sentence in Section 3.2 or 3.3 explaining how your model handles this uncertainty – for example, "the 6.72 km grid cell size was chosen to absorb this displacement."

Response: Thank you for pointing this out. Following your suggestions, we have added a short explanation in Section 3.3 clarifying that the 6.72 km grid-cell resolution was selected in part to reduce the influence of DHS coordinate displacement. The revisions are as follows (Lines348-352):

Section 3.3

…Since DHS cluster coordinates are randomly displaced for confidentiality purposes, with a maximum displacement of 2 km for urban clusters and 5 km for rural clusters, the approximately 6.72 km grid-cell size helps absorb part of this displacement and reduce the influence of positional uncertainty on model training and prediction.

7. Line 177–181 – The description of the NPP-VIIRS-like nighttime light dataset is technical. Add a plain-English statement: "Brighter nighttime lights generally indicate more developed areas, so this helps the model distinguish slums from better-served neighbourhoods."

Response: Thank you for this helpful comment. Following your suggestions, we have added a plain-language explanation of the NPP-VIIRS-like nighttime light dataset. The revisions are as follows (Lines 196-197):

…Brighter nighttime lights generally indicate more developed areas, so this helps the model distinguish slums from better-served neighborhoods.

8. Line 228–232 (SDG 11.1 and "access to electricity") – Explain why electricity was added while "security of tenure" was excluded, in one simple sentence. Currently it reads as an arbitrary choice.

Response: Thank you for pointing this out. We have clarified the reasons why we included the electricity and excluded the security of tenure. The revisions are as follows (Lines 249-253):

…we incorporated "access to electricity" into our framework because it is a key dimension of infrastructure services, consistent with the broader focus of SDG 11.1 on adequate, safe and affordable housing and basic services. However, we excluded the "security of tenure" dimension due to the lack of internationally comparable data.

9. Line 273–276 – "ResNet-34" and "transfer learning" will be opaque to many. Add a short intuitive explanation: "Like reusing a general-purpose image-recognition system (trained on everyday photos) and adapting it to recognise slum-like patterns in satellite images."

Response: Thank you for pointing this out. Following your suggestion, we have added the intuitive explanation of ResNet-34 and transfer learning. The revisions are as follows (Lines297-299):

…Transfer learning in deep learning leverages pre-trained models, which have been trained on large datasets, to capture general features with high accuracy (Pan and Yang., 2009). Like reusing a general-purpose image-recognition system (trained on everyday photos) and adapting it to recognize slum-like patterns in satellite images.

10.Line 385–388 – "Excluded grid-cells with population below 1,000" – Explain the reason (e.g., "to avoid unreliable predictions in very sparsely populated areas"). This reassures readers that you did not arbitrarily cut off data.

Response: Thank you for pointing this out. We have added a sentence explaining that grid cells with populations below 1,000 were excluded to avoid unreliable predictions in very sparsely populated areas. The revisions are as follows (Lines 411-412):

… *any image slice covering an area with a total population below 1,000 people was excluded to avoid unreliable predictions in very sparsely populated areas.*

11.Line 402–409 – The assumption that slum and non-slum households have the same average size is crucial. The discussion already does a good sensitivity analysis. Please move a short version of that sensitivity (e.g., "assuming ±10% difference changes national estimates by ~8%") into the main results section (Section 4.2 or 4.4), not only in the Discussion. This makes it more visible to readers who skim.

Response: Thank you for this helpful comment. To make the implication of this assumption more visible, we have moved a short summary of the household-size sensitivity analysis into the Section 4.4. The revisions are as follows (Lines 636-642):

Section 4.4

…To quantify the potential influence of the equal-household-size assumption, we conducted a sensitivity analysis using alternative household-size scenarios. Specifically, we assumed that slum households were, on average, either 10% larger or 10% smaller than non-slum households and recalculated the resulting national slum-population estimates. The results show that assuming a ±10% difference in slum household size changes the national slum population estimates by about 8%. Detailed results are shown in Figure S5.

12.Figure 5a and 6a (maps) – The maps use color gradients. For colour-blind or greyscale printing, add a note in the caption that patterns (e.g., hatching) are not used, but the digital version relies on colour. Also, define "dark blue" as "highest absolute slum population" explicitly in the caption.

Response: Thank you for this suggestion. Following your comments, we have revised the captions of Figures 5a and 6a to note that the digital version relies on color gradients and that hatching or other pattern overlays are not used. We have also explicitly clarified that darker blue indicates higher absolute slum population. The revisions are as follows (Lines 549-551, 583-585):

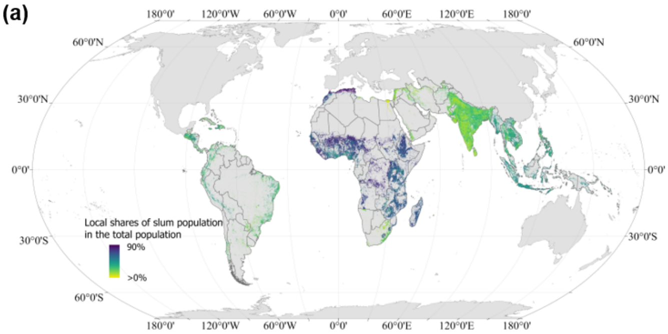
Figure 5 …(a) *Map of slum population with a resolution of 6.72km. The digital version relies on color gradients; hatching or other pattern overlays are not used. Darker blue indicates a higher absolute slum population.*

Figure 6…(a) *Map displaying local shares of the slum population at a resolution of 6.72km. The digital version relies on color gradients; hatching or other pattern overlays are not used. Darker blue indicates a higher local share of the slum population.*

13.Line 532–541 – The text mentions "very low (60%)" concentration levels. Please add these exact thresholds to the legend of Figure 6a, not only in the text.

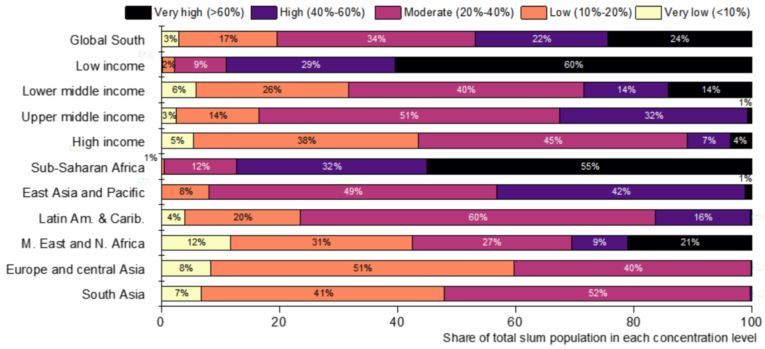
Response: Thank you for your suggestion. We have added the exact threshold ranges to the relevant legend in Figure 6. The revisions are as follows (Lines580):

(a)



Local shares of slum population in the total population
90%
>0%

(b)



Very high (>60%)
High (40%-60%)
Moderate (20%-40%)
Low (10%-20%)
Very low (<10%)
Global South
Low income
Lower middle income
Upper middle income
High income
Sub-Saharan Africa
East Asia and Pacific
Latin Am. & Carib.
M. East and N. Africa
Europe and central Asia
South Asia
Share of total slum population in each concentration level

Notification to the authors:

Regarding figure 7: Google Maps/Earth Imagery should always be exported with the copyright statement visible as embedded in the snippet. If you adapted from Google Maps Imagery the copyright of the underlying map material must still be provided and visible in the map. Furthermore, Google allows to give the attribution (copyright) close to the map, if it was not possible to keep the embedded copyright statement within the map. Then, you need to ensure that the caption uses the identical copyright statement, here "Map data © 2026 Google".

Response: Thank you for pointing this out. Because the embedded copyright statement could not be retained in the exported image, we have added the identical attribution statement, "Map data © 2026 Google", at the end of the figure caption. The revisions are as follows (Lines 598):

Figure 7. 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. Map data © 2026 Google.