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

This manuscript presents a technically sophisticated effort to create a first-of-its-kind, spatially explicit inventory of slum populations for the Global South. The authors' framework, which integrates diverse data sources using a hybrid deep learning approach, is novel and addresses a topic of significant importance for sustainable development and urban studies.

However, the manuscript's claims of robustness currently rest on a methodological foundation that requires significant clarification and justification. Several critical design choices, from the definition of the ground-truth variable to the modeling architecture and data processing steps, are not sufficiently defended. Therefore, it requires a major revision to address these fundamental issues before it can be considered for publication.

Overall Response: Thank you for your thoughtful evaluation of our manuscript and for recognizing the novelty and potential value of our work. In the revised manuscript, we have strengthened the clarification and justification including the definition of the ground-truth variable, modeling architecture and data processing procedure. Please find our point-by-point responses below.

Major Comments

1. Line 241-252: The formula appears to calculate a weighted average of three within-dimension proportions. Why is this structure superior to more standard aggregation methods?

Response: Thank you for this thoughtful comment**.** We agree that the rationale for the weighted average aggregation method required clearer explanation. Our intention was not to claim that this aggregation structure is universally superior to other aggregation methods, but to justify why it is appropriate for constructing a consistent, interpretable, and cross-country comparable measure of multidimensional infrastructure deprivation.

We assigned equal weight to the three dimensions—adequate housing, water and sanitation services, and energy services—because we treat them as equally important

components of slum-related deprivation. Within each dimension, the relevant sub-indicators are averaged before being combined into the overall index. This avoids a mechanical weighting bias that would arise from a simple average across all ten sub-indicators. In our framework, the three dimensions contain different numbers of sub-indicators: two for housing, six for water and sanitation services, and two for energy services. A simple unweighted average across all sub-indicators would therefore assign greater influence to the water and sanitation dimension simply because it includes more indicators, rather than because it is conceptually more important.

We also considered data-driven weighting methods, such as principal component analysis. However, such methods derive weights from the covariance structure of the sample. As a result, the weights may vary across countries, regions, or survey samples, which would reduce the interpretability and cross-country comparability of the resulting index. For this reason, we did not use data-driven weights in the main analysis.

To address the sensitivity of our results to alternative weighting structures, we evaluated two additional weighting schemes in the robustness analysis: the "Basic Service" scenario and the "UN-Habitat" scenario, as described in Section 3.5 and reported in Section 4.4. The manuscript now specifies that the Basic Service scenario assigns greater weight to water and sanitation and energy services, whereas the UN-Habitat scenario gives greater emphasis to housing and water and sanitation services. The results show that, for most countries in Africa, South Asia, and Brazil, the estimated slum population changes by less than ±5% under these alternative schemes, supporting the robustness of our main equal-weight specification. This robustness analysis is reported in Figure 8. The relevant methodological description appears in the revised manuscript around the slum-indicator framework and robustness sections.

We have therefore added the following clarification to Section 3.1 (Lines 239-248):

The indicator calculates a weighted average of the proportion of households lacking services in each dimension. We did not use a simple average across all sub-indicators or data-driven weighting methods such as principal component analysis. A simple average across all sub-indicators would give disproportionate influence to dimensions with more indicators, particularly water and sanitation services, while data-driven weights would depend on the covariance structure of specific samples and could therefore reduce cross-country comparability. We therefore treated adequate housing, water and sanitation services, and energy services as equally important dimensions of multidimensional infrastructure deprivation and assigned each dimension an equal weight of 0.33.

2. Lines 289-315 & 638-642: The manuscript employs a complex, two-stage model (ResNet34 for feature extraction + XGBoost for regression), but its necessity and true performance are not adequately proven. The high R^2 values reported (82-96%) raise concerns about potential overfitting, especially given the presence of spatial autocorrelation in the data. The authors

mention constructing regionally stratified models" in the Discussion (Lines 638-642), which is a crucial methodological detail. Was a formal spatial cross-validation procedure implemented (e.g., holding out entire geographic regions or countries for testing)? If not, the reported performance metrics are likely inflated and do not reflect the model's true generalizability.

Response: Thank you for raising this important concern. We agree that the previous version of the manuscript did not sufficiently justify the two-stage modelling framework or clearly distinguish between within-country predictive performance and out-of-country generalizability. We have therefore revised the manuscript to clarify both the methodological rationale and the validation design, particularly in relation to potential inflation of performance metrics due to spatial autocorrelation.

First, we have clarified that the ResNet-34 + XGBoost framework is not intended simply as a more complex architecture, but as a transfer-learning-based feature extraction and regression pipeline. In the first stage, the fine-tuned ResNet-34 model extracts high-dimensional spatial–spectral features from the seven-band satellite image inputs. In the second stage, XGBoost maps these image-derived embeddings to the DHS-derived slum indicator. This structure is consistent with prior satellite-based socioeconomic prediction studies, where convolutional neural networks are used to extract informative image features and a separate regression model is then used for prediction. We used XGBoost because it can capture nonlinear relationships among the CNN-derived features and the household-based slum indicator, while avoiding reliance on a fully end-to-end deep-learning model trained only on the available survey labels. We have revised the model description to make clear that this is a pragmatic hybrid modelling choice rather than a claim that the architecture is universally superior. The current manuscript describes the ResNet-34 feature extraction and XGBoost regression structure in Section 3.2.

Second, we agree that observation-level cross-validation alone may produce overly optimistic performance estimates when spatial autocorrelation is present. We have therefore revised the validation section to explicitly distinguish two levels of generalizability: within-country spatial prediction and out-of-country prediction. The high R^2 values reported in the main performance table refer to the within-country setting, where the model predicts unsampled grid cells within countries represented in the training data. We now state this scope clearly and avoid interpreting these values as evidence of full cross-country generalizability. In the revised manuscript, the within-country models achieve R^2 values of 0.82–0.95, with a median R^2 of 0.89.

Third, to address the concern about spatially independent validation, we implemented and now report a formal country-group holdout cross-validation procedure within each regional stratum. In this procedure, all observations from the held-out country or countries are excluded from training and validation and are used only for testing. Where sufficient countries are available, approximately 20% of countries are held out in each fold; for strata

with fewer than five countries, leave-one-country-out validation is used. This procedure directly evaluates the model's ability to predict countries not used during model fitting and substantially reduces the risk that nearby or same-country samples inflate the estimated performance. We did not hold out entire regional strata because each regional model is designed to operate within its corresponding geographic–income stratum; instead, country-level holdout within each stratum provides the most relevant test of the model's intended out-of-country application. The revised Section 3.3 now describes this validation design explicitly.

The new out-of-country validation results are reported in Figure 4 and Table S5. Under country-group holdout validation, the models achieve a median R^2 of 0.85 and a median RMSE of 6.89% across holdout folds. Performance is strongest in several African regional models, where held-out-country R^2 values frequently exceed 0.85, but weaker in some Latin American holdout cases, where R^2 values are closer to 0.50. We now report these results separately from the within-country validation metrics and explicitly note that predictive performance varies across regions.

We have also revised the Abstract (Lines 31-35), Section 3.3 (Lines 338-355), Section 4.1 (Lines 463-474), and Conclusions (Lines 929-931) to distinguish within-country prediction from out-of-country generalization. This clarification ensures that the reported performance metrics are interpreted appropriately: the higher R^2 values reflect within-country spatial prediction, whereas the country-group holdout results provide a more stringent assessment of generalizability to countries not used in training.

Abstract

… Our models explain 82–95% of the variation in within-country spatial prediction and 45–91% in country-level holdout validation, corresponding to median R^2 values of 0.89 and 0.85, respectively. These results indicate strong predictive performance and remain broadly comparable with, and in some cases higher than, previously reported benchmarks.

3 Methods

3.3 Model training, cross validation and testing

…

We evaluated model performance under two validation settings: within-country spatial prediction and out-of-country generalization. For within-country prediction, spatiotemporally matched satellite images and DHS-derived slum indicators were split at the observation level using five-fold stratified cross-validation. The folds were constructed to preserve both the urban–rural composition and the country-level representation of the samples within each regional stratum. This validation setting assesses the ability of the models to predict unsampled grid cells in countries that are represented in the training data. To further assess spatial generalizability and reduce the risk of performance inflation due to spatial autocorrelation, we implemented country-group holdout cross-validation within each regional

stratum. In this setting, entire countries, rather than individual observations, were held out for testing. Where a regional stratum contained a sufficient number of countries, approximately 20% of countries were excluded in each fold; for strata with fewer than five countries, we applied leave-one-country-out validation. All observations from the held-out country or countries were excluded from both model training and validation and were used only for final testing. This design provides a stricter evaluation of model transferability to countries not used during model fitting.

4.1 Model performance

… Figure 4 presents the out-of-country predictive performance of the regionally stratified models evaluated using country-group holdout cross-validation. Across holdout folds, R^2 values range from 0.45 to 0.91, with a median of 0.85, while RMSE values range from 4.52% to 12.96%, with a median of 6.89%. Predictive performance is generally strongest for the African regional models, where several held-out-country tests yield R^2 values above 0.85. By contrast, transferability is weaker in several Latin American holdout cases, where R^2 values are closer to 0.50. Full validation results are provided in Table S5. Overall, these results indicate that the modelling framework can generalize to countries not used during model fitting, although predictive accuracy varies across regional contexts. The performance is broadly comparable with, and in some cases higher than, benchmarks reported in previous satellite-based socioeconomic prediction studies.

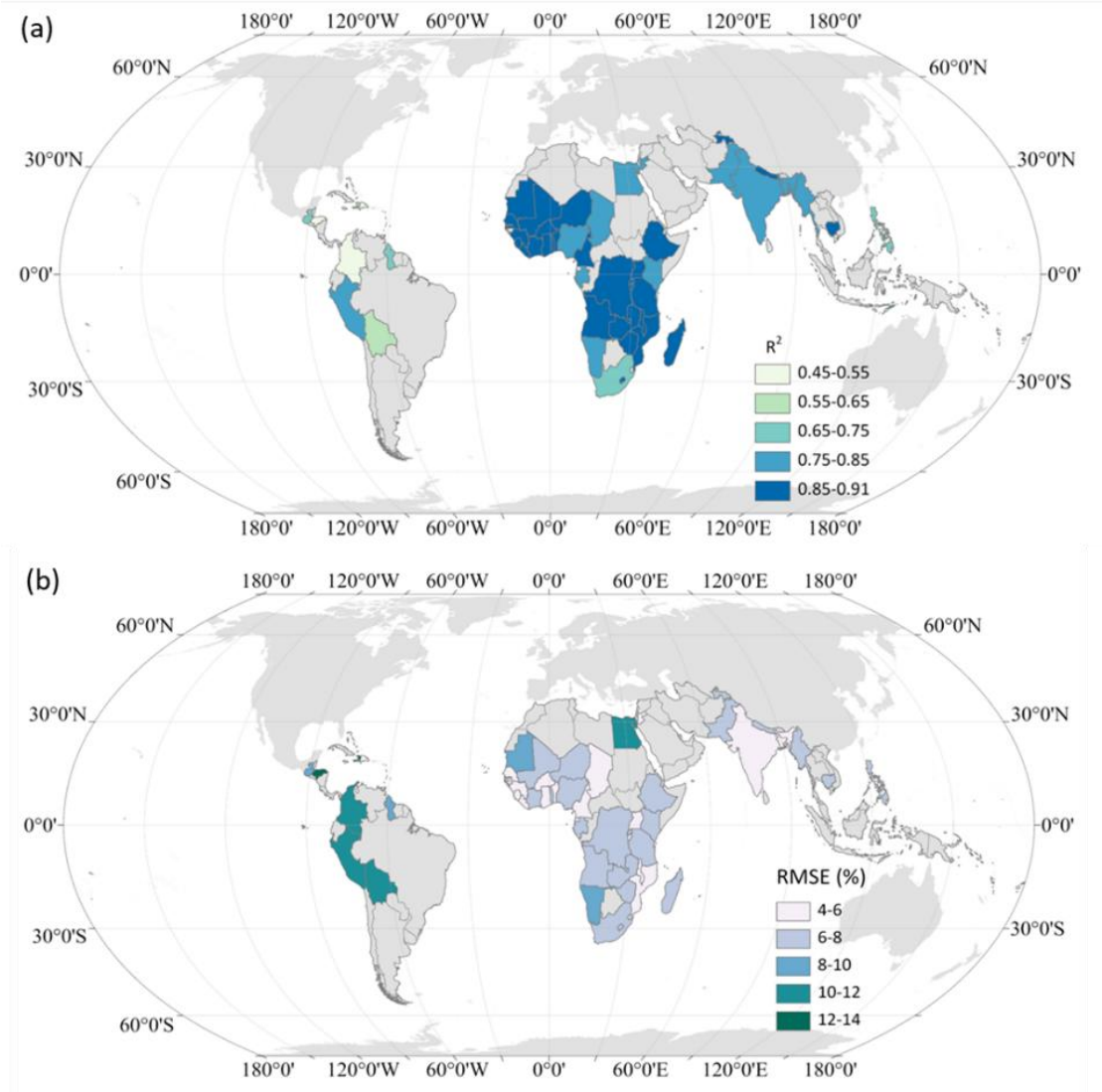

Figure 4. Predictive performance of regionally stratified models under country-group holdout cross-validation. For regions with sufficient country coverage, approximately 20% of countries were held out in each fold; for regions with fewer than five countries, leave-one-country-out cross-validation was used. Performance was measured by (a) the coefficient of determination (R^2) and (b) root mean squared error (RMSE). In each validation fold, all observations from the held-out country or countries were excluded from model training, validation, and hyperparameter tuning, and were used only for testing.

Conclusions

… The models exhibit strong within-country and out-of-country spatial predictive performance, achieving median R^2 values of 0.89 and 0.85, respectively.

Supplementary Information

Table S5 Country-group holdout cross-validation results for assessing out-of-country model generalizability.

Regional model	Fold	Held-out countries	Training countries	No. training samples	No. test samples	RMSE(%)	R^2
West Africa – Low Income	1	BF, LB	CD, GM, ML, NI, SL, TD, TG	3071	821	5.63	0.86
	2	GM, SL	BF, CD, LB, ML, NI, TD, TG	3056	836	5.99	0.89
	3	TD	BF, CD, GM, LB, ML, NI, SL, TG	3268	624	5.96	0.82
	4	CD, ML	BF, GM, LB, NI, SL, TD, TG	3087	805	6.22	0.90
	5	NI, TG	BF, CD, GM, LB, ML, SL, TD	3086	806	6.98	0.88
East Africa – Low Income	1	MW	BU, ET, MD, MZ, RW, UG	3591	850	4.52	0.87
	2	UG	BU, ET, MD, MW, MZ, RW	3757	684	5.01	0.91
	3	MD	BU, ET, MW, MZ, RW, UG	3791	650	6.48	0.88
	4	ET, RW	BU, MD, MW, MZ, UG	3329	1112	7.02	0.90
	5	BU, MZ	ET, MD, MW, RW, UG	3296	1145	5.25	0.91
West Africa – Lower-middle Income	1	NG	AO, BJ, CI, CM, GH, GN, MR, SN	4064	1379	7.39	0.84
	2	MR	AO, BJ, CI, CM, GH, GN, NG, SN	4245	1198	8.64	0.90
	3	AO, CI	BJ, CM, GH, GN, MR, NG, SN	4567	876	7.81	0.89
	4	BJ, GN	AO, CI, CM, GH, MR, NG, SN	4502	941	5.39	0.89
	5	CM, GH, SN	AO, BJ, CI, GN, MR, NG	4394	1049	5.86	0.91
East Africa – Lower-middle Income	1	EG	KE, KM, LS, TZ, ZM, ZW	3643	1817	11.30	0.85
	2	KE	EG, KM, LS, TZ, ZM, ZW	3875	1585	7.75	0.85
	3	TZ	EG, KE, KM, LS, ZM, ZW	4853	607	6.31	0.88
	4	KM, ZM	EG, KE, LS, TZ, ZW	4808	652	6.12	0.90
	5	LS, ZW	EG, KE, KM, TZ, ZM	4661	799	6.80	0.91
Africa – Upper-middle Income	1	ZA	GA, NM	720	746	7.68	0.69
	2	NM	GA, ZA	916	550	9.79	0.82
	3	GA	NM, ZA	1296	170	7.61	0.76
Asia – Lower West	1	IA	BD, JO, NP, PK, TJ	2950	19554	5.76	0.80
	2	JO	BD, IA, NP, PK, TJ	21534	970	5.72	0.81
	3	BD	IA, JO, NP, PK, TJ	21832	672	5.47	0.83
	4	PK	BD, IA, JO, NP, TJ	21944	560	7.94	0.83
	5	NP, TJ	BD, IA, JO, PK	21756	748	6.10	0.86
Asia – Lower-middle East	1	PH	KH, MM, TL	1507	1198	6.08	0.75
	2	KH	MM, PH, TL	2094	611	7.10	0.88
	3	TL	KH, MM, PH	2250	455	6.60	0.81
	4	MM	KH, PH, TL	2264	441	7.75	0.76
Latin America – High/Upper-middle Income	1	CO	DR, GU, GY, PE	2761	4408	12.03	0.52
	2	PE	CO, DR, GU, GY	6088	1081	10.18	0.76
	3	GU	CO, DR, GY, PE	6316	853	8.21	0.71
	4	DR	CO, GU, GY, PE	6646	523	5.41	0.45
	5	GY	CO, DR, GU, PE	6865	304	9.35	0.73
Latin America – Lower/Low Income	1	HN	BO, HT	1442	1093	12.78	0.53
	2	BO	HN, HT	1543	992	11.54	0.58
	3	HT	BO, HN	2085	450	14.96	0.64

3. Lines 200-201: The manuscript fails to state how the 1km GHS-POP data was aggregated to the 6.72km analysis grid. For population counts, summation is the only method that preserves demographic totals.

Response: Thank you for pointing out this important issue. We agree that GHS-POP population counts are extensive variables and should therefore be aggregated by summation rather than resampled using interpolation. The previous version of the manuscript did not describe this step sufficiently clearly, and the wording could have implied that interpolation-based resampling was applied to population counts. Following this suggestion, we have revised the population aggregation procedure, clarified the corresponding description in the Methods section, and updated the affected results accordingly. The specific revisions are as follows (Lines 392-400, 505-516, 619-629):

***3 Methods** (section 3.4):*

… All auxiliary datasets were harmonized to the 6.72 km analysis grid defined by the predicted slum-indicator map. The predicted slum-indicator grid was used as the target grid for population aggregation. To preserve demographic totals, the GHS-POP raster was aggregated by summing the 1 km population counts within each 6.72 km target grid cell. No interpolation-based resampling was applied to the GHS-POP population counts. Categorical datasets, including GHS-SMOD and CGLS-LC100, were resampled using nearest-neighbor interpolation to retain categorical integrity. All datasets were projected to WGS 1984 for spatial consistency.

4 Results:

4.2 Spatially explicit estimates of slum population in Global South

Figure 5a shows the spatial distribution of slum population across Global South countries. In terms of absolute numbers, slum population is concentrated in areas with expected high population densities, such as northern India in Asia, Rwanda and northern Morocco in Africa, and the coastal regions of Rio De Janeiro in Latin America, highlighted in dark blue on the map. Notably, more than 50% of the total slum population resides in less than 8% of the total grid cells.

Figure 5b highlights the administrative statistics and geographic disparities in slum population distribution. At the regional level, we estimate a total of over 0.80 billion people living in slums across the Global South. More than 60% of this population resides in South Asia and Sub-Saharan Africa, followed by East Asia and the Pacific (16.3%), the Middle East and North Africa (8.6%), Latin America and the Caribbean (7.8%), and Europe and Central Asia (0.15%).

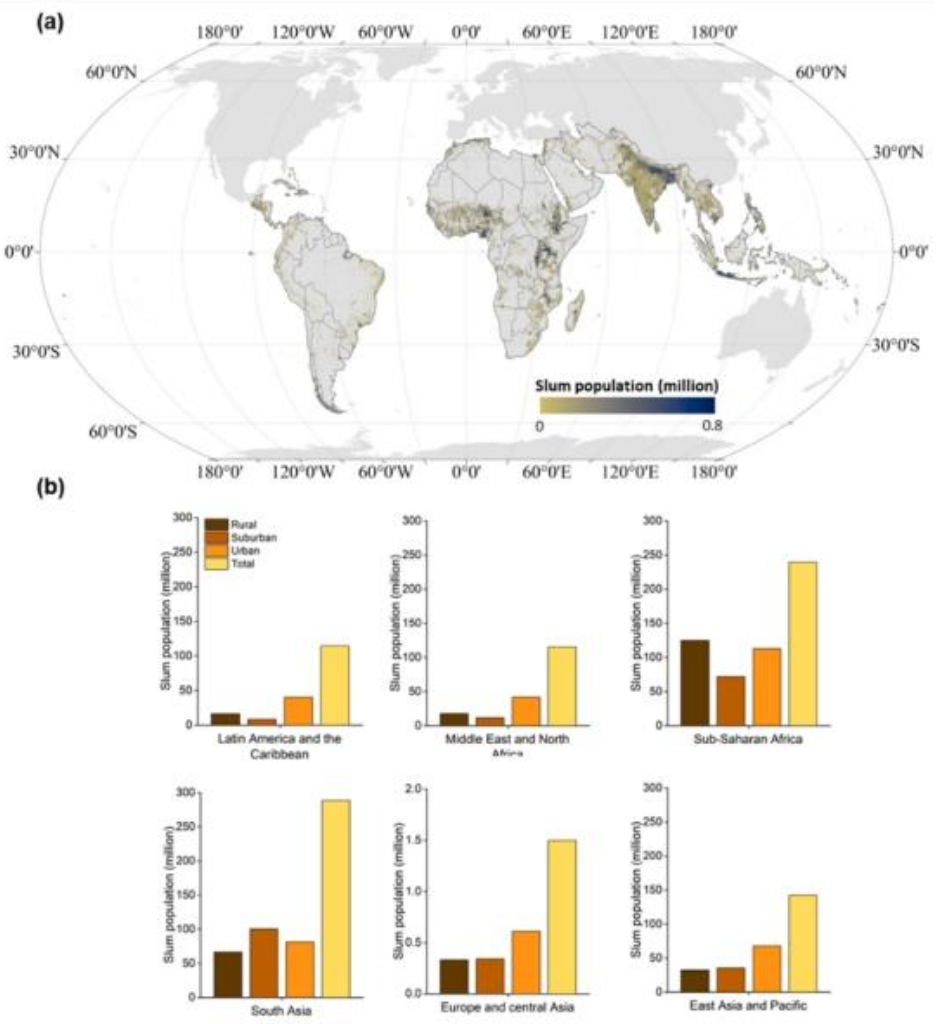

Figure 5 Spatial distribution of slum population in the Global South and their geographic disparities. (a) Map of slum population with a resolution of 6.72km. The grid cells with a population of less than 1,000 are not included. (b) Statistics of slum population across geographic regions and disparities between rural and urban areas.

4.5 Comparison of results with statistical data

We compare our slum population estimates with previous estimates from the literature, which are derived by scaling population counts at individual slum level as well as from city-scale slum population statistics, as summarized in Table 6 (Butera et al., 2019; Byuro, 2015; Davis, 2013; Guilmoto and Rajan, 2013; Medeiros et al., 2012; Patel et al., 2019; Sabry, 2009; Taubenböck and Wurm, 2015). Figure 9a shows our estimates, indicated by an asterisk (*), fall within the range of these scaled slum population estimates. Moreover, when compared to city-scale slum population statistics, our results show strong alignment, particularly in Hyderabad, Johannesburg and São Paulo (Trindade et al., 2021) (Figure 9b).

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	**0.90**	0.92, 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	**1.99**	1.3, 1.41, 3.02, 5.67, 8.62	2.16	Breuer et al. (2024); Trindade et al. (2021)
Cape Town	**0.77**	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	**0.55**	0.01, 0.3, 0.71, 3.75, 4.55, 7.47	/	Breuer et al. (2024); Sabry (2009);
Mumbai	**1.60**	1.65, 1.68, 5.2	/	Breuer et al. (2024); Taubenböck and Wurm (2015)
Dhaka	**1.83**	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	**0.70**	/	0.6	Trindade et al. (2021)
Hyderabad	**2.00**	/	2.3	Trindade et al. (2021)

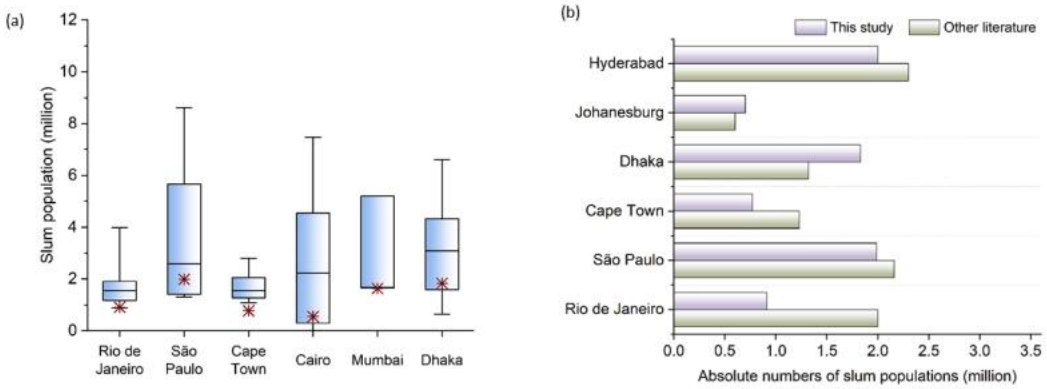

Figure 9 Comparison of our slum population estimates with data from previous literature and official statistics, presented as both absolute numbers and relative shares. (a) Comparison with scaled-up values from previous studies for six cities: Rio de Janeiro (n=9), São Paulo (n=5), Cape Town (n=12), 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.

Minor Comments

- Lines 638-642: Key details about the modeling strategy (e.g., the use of regionally stratified models) are currently in the Discussion section. All descriptions of what was done should be consolidated within the Methods section to ensure a logical flow.

Response: We have moved the key methodological details regarding the regionally stratified models from the Discussion section to the Methods section (Lines 327-331).

- Section 2.1 vs 2.2: The Methods section inconsistently switches between present tense (This study uses...) and past tense (We obtained...). The established convention is to use the past tense to describe the work performed. This should be corrected for professionalism.

Response: We have revised Methods sections to ensure consistent use of the past tense.

- Lines 46 & 835: The data is repeatedly cited as being in the Zalando repository. The correct name is Zenodo.

Response: Thank you for pointing out this error. We have corrected "Zalando" to "Zenodo" throughout the manuscript.

- Lines 767-831: The Implications and future research section is good but could be more impactful by proposing 1-2 concrete, actionable research questions that this new dataset uniquely enables (e.g., How does the spatial distribution of slum populations correlate with exposure to specific climate hazards at a regional scale?)

Response: We have extended the Discussion to provide two actionable research pathways directly supported by this product. The revisions are as follows (Lines 842-852):

5.3 Implications and future research

… At this resolution, the dataset is best suited to identifying broad concentrations of slum populations for subsequent research and policy analysis. For example, one actionable research pathway is to integrate the gridded slum population estimates with commensurate-scale climate and environmental hazard maps, such as floods, heat exposure, drought, or air pollution, to assess compound risks among populations living in slum-like conditions. Another pathway is to use the dataset as a spatially explicit vulnerable-population baseline to evaluate whether public investment, health services, and climate adaptation efforts are equitably aligned with the distribution of slum populations. Together, these applications demonstrate the value of the dataset for assessing compound vulnerability and resource equity among slum populations.

Reviewer's Comments:

Anonymous Referee #4

Summary

This manuscript presents a highly significant and methodologically rigorous study. The authors have successfully developed a novel, scalable framework that integrates household surveys, multi-spectral satellite imagery, and gridded population data to produce the first spatially explicit, gridded inventory of slum populations across 129 Global South countries. The hybrid machine learning approach (ResNet-34 + XGBoost) is well-justified and state-of-the-art. The manuscript is well-structured, the validation is comprehensive, and the work addresses a critical data gap. It represents a substantial contribution to urban geography, sustainable development, and applied remote sensing. I recommend publication after the authors address the following points to further strengthen an already excellent manuscript.

Overall Response: Thank you for recognizing the contribution of our work and for your constructive comments, which have helped us further strengthen the manuscript. Below, we provide our point-by-point responses to each comment.

Major Comments

1. **Lines 716-739 (Spatial Resolution & Policy Relevance):** The technical rationale for the 6.72 km resolution is clear. However, the discussion would benefit from a more explicit analysis of its policy and practical implications. Please briefly elaborate on what scale of planning or analysis (e.g., national/regional resource allocation, identifying broad vulnerability hotspots) this product is optimally designed for, and contrast this with applications for which it is not suitable (e.g., intra city infrastructure planning). This will greatly aid end-users in correctly applying your dataset.

Response: Thank you for this valuable suggestion. We agree that, beyond explaining the technical basis of the 6.72 km resolution, the manuscript should more clearly specify the appropriate policy and practical uses of the dataset. We have therefore revised the Discussion section to clarify the spatial scale at which the product is most useful, as well as the types of applications for which it is not intended. Specifically, we now emphasize that the dataset is designed primarily for national- and regional-scale analysis, including the identification of broad vulnerability hotspots, cross-country and cross-regional comparisons, and strategic prioritization of resources for basic service provision, poverty reduction, public health, and climate adaptation. At the same time, we clarify that because the product is based on cluster-level slum proxy indicators and has a spatial resolution of 6.72 km, it cannot capture the fine-grained morphology, boundaries, or internal heterogeneity of individual slum settlements. It is therefore not suitable for site-specific operational applications, such as intra-city

infrastructure planning, settlement-boundary delineation, or building-level slum upgrading. The revised text in the Discussion section (on spatial resolution and dataset) as follows (Lines 782-797):

… Our method relies on cluster-level slum proxy indicators rather than precise morphological boundaries, which shapes the downstream policy and practical implications of the dataset. The resulting product provides estimates of slum populations at a spatial resolution of 6.72 km under current data constraints, and is therefore primarily intended for national- and regional-scale analyses. Potential applications include identifying broad vulnerability hotspots where slum-like living conditions coincide with climate-related hazards, environmental risks, or public health concerns, as well as supporting the strategic prioritization of resources for basic service provision, poverty reduction, and climate adaptation. However, because this spatial resolution cannot capture the fine-grained morphology, precise boundaries, or internal heterogeneity of slum settlements, the dataset is not intended for site-specific operational applications, such as intra-city infrastructure planning, settlement-level boundary delineation, or building-level slum upgrading. Future research could address this limitation by developing higher-resolution slum datasets and more refined approaches for delineating settlement boundaries.

2. **Lines 365-370 & 192-201 (Population Data Processing - Critical Clarification):** In Section 3.4, you mention resampling the GHS-POP data using "cubic convolution." For population **counts** (as opposed to density), the standard and conceptually correct method for aggregation to a coarser resolution is **summation**, which preserves the total population. Please explicitly clarify: o Was the cubic convolution applied to a population density layer? o If it was applied to counts, please justify this methodological choice in Sections 2.3 or 3.4 and discuss its potential impact on the final slum population totals.

Response: Thank you for raising this important methodological issue. We apologize that the previous version of the manuscript did not describe the GHS-POP processing procedure clearly enough. We agree that, because GHS-POP represents population counts rather than a continuous density surface, cubic convolution or other interpolation-based resampling methods are not appropriate for aggregation, as they do not preserve demographic totals. Following the reviewer's and editor's suggestions, we revised the GHS-POP processing procedure. Specifically, we now aggregate the 1 km GHS-POP population counts to the 6.72 km prediction grid by summation, using the predicted slum-indicator grid as the target grid. No interpolation-based resampling is applied to the GHS-POP population counts. This ensures that population totals are preserved when generating the grid-cell-level slum population estimates. We have clarified this procedure in the Methods section and updated the related results accordingly. The specific revisions are as follows (Lines 392-400, 505-516, 619-629):

***3 Methods** (section 3.4):*

… All auxiliary datasets were harmonized to the 6.72 km analysis grid defined by the predicted slum-indicator map. The predicted slum-indicator grid was used as the target grid for population aggregation. To preserve demographic totals, the GHS-POP raster was aggregated by summing the 1 km population counts within each 6.72 km target grid cell. No interpolation-based resampling was applied to the GHS-POP population counts. Categorical datasets, including GHS-SMOD and CGLS-LC100, were resampled using nearest-neighbor interpolation to retain categorical integrity. All datasets were projected to WGS 1984 for spatial consistency.

4. Results:

4.2 Spatially explicit estimates of slum population in Global South

Figure 5a shows the spatial distribution of slum population across Global South countries. In terms of absolute numbers, slum population is concentrated in areas with expected high population densities, such as northern India in Asia, Rwanda and northern Morocco in Africa, and the coastal regions of Rio De Janeiro in Latin America, highlighted in dark blue on the map. Notably, more than 50% of the total slum population resides in less than 8% of the total grid cells.

Figure 5b highlights the administrative statistics and geographic disparities in slum population distribution. At the regional level, we estimate a total of over 0.80 billion people living in slums across the Global South. More than 60% of this population resides in South Asia and Sub-Saharan Africa, followed by East Asia and the Pacific (16.3%), the Middle East and North Africa (8.6%), Latin America and the Caribbean (7.8%), and Europe and Central Asia (0.15%).

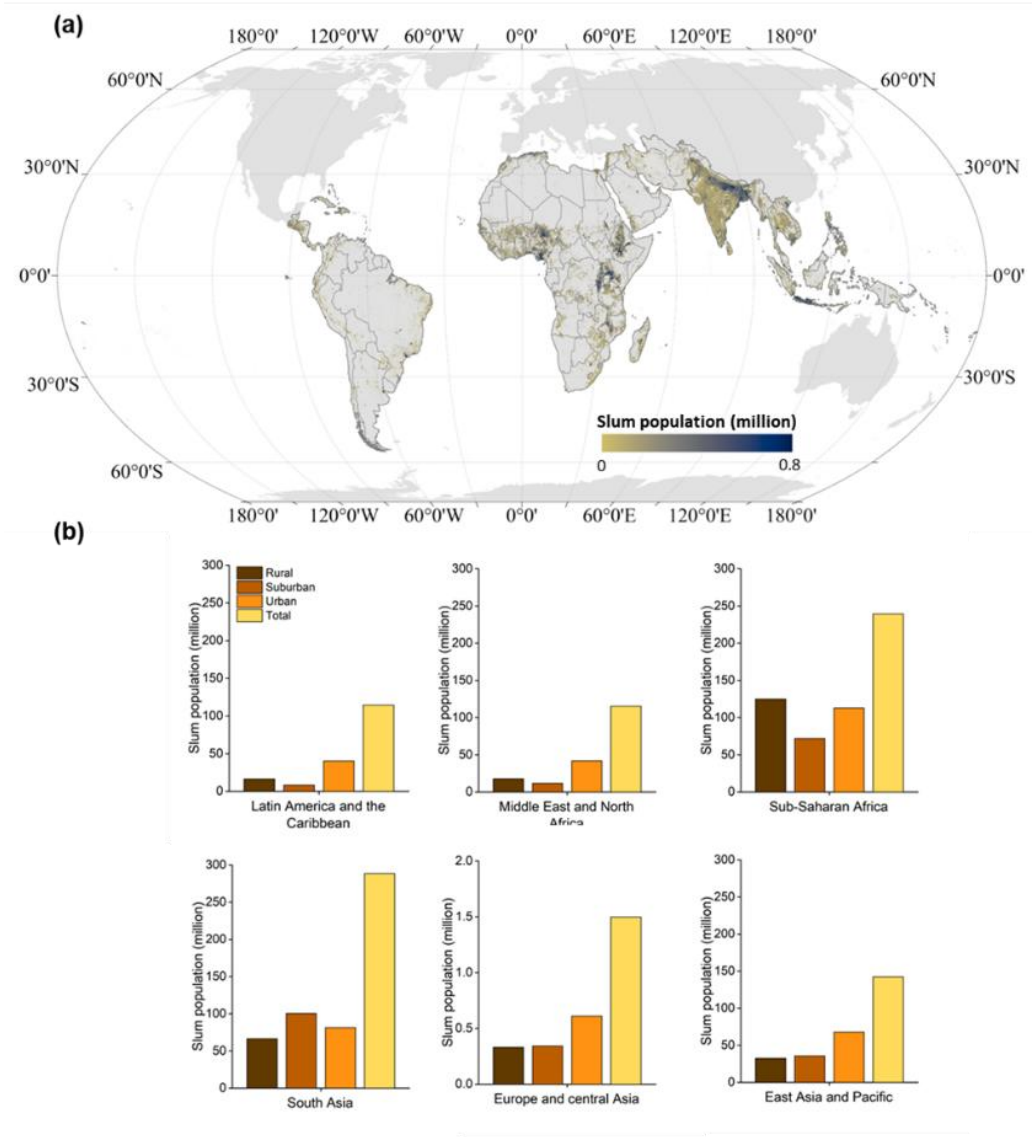

Figure 5 Spatial distribution of slum population in the Global South and their geographic disparities. (a) Map of slum population with a resolution of 6.72km. The grid cells with a population of less than 1,000 are not included. (b) Statistics of slum population across geographic regions and disparities between rural and urban areas.

4.5 Comparison of results with statistical data

We compare our slum population estimates with previous estimates from the literature, which are derived by scaling population counts at individual slum level as well as from city-scale slum population statistics, as summarized in Table 6 (Butera et al., 2019; Byuro, 2015; Davis, 2013; Guilmoto and Rajan, 2013; Medeiros et al., 2012; Patel et al., 2019; Sabry, 2009; Taubenböck and Wurm, 2015). Figure 9a shows our estimates, indicated by an asterisk (*), fall within the range of these scaled slum population estimates. Moreover, when compared to city-scale slum population statistics, our results show strong alignment, particularly in Hyderabad and Johannesburg (Trindade et al., 2021) (Figure 9b).

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	**0.90**	0.92, 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	**1.99**	1.3, 1.41, 3.02, 5.67, 8.62	2.16	Breuer et al. (2024); Trindade et al. (2021)
Cape Town	**0.77**	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	**0.55**	0.01, 0.3, 0.71, 3.75, 4.55, 7.47	/	Breuer et al. (2024); Sabry (2009);
Mumbai	**1.60**	1.65, 1.68, 5.2	/	Breuer et al. (2024); Taubenböck and Wurm (2015)
Dhaka	**1.83**	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	**0.70**	/	0.6	Trindade et al. (2021)
Hyderabad	**2.00**	/	2.3	Trindade et al. (2021)

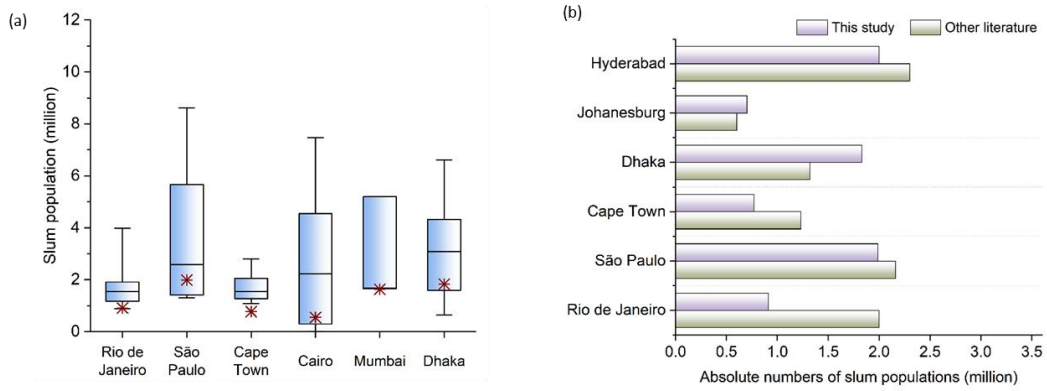

Figure 9 Comparison of our slum population estimates with data from previous literature and official statistics, presented as both absolute numbers and relative shares. (a) Comparison with scaled-up values from previous studies for six cities: Rio de Janeiro (n=9), São Paulo (n=5), Cape Town (n=12), 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.

3. **Lines 807-817 (Quantifying the "Equal Household Size" Assumption)**: You correctly identify this as a key limitation and provide helpful data from India. To strengthen this discussion, I strongly recommend adding **a simple sensitivity or bounding analysis** (e.g., in the

Supplementary Materials). Providing a quantitative estimate (e.g., "If slum households were, on average, 10% larger, our national estimate for Country X would change by approximately Y%") would powerfully contextualize the potential magnitude of bias introduced by this assumption.

Response: Thank you for your constructive comments on the "Equal Household Size" Assumption. We agree that the potential bias introduced by this assumption should be quantified more explicitly. In response, we have added a sensitivity analysis in the Supplementary Materials to evaluate how the estimated slum populations would change if slum households were, on average, larger or smaller than non-slum households, including the adjustment formula and report quantitative results. The specific revisions are as follows (Lines 868-874, 881-890):

5. Discussion

In this study, we assume that the proportion of slum households within each grid cell is equivalent to the proportion of the population living in slums or slum-like conditions. This assumption implies that slum and non-slum households have the same average household size within each 6.72 km grid cell. Although this is a practical approximation given the absence of spatially explicit household-size data at the grid-cell scale, it may introduce uncertainty where household sizes differ systematically between slum and non-slum populations. …

To quantify the potential influence of this 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 (Figure S5). For India, the results show that assuming slum households to be 10% larger would increase the national estimate by approximately 7.82%, whereas assuming them to be 10% smaller would decrease the estimate by approximately 8.14%. These results suggest that household-size differences can affect absolute population estimates, while also providing a quantitative bound on the uncertainty associated with this assumption.

Supplementary Information

Sensitivity analysis for the equal household-size assumption

To quantify the potential influence of the equal household-size assumption, we conducted a simple sensitivity analysis. Let *N* denote the total population of a country, *H* the total number of households, and *p* the baseline estimated slum proxy under the equal household-size assumption. Under this baseline assumption, the share of slum households is treated as equivalent to the share of the population living in slum-like conditions.

To relax this assumption, we allowed the mean household size of slum households to differ from that of non-slum households. Let denote the assumed mean household size of slum households and $\bar{h}_n$ denote the assumed mean household size of non-slum households. The

number of slum households is *pH*, while the number of non-slum households is (1-*p*) *H*. The implied slum population is therefore $pH\bar{h}_s$, and the implied non-slum population is $(1-p)H\bar{h}_n$. The adjusted slum population share can thus be expressed as:

$$p_{adj} = \frac{pH\bar{h}_s}{pH\bar{h}_s + (1-p)H\bar{h}_n} \quad (1)$$

Equivalently, defining $r = \bar{h}_s/\bar{h}_n$as the relative mean household-size ratio between slum and non-slum households, the adjusted slum population share can be written as:

$$p_{adj} = \frac{pr}{pr + (1-p)} \quad (2)$$

The percentage change relative to the baseline estimate under equal household-size assumption is:

$$\Delta(\%) = (\frac{p_{adj}}{p} - 1) \times 100 \quad (3)$$

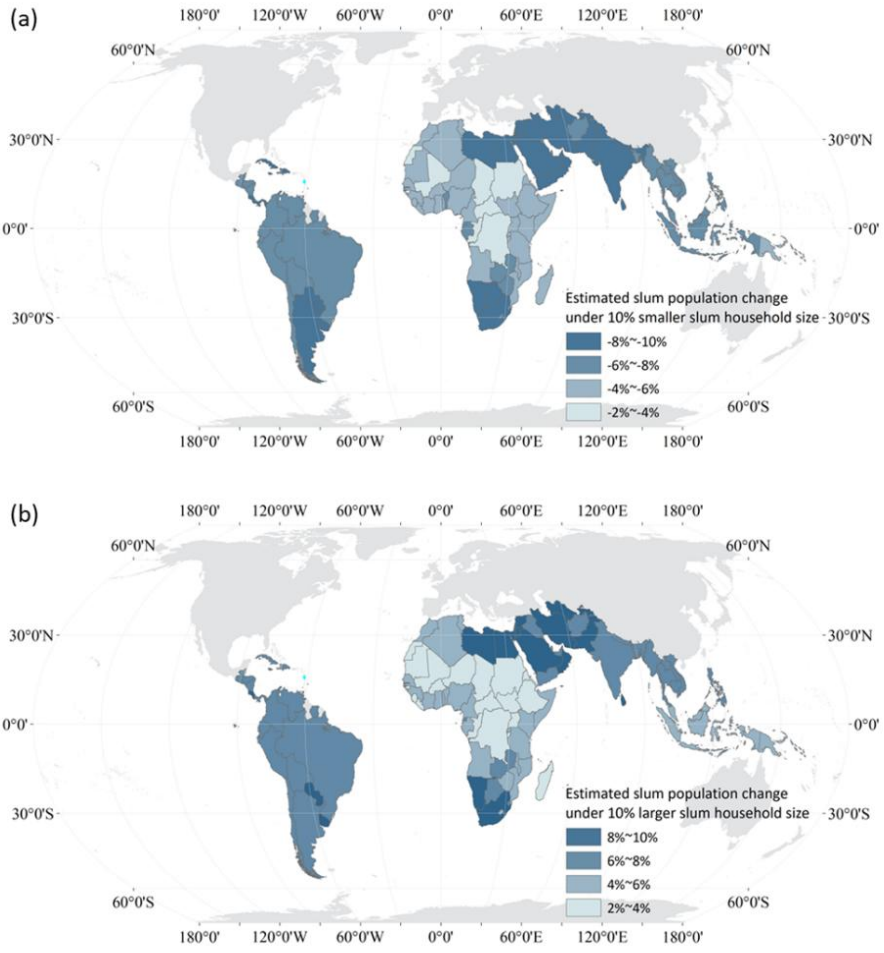

Figure S5. Sensitivity of estimated slum populations to a 10% larger or smaller slum household-size assumption. Percentage changes are calculated relative to the baseline estimates, which assume equal mean household size between slum and non-slum households. The sensitivity scenarios assume that slum households are, on average, either 10% larger or 10% smaller than non-slum households.

4. **Lines 648-651 (Characterizing Uncertainty for Data-Scarce Regions)**: Your note on countries where estimates may be less reliable is important. To enhance transparency and guide future work, please consider: o **Listing or mapping these specific countries** in the Supplementary Materials. o Briefly suggesting how future model iterations might reduce this uncertainty (e.g., incorporating spatial context or other proxy data).

Response: Thank you for your constructive comments regarding the Uncertainty for Data-Scarce Regions. Following your suggestion, we have added a Supplementary Table listing the countries where estimates may be less reliable and should therefore be interpreted with greater caution. We have also revised the manuscript to clarify how future model iterations could reduce this uncertainty. The revisions are as follows (Lines 697-702):

Supplementary Information

Table S7. Countries and areas with changes exceeding ±10% in estimated slum populations under alternative indicator frameworks

Number	Country or area	Number	Country or area
1	Antigua and Barbuda	16	Jamaica
2	Argentina	17	Lao People's Democratic Republic
3	Bahamas	18	Malaysia
4	Brunei Darussalam	19	Occupied Palestinian Territory
5	Chile	20	Oman
6	Cuba	21	Panama
7	Djibouti	22	Peru
8	Dominica	23	Qatar
9	Dominican Republic	24	Saint Kitts and Nevis
10	Egypt	25	Saint Lucia
11	El Salvador	26	Seychelles
12	Equatorial Guinea	27	Thailand
13	Eritrea	28	Vietnam
14	Grenada	29	Western Sahara
15	Honduras		

5. Discussion

5.1 Robustness and external validation

…It is important to note that countries or areas with changes exceeding ±10% are generally those where nationwide household-based surveys are unavailable or limited, or where the

geographic extent is relatively small (Table S7). Future model iterations could reduce this uncertainty by incorporating additional spatial-contextual covariates and proxy data, such as settlement morphology, built-up density, and other socioeconomic indicators.

5. **Lines 767-831 & 845-863 (Connecting Results to Actionable Research & Monitoring):** The implications are well-discussed. Please consider extending this slightly to propose 1-2 concrete, actionable research pathways directly enabled by this new dataset (e.g., integration with commensurate-scale climate hazard maps for compound risk assessment). Furthermore, in the Conclusion, you could more forcefully frame this work not just as a static map, but as the foundation for a future monitoring system, highlighting its readiness for updates with new DHS and satellite data (e.g., Landsat Next, Sentinel-2).

Response: Thank you for your constructive comments. Following your suggestions, we have extended the Discussion to provide two actionable research pathways directly supported by this product, and also strength the Conclusion. The revisions are as follows (Lines 842-852, 928-931):

5.3 Implications and future research

… At this resolution, the dataset is best suited to identifying broad concentrations of slum populations for subsequent research and policy analysis. For example, one actionable research pathway is to integrate the gridded slum population estimates with commensurate-scale climate and environmental hazard maps, such as floods, heat exposure, drought, or air pollution, to assess compound risks among populations living in slum-like conditions. Another pathway is to use the dataset as a spatially explicit vulnerable-population baseline to evaluate whether public investment, health services, and climate adaptation efforts are equitably aligned with the distribution of slum populations. Together, these applications demonstrate the value of the dataset for assessing compound vulnerability and resource equity among slum populations.

7 Conclusions

…More than just a static map, our study provides a scalable foundation for a future monitoring system of slum populations. The framework can be readily updated with new DHS survey rounds and satellite data (e.g., forthcoming Landsat Next and Sentinel-2).

Minor Comments

• **Formatting Consistency:** Please ensure consistent formatting of references and acronyms throughout (e.g., "UN-Habitat" vs. "UN-Habitat" in Lines 49, 225).

Response: Thank you for pointing this out. We have carefully checked the manuscript and revised the formatting of references, and acronyms for consistency throughout the text using

"UN-Habitat".

• **Lines 233-245 (Clarifying the Slum Indicator Formula):** Adding a one sentence intuitive explanation of the formula in the text (e.g., "The indicator calculates a weighted average of the proportion of households lacking services in each dimension") would improve readability for a broader audience.

Response: Thank you for this helpful comment. Following your suggestion, we have added this sentence to make it clear (Lines 239-240).

• **Typographical Errors:** A final proofread is recommended to catch minor errors (e.g., "grided" should be "gridded" in Lines 125, 850).

Response: Thank you for pointing out this type of errors**.** We have conducted a proofread of the manuscript and corrected typographical errors, including changing "grided" to "gridded".