Response to Reviewers' Comments

Paper #: essd-2026-381

Title: A global urban built-up area dataset for cities with populations exceeding 300,000 (2000–2025)

Journal: Earth System Science Data

Authors: Dinoo Gunasekera, Zhongchang Sun, Wenjie Du, Jian Gao, Yunpeng Zhang, Yuyu Zhou, Robert Ndugwa, Huadong Guo

Correspondence: Zhongchang Sun (sunzc@aircas.ac.cn), Wenjie Du

Dear Reviewer,

We thank the reviewer for the careful, critical and constructive assessment of our manuscript. The comments correctly identify the central methodological risks of a multi-epoch, multi-sensor dataset — the 2015→2020 sensor and feature transition, the potential circularity of population constraints, the monotonicity imposed by isotonic regression, the aggregation of 10 m impervious surface area (ISA) to 30 m, and the interpretation of the GUBAD–WUP comparison — and they prompted us to replace several qualitative arguments with controlled, quantified experiments. The main revisions are:

- A controlled sensor-switching experiment on pilot cities across six continents, in which the 2015, 2020 and 2025 epochs were processed under both a Landsat-only route and the production switching route with identical downstream processing (Sections 4.3 and 5.3 revised).
- A quantification of the effect of pixel-level Pool-Adjacent-Violators Algorithm (PAVA) isotonic regression on classifications, urban built-up area (UBA) and land consumption rate (LCR) under three weighting schemes (Sections 3.1 and 5.3 revised).
- A branch-by-branch audit of the ISA-to-UBA production code establishing the actual role of population constraints in the released footprints (Section 3.2 corrected; Sections 5.1 and 5.3 expanded).
- A sub-pixel aggregation-threshold sensitivity experiment at thresholds of 0.25, 0.50 and 0.75 (Sections 3.1, 3.2 and 5.3 revised).
- A decomposition of the GUBAD–WUP comparison behind Figure 8 into a systematic definitional offset, its scaling with city size, and the driver of the declining R^2 (Section 4.2.2 expanded).
- Removal of comparative-superiority language throughout and reframing of the contribution as complementarity rather than competition with existing products (Sections 4.2.3, 5.2 and Conclusions revised).

Point-by-point responses follow. Reviewer comments are quoted verbatim in italics; our responses follow each comment; the corresponding changes in the revised manuscript are listed at the end of each response. Blue marks Reviewer 1 amendments. Purple marks Reviewer 2 amendments.

Reviewer #1

We are grateful to Reviewer #1 for the detailed and technically precise comments. Addressing them has substantially improved the methodological transparency and the honesty of the uncertainty statements in the manuscript.

Comment RC1.1 — Sensor transition and the 2020–2025 LCRPGR increase

> *"The feature stack shifts fundamentally between the 2000–2015 epochs (Landsat-only, spectral indices) and the 2020–2025 epochs (Sentinel-2 MSI at 10 m, Sentinel-1 SAR backscatter, and VIIRS nighttime lights). Adding radar backscatter and sub-pixel optical resolution substantially increases sensitivity to fragmented, peri-urban built fabric and structural edges. The authors observe a dramatic spike in land consumption rate and population growth rate (LCRPGR) during 2020–2025 and dismiss this as real-world divergence driven by post-2020 infrastructure investment. In all likelihood, this is an artifact of sensor transition and feature expansion. Applying the Pool-Adjacent-Violators Algorithm only enforces monotonicity; it does not correct baseline step-changes introduced by adding high-sensitivity sensors in 2020. A controlled sensitivity study holding sensor inputs constant (e.g., Landsat-only indices across all years) is necessary to prove this jump is physical rather than methodological."*

Response

We agree with the reviewer on two points without reservation: the 2015→2020 sensor and feature transition is the single most important threat to the temporal consistency of GUBAD, and PAVA enforces monotonicity but cannot correct a step-change in the detection baseline. We also agree that the original wording attributed the 2020–2025 LCRPGR increase to real-world drivers more confidently than our evidence justified. We assessed sensitivity to the sensor and feature transition through two complementary analyses covering 85 distinct cities. The multi-epoch comparison of urban built-up area (UBA) and land consumption rate (LCR) included 65 cities, and the 2020 map-level comparison of impervious surface area (ISA) included 25 cities. Five cities were included in both analyses.

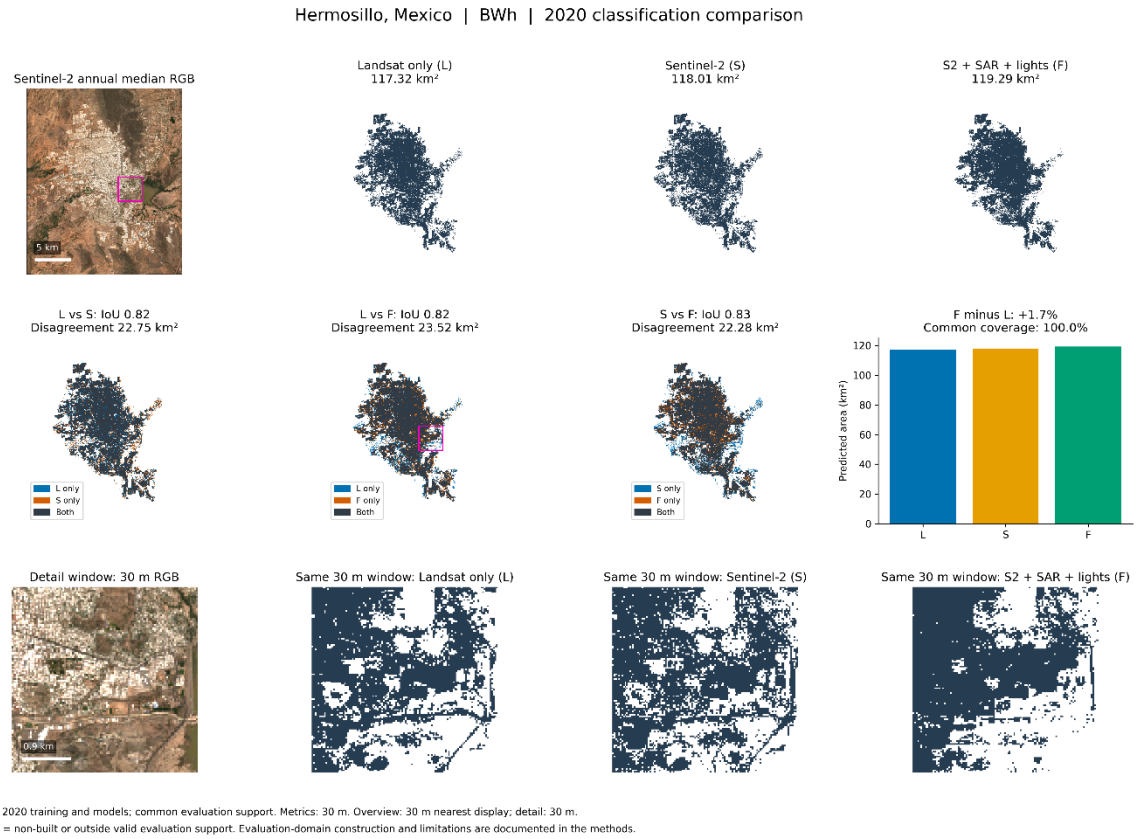
(a) Controlled switching-route experiment. For 65 cities spanning six continents (20 in Asia, 12 in North America, 10 in Europe, 9 in Africa, 8 in South America and 6 in Oceania), we processed the 2015, 2020 and 2025 epochs twice: (i) a Landsat-only route, in which the Landsat-style feature stack is used at all three epochs; and (ii) the production switching route, in which 2015 uses Landsat and 2020–2025 use Sentinel-2 MSI, Sentinel-1 SAR, and VIIRS nighttime lights. Crucially, both routes share identical downstream processing (30 m aggregation, morphological refinement, metropolitan-context screening, population-constrained masking and UBA nesting) and identical population inputs, so the comparison isolates the sensor/feature configuration. The result is not a simple upward step. Relative to the Landsat-only route, the switching route produces a higher land consumption rate (LCR) in 2015–2020 in 64 of the 65 cities (range −0.037 to +14.279 percentage points per year; median +1.805), but a lower LCR in 2020–2025 in 55 of the 65 cities (range −10.145 to +1.175 percentage points per year; median −0.785). At the area level, the median 2020 UBA is 8.72% larger under the switching route (larger in 52 of 65 cities) and the median 2025 UBA is 4.67% larger (larger in 45 of 65). In other words, within this controlled experiment the introduction of Sentinel-1/2 and VIIRS does not mechanically inflate the post-2020 growth rate; in the majority of cities the richer feature stack yields the more conservative recent-epoch estimate, plausibly because radar backscatter and nighttime

lights also improve the exclusion of spectrally confusing non-urban surfaces (bare soil, dry cropland) that inflate Landsat-only ISA. The direction and magnitude of the bias introduced by the sensor transition are therefore city- and period-dependent, not a uniform upward shift. Table R1-1 summarizes the experiment.

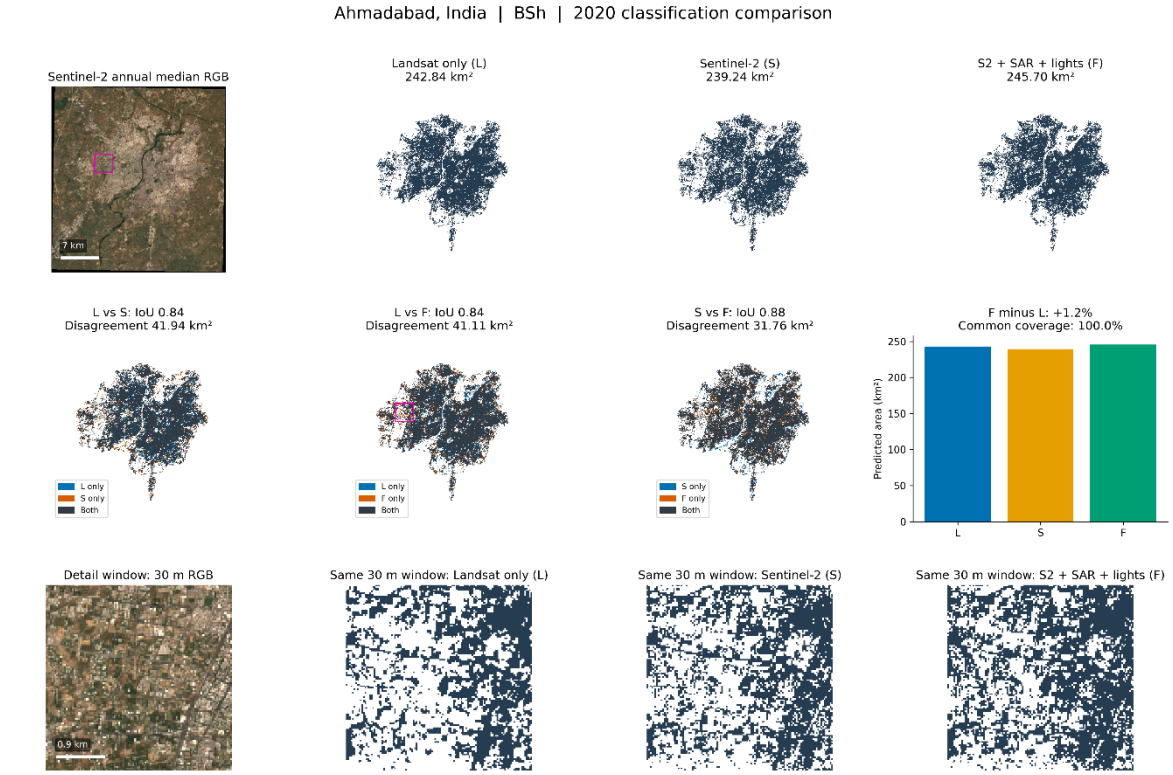
Table R1-1. Summary of the controlled sensor-switching experiment for 65 cities: difference in LCR between the production switching route and the Landsat-only route (switching minus Landsat-only, percentage points per year).

Period	Cities with higher LCR under switching route	Minimum (pp/yr)	Median (pp/yr)	Maximum (pp/yr)
2015–2020	64 of 65	−0.037	+1.805	+14.279
2020–2025	10 of 65	−10.145	−0.785	+1.175

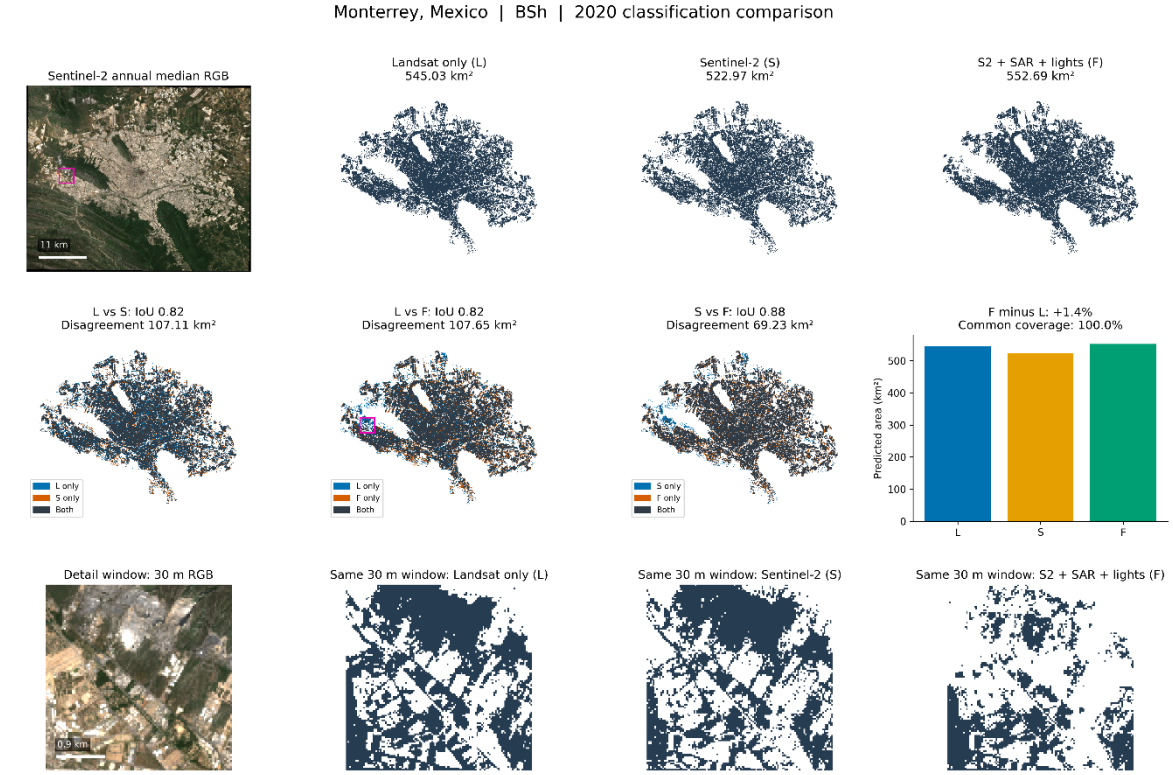
(b) Map-level comparison of 2020 ISA. For 25 cities, we compared 2020 ISA maps from Landsat 8, Sentinel-2, and a combined configuration using Sentinel-2, Sentinel-1, and VIIRS. Within domains derived from the 2020 GUBAD footprints, the combined configuration mapped more ISA than Landsat 8 in 23 of 25 cities (92%). The median relative ISA area difference was +13.10%, and the median spatial intersection-over-union was 0.638.



Same 2020 training and models; common evaluation support. Metrics: 30 m. Overview: 30 m nearest display; detail: 30 m.
White = non-built or outside valid evaluation support. Evaluation-domain construction and limitations are documented in the methods.



Same 2020 training and models; common evaluation support. Metrics: 30 m. Overview: 90 m nearest display; detail: 30 m.
White = non-built or outside valid evaluation support. Evaluation-domain construction and limitations are documented in the methods.



Same 2020 training and models; common evaluation support. Metrics: 30 m. Overview: 120 m nearest display; detail: 30 m.
White = non-built or outside valid evaluation support. Evaluation-domain construction and limitations are documented in the methods.

What this evidence does and does not show.

Together, (a) and (b) demonstrate material methodological sensitivity — the reviewer is right that the transition matters — but they do not establish that the global 2020–2025 growth pattern is physical. The 65 experimental cities are not a representative sample of the 1,611 GUBAD cities, and a representative six-epoch constant-input (Landsat-only across all epochs) reprocessing remains incomplete. We therefore do not claim that the post-2020 upturn is proven to be real; equally, the controlled evidence does not support the opposite assumption of a uniform upward artifact. Both possibilities remain open, and the revised manuscript now says so explicitly.

Changes in the manuscript.

(i) Section 5.1 now explains that the sensor and feature transition can affect mapped built-up area and the resulting LCRPGR values. The relative contributions of methodological effects, physical built-up expansion, and population change to the observed 2020–2025 LCRPGR increase across the full dataset remain unresolved.

(ii) Section 5.3 now states the epoch asymmetry as a first-order limitation, quantifies it with the numbers above, and advises users to treat epoch-to-epoch changes that straddle the 2015→2020 transition with explicit caution.

Revised manuscript text in context

Section 5.1 · p.19, after Fig 12 · replace the paragraph beginning "The increase in LCRPGR"

The distribution of LCRPGR shifted toward higher values during 2020–2025. The share of cities with $0 < \text{LCRPGR} < 1$ declined to 16%, while the categories $1 < \text{LCRPGR} < 2$ and $\text{LCRPGR} > 2$ accounted for 41% and 38%, respectively. The estimates depend on both mapped built-up area and population growth. Changes in sensor inputs and features can affect the mapped built-up area series and the resulting LCRPGR values. The contribution of the sensor transition to this shift across the full dataset remains unresolved.

Section 5.1 · p.19 · insert immediately after the revised paragraph above

We assessed the sensitivity of the mapped UBA time series to the sensor transition by processing the 2015, 2020, and 2025 epochs for 65 cities across six continents using two workflows. The Landsat-only workflow used Landsat spectral indices at all three epochs. The production workflow used Landsat inputs in 2015 and combined Sentinel-2 MSI, Sentinel-1 SAR backscatter, and VIIRS nighttime lights in 2020 and 2025. The paired runs used common spatial support, the same downstream processing settings, and identical population inputs.

The median paired difference in LCR (production minus Landsat-only) was +1.805 percentage points per year in 2015–2020 and −0.785 percentage points per year in 2020–2025 (Table 8). In a complementary comparison for 25 cities, the median relative difference in 2020 ISA between the combined Sentinel-2, Sentinel-1, and VIIRS configuration and Landsat 8 was +13.10% within domains derived from the 2020 GUBAD footprints. The median spatial IoU was 0.638. The contribution of the sensor transition to the 2020–2025 increase in LCRPGR across the full dataset remains uncertain.

Table 8. Paired LCR differences (switching minus Landsat-only) in the 65-city sensor experiment, expressed in percentage points per year.

Period	Cities with higher LCR under switching route	Minimum (pp/yr)	Median (pp/yr)	Maximum (pp/yr)
2015—2020	64 of 65	–0.037	+1.805	+14.279
2020—2025	10 of 65	–10.145	–0.785	+1.175

Section 5.1 · p.19 · lines 378—382 · replace the paragraph after the sensor experiment

Across all intervals, cities with negative LCRPGR values accounted for less than 12% of the sample. These ratios should be interpreted alongside the component rates, particularly when PGR is close to zero. Figure 13 shows the spatial distribution of LCRPGR and its temporal changes across continents.

Section 5.3 · p.22 · lines 452—462 · fill opening limitations paragraph with RC1.1 and RC1.4 combined

Across the 1,611 cities, the mean per-city coefficient of variation (CV) for UBA estimates was 8.3% (SD = 4.1 percentage points). The mean CV was higher for small cities (12.4%, (n = 342)) than for megacities (5.7%, (n = 143). Mixed pixels, heterogeneous land-use patterns, and persistent cloud cover contribute to mapping uncertainty. Anthropogenic but permeable surfaces, such as golf courses and parks with impervious infrastructure, can also introduce ambiguity in boundary delineation.

Morphological processing affects the representation of enclosed open spaces and fragmented settlements. Hole filling can include enclosed parks or other non-built-up spaces in UBA when the holes meet the area and boundary-protection criteria in Sect. 3.2. Minimum-area filtering can remove small, isolated built-up patches in peri-urban areas, while morphological closing can bridge narrow gaps between nearby patches. Uniform processing parameters can therefore affect footprints differently across urban forms.

Temporal monotonicity can suppress genuine ISA loss, a limitation also noted for GISD30 (Zhang et al., 2022). Local impervious cover can decline within a persistent urban footprint. Final UBA nesting separately prevents contraction of the mapped footprint. These constraints limit the use of GUBAD for estimating gross built-up loss or physical retreat.

The transition from Landsat optical inputs in 2000–2015 to Sentinel-2, Sentinel-1, and VIIRS inputs in 2020–2025 can shift the baseline for ISA detection. In the 65-city experiment, the direction and magnitude of the paired LCR differences varied across cities and intervals (Sect. 5.1). PAVA enforces temporal monotonicity and cannot correct systematic differences in detection associated with the change in inputs. Estimates for 2015–2020 span the transition, and estimates for 2020–2025 depend on the 2020 baseline. Reprocessing all six epochs with constant sensor inputs for a representative sample of cities would help quantify this source of uncertainty.

Comment RC1.2 — Validation sample size and stratification

> *"Validating 1,611 complex urban environments with only 10 validation points per city (5 urban, 5 non-urban) is statistically insufficient for local error characterization. While the authors aggregate across all points to claim a narrow global confidence interval, a 10-point sample per city cannot meaningfully capture classification accuracy, edge commission/omission errors, or morphological variance across heterogeneous urban landscapes. Stratified sampling should be scaled proportionally with city footprint, and accuracy must be reported across different urban morphology tiers and climate/biophysical zones (say, Köppen climate zones)."*

Response

We thank the reviewer for this constructive comment. We would like to clarify that the validation scheme was designed to provide consistent class representation across the global set of cities, rather than to estimate precise accuracy separately for every individual city. This balanced allocation prevents either class from being entirely absent from a city's validation sample and supports comparison of classification outcomes across the geographically distributed city set.

We retained the fixed allocation of ten points per city and epoch. The pooled metrics describe ISA classification agreement under this sampling design. The allocation provides limited coverage of urban-edge errors and local morphological variation, particularly in large and spatially complex cities. The revised manuscript clarifies the scope of the assessment. Accuracy by urban morphology and Köppen climate zone remains uncharacterized.

Changes in the manuscript

(i) The validation methods now explain the five/five allocation, identify the sampling unit, describe how reference labels were assigned, and specify how observations were pooled across cities and epochs.

(ii) The accuracy results now identify the reported statistics as measures of ISA classification agreement. We reviewed all reported aggregate confidence intervals against the documented class allocation and the grouping of observations by city, and retained them only where the calculation and interpretation are supported by the sampling design.

(iii) The limitations section clarifies the scope of the pooled ISA accuracy statistics and identifies the uncertainties that remain at the city level, at urban edges, and across morphology and climate zones.

Revised manuscript text in context

Section 3.3 · p. 9, lines 203–205 · opening paragraph

The accuracy of the extracted impervious surface area was evaluated using a confusion-matrix-based validation framework. Spatial patterns were further assessed by comparing ISA and UBA expansion patterns against Google Earth reference imagery for representative cities across diverse urban contexts (Fig. 5).

Section 3.3 · p. 9, lines 209–216 · sampling and interpretation

Validation was performed using stratified random sampling with a planned allocation of ten points per city and epoch, five assigned to the ISA class and five to the non-ISA class (Foody, 2010; Sun et al.,

2019). The sampling unit was a location within a city and epoch. The reference labels were acquired from visual interpretation of concurrent high-resolution satellite imagery. The map and reference labels were pooled across sampled cities within each epoch to compute OA, PA, UA and kappa. These statistics describe ISA classification agreement under the fixed allocation. The design gives limited coverage of local edge errors and does not support precise city-specific accuracy estimates.

Section 4.1 · p. 11, lines 238–245 · interpretation of the accuracy statistics

The mean overall accuracy (OA), the average producer's accuracy (PA) and the average user's accuracy (UA) as well as the kappa coefficient reached 95.18%, 95.40%, 94.95% and 0.90, respectively. In the individual years, OA ranged from 94.62% to 96.57% and kappa ranged from 0.89 to 0.93. These values describe ISA classification agreement for the pooled validation sample. Independent reference footprints would be required to assess the accuracy of the derived UBA boundaries.

Table 4 · p. 11 · full caption

Table 4. ISA classification accuracy metrics by epoch (2000–2025), with confusion-matrix counts for the pooled validation sample.

Figure 6 · p. 12 · full caption

Figure 6. ISA classification agreement by epoch (2000–2025) for the pooled balanced sample.

Section 5.3 · p. 22 · insert after the discussion of classification uncertainty

The fixed allocation supports a pooled assessment of ISA classification agreement for the sampled points. City-specific accuracy and urban-edge errors remain poorly characterized, particularly in large and spatially complex cities. Variation in accuracy by urban morphology and Köppen climate zone was outside the scope of the assessment.

Comment RC1.3 — Population constraints and circularity in LCRPGR

> *"The conversion from impervious surface area to urban built area explicitly incorporates 1 km LandScan population density thresholds to delineate the urban footprint. When this footprint is subsequently combined with UN population growth rates to estimate LCRPGR, the analysis borders on circular reasoning. This is because population data are used both to constrain whether land is included in the built-up footprint and as the denominator against which that footprint's growth is evaluated. The authors must clarify and bound how demographic assumptions in LandScan bias the resulting land consumption rate calculations."*

Response

The reviewer identifies a real risk, and we address it on three levels: the provenance of the population data at the two ends of the calculation, the audited behavior of the production code, and an honest statement of what we cannot yet exclude.

(a) Population inputs. City-level UN WUP population estimates supply the population growth rate used in the LCRPGR denominator. LandScan provides a separate 1 km gridded population input for the diagnostic mask in the inspected delineation workflow. Both sources may draw on overlapping census information. The production audit establishes where the LandScan-derived mask enters the inspected workflow.

(b) Production code audit. In the inspected automatic export, ISA-based metropolitan-context screening determines component inclusion. The LandScan-derived mask is generated as a diagnostic intermediate and does not directly gate components in that branch. Section 3.2 now distinguishes the diagnostic mask from the export screen.

(c) Scope of the audit. The processing history of every released footprint has yet to be verified. The audit establishes how the inspected export branch behaves. A production-matched comparison with LandScan disabled or held fixed would be needed to quantify any residual effect on released UBA and LCR.

Changes in the manuscript

Section 3.2 distinguishes the LandScan diagnostic branch from the automatic export. Section 5.1 identifies the city-level WUP series used to compute PGR, and Section 5.3 states the scope of the code audit and the remaining release-level uncertainty.

Revised manuscript text in context

Section 3.2 · p. 8, lines 163–166 · opening definition

In this study, urban built-up area (UBA) is defined as a spatially contiguous built-up footprint derived from impervious surfaces using morphological refinement and ISA-based metropolitan-context screening (Fig. 4). The LandScan-derived urban-centre mask is a separate diagnostic intermediate in the inspected automatic workflow.

Figure 4 · p. 8 · full caption

Figure 4. Spatial-morphological conversion workflow from ISA to UBA, including resolution harmonization, morphological refinement and metropolitan-context screening. LandScan inputs contribute to the separate diagnostic branch.

Section 3.2 · p. 9, lines 191–201 · LandScan processing paragraph

A population-constrained urban-center mask was further constructed using LandScan population data on the same 1 km grid. The diagnostic mask used a density threshold of 300 persons km^{-2} and a minimum population of 5,000 per contiguous component (Dijkstra et al., 2021), with eight-neighbour contiguity, ISA-based bridging and one-cell dilation. In the inspected automatic export, the ISA-derived metropolitan-context screen determines component inclusion. The LandScan-derived mask is retained as a diagnostic intermediate. Final UBA masks were vectorized, topologically corrected, and subjected to the same minimum-area filtering across all cities and epochs. Intermediate rasters were retained for workflow auditing.

Section 5.1 · p. 17, lines 341–343 · land and population inputs

City-level population estimates from the United Nations World Urbanization Prospects were used to compute population growth rates for SDG indicator 11.3.1. GUBAD supplied the urban land areas used to compute land consumption rates. The LandScan 1 km ambient population grids were used for the diagnostic mask described in Sect. 3.2. The sources have different spatial supports and may draw on overlapping census inputs.

Section 5.3 · p. 22 · insert after the validation limitation

The workflow audit identifies the direct role of LandScan in the inspected automatic branch. The processing history of every released footprint remains to be verified. A production-matched comparison of the original workflow against no-population and fixed-population configurations would be needed to quantify residual demographic effects on the released LCR values.

Comment RC1.4 — Monotonicity imposed by pixel-level PAVA

> *"The pixel-level PAVA isotonic regression strictly enforces monotonic urban expansion, explicitly forbidding transitions from built-up to non-built-up land. While global urbanization dominates, this hard constraint artificially masks real urban abandonment, structural retreat, conflict-driven displacement, and brownfield re-greening. The manuscript needs to quantify how much true land-use change or sensor noise is overridden by this monotonic assumption."*

Response

We thank the reviewer for raising this concern. We agree that imposing monotonicity prevents the mapped time series from representing genuine losses of impervious surface. The original wording attributed the overridden classifications primarily to noise without independent evidence for that interpretation. To quantify the effect of the constraint, we expanded the sensitivity experiment to 65 pilot cities across six continents. We compared three configurations for 2015, 2020, and 2025 under both the Landsat-only route and the switching sensor route described in RC1.1: no PAVA, equal-weight PAVA (1:1:1), and PAVA with double weight assigned to the first epoch (2:1:1). Comparisons used common spatial support within a fixed 2025 analysis domain for each city.

(a) Pixel-level classification changes. For each city and sensor route, we divided the number of pixel locations whose binary label changed in at least one epoch after equal-weight PAVA by the number classified as impervious in at least one epoch immediately before PAVA. The denominator was calculated after spatial mode filtering. The affected proportion ranged from 7.80% to 55.80% across the 65 cities for the Landsat-only route (median 20.15%) and from 5.30% to 43.15% for the switching route (median 16.71%).

The sequence audit shows how PAVA resolves apparent losses. For example, under equal weighting, 110 becomes 111, assigning an impervious label in 2025, while 101 becomes 001, removing the impervious label in 2015. Here 1 denotes impervious surface and 0 denotes non-impervious surface, in the order 2015, 2020, and 2025. Under the first-epoch double-weight configuration, 101 instead becomes 111. The treatment of an apparent loss can therefore depend on the weighting scheme. Across the tested domains, equal-weight PAVA removed all 16,862,833 apparent adjacent-epoch ISA loss transitions in the Landsat-only route and all 14,884,842 in the switching route. These are counts of mapped pixel transitions across the two adjacent epoch pairs.

(b) Effects on urban built-up area and LCR. With the water mask and final UBA nesting held constant, equal-weight PAVA increased the 2020–2025 LCR relative to the no-PAVA configuration in all 65 pilot cities under both sensor routes. Increases ranged from 0.159 to 4.055 percentage points per year for the Landsat-only route (median 0.628) and from 0.024 to 1.408 percentage points per year for the switching route (median 0.353). Median increases in final 2025 UBA were 4.90% and 2.07%, respectively. Before final nesting, with the water mask held constant, the corresponding median LCR increases were 2.523 and 1.491 percentage points per year. The comparison after nesting measures PAVA's effect conditional on a separate constraint that also prevents mapped UBA contraction.

What this evidence does and does not show

The experiment quantifies how often PAVA changes mapped impervious-surface classifications and how those changes affect derived UBA and LCR. The proportion of overrides attributed to genuine surface loss versus sensor or classification error remains unknown: estimating that proportion requires

independent interpretation of a sample of candidate transitions against dated reference imagery, and attributing physical losses to processes such as abandonment or conflict-driven displacement would require additional evidence. The experiment covers three epochs using a controlled automated reconstruction within the common analysis domain, and the reported magnitudes apply to those experimental conditions.

Changes in the manuscript.

(i) Section 3.1 now specifies the PAVA weighting scheme, explains how the algorithm resolves non-monotonic pixel sequences, and replaces the claim that it primarily suppresses noise with the measured classification changes.

(ii) Section 5.3 now describes suppressed contraction as a structural limitation of both pixel-level PAVA and final UBA nesting, and clarifies that the final GUBAD layers are monotonic by construction: they cannot represent urban abandonment, conflict-driven retreat or brownfield re-greening, and should not be used to study urban contraction. The quantified magnitudes from the 65-city experiment are reported alongside this statement.

(iii) The response reports the apparent ISA loss counts and the effects on final UBA and LCR for the tested configurations.

Revised manuscript text in context

Section 3.1 · p. 8, lines 155–161 · full classification paragraph with RC1.4 and RC1.5 combined

ISA classification employed Random Forest ensembles (100 trees) implemented on Google Earth Engine. For the Sentinel-2 epochs (2020 and 2025), the area-weighted fraction of each 30 m cell classified as impervious at 10 m was calculated before thresholding. Cells with an impervious fraction of at least 0.50 were retained in the binary ISA mask on the 30 m grid. Pool-adjacent-violators algorithm (PAVA) isotonic regression was applied to impose a non-decreasing ISA time series at each pixel (Ayer et al., 1955; Barlow, 1980). For the 2015, 2020 and 2025 sensitivity test, we compared no PAVA, equal weights (1:1:1) and double weight on the first epoch (2:1:1). With 1 denoting ISA, equal weighting changed 110 to 111 and 101 to 001, whereas double weighting of the first epoch changed 101 to 111. Thus, the epoch adjusted by PAVA depends on the input sequence and weighting.

Section 5.3 · p. 22, lines 452–462 · full opening limitations paragraph with RC1.1 and RC1.4 combined

Limitations include quantified classification uncertainties: per-city coefficient of variation (CV) for UBA estimates averaged 8.3% ± 4.1% (mean ± SD on 1,611 cities), and greater uncertainty in Small Cities (CV=12.4%, n=342) as compared with Megacities (CV=5.7%, n=143). Increased uncertainty in smaller cities is attributable to higher spectral heterogeneity relative to sensor resolution. Despite these uncertainties, the spatial and temporal completeness of GUBAD advances current global monitoring capacities beyond fragmented national reporting systems. Remaining uncertainties are present in areas with complicated land-use patterns, mixed pixels, or continuous cloud coverage. The monotonicity constraint can suppress genuine ISA loss, as also recognized for GISD30 (Zhang et al., 2022). Local loss of impervious cover can occur within a persistent urban footprint. Final UBA nesting separately prevents contraction at the footprint scale. The resulting series is therefore unsuitable for estimating gross built-up loss or physical retreat. Additionally, anthropogenic but permeable surfaces (e.g., golf

courses, parks with impervious infrastructure) introduce subjectivity in boundary delineation. The change from Landsat optical inputs in 2000–2015 to Sentinel-2, Sentinel-1 and VIIRS inputs in 2020–2025 can shift the baseline for ISA detection. The 65-city experiment in Sect. 5.1 found that the magnitude and direction of LCR differences vary across intervals and cities. PAVA constrains temporal sequences but cannot correct systematic cross-sensor differences. Changes spanning 2015–2020 should therefore be interpreted with particular caution. A representative six-epoch constant-input reconstruction would provide a more complete assessment.

Section 5.3 · p. 22 · insert after the full opening limitations paragraph above

In the 65-city experiment, equal-weight PAVA changed at least one epoch label at a city median of 20.15% of locations classified as ISA at least once before PAVA under the Landsat-only route (range 7.80%–55.80%) and 16.71% under the switching route (5.30%–43.15%). With final nesting and water masking held constant, median 2020–2025 LCR increased by 0.628 and 0.353 percentage points per year, respectively. Median final 2025 UBA increased by 4.90% and 2.07%. Before nesting, the median LCR increases were 2.523 and 1.491 percentage points per year, respectively. Across the tested domains, equal-weight PAVA removed all 16,862,833 apparent adjacent-epoch ISA losses in the Landsat-only route and all 14,884,842 in the switching route. These counts and effect sizes describe the tested reconstructions. Independent dated reference imagery is needed to distinguish corrections of classification errors from suppression of genuine ISA loss.

References · p. 25 · insert alphabetically among the Zhang references

Zhang, X., Liu, L., Zhao, T., Gao, Y., Chen, X., and Mi, J.: GISD30: global 30 m impervious-surface dynamic dataset from 1985 to 2020 using time-series Landsat imagery on the Google Earth Engine platform, Earth Syst. Sci. Data, 14, 1831–1856, https://doi.org/10.5194/essd-14-1831-2022, 2022.

Comment RC1.5 — Aggregation of 10 m Sentinel-2 ISA to 30 m

> *"Aggregating 10 m Sentinel-2 ISA down to 30 m using a simple 0.5 area-weighted majority threshold introduces clear scale bias compared to native Landsat 30 m mixed pixels. A native 30 m pixel integrates reflectance over a 900 m2 instantaneous field of view, whereas a binary threshold on 10 m sub-pixels systematically eliminates sub-pixel impervious elements below 50% fraction (e.g., suburban residential streets or sparse informal settlements). A sub-pixel impervious fraction comparison or a soft-aggregation scheme should be explored."*

Response

We assessed the sensitivity of binary ISA and derived UBA to the 10 m-to-30 m aggregation threshold. We compared thresholds of 0.25, 0.50 and 0.75 in the 2025 pilot sample and a smaller 2020 subset. We assessed area preservation against the continuous source fraction and explored agreement with recorded built-up-area reference labels.

Threshold sensitivity. The binary footprint is strongly threshold-sensitive, in the 2025 pilot sample. relative to 0.50, a threshold of 0.25 increases the median ISA and UBA area by 27.01% and 26.15% respectively, while 0.75 reduces them by 24.88% and 25.99%

Why 0.5 is retained — and what that justification is worth

Among the three thresholds tested, 0.50 most closely preserved the continuous source area in most pilot cities, with a median absolute discrepancy of 2.35%. Exploratory agreement with the recorded UBA reference labels was 89.59% at 0.25 and 87.92% at 0.50. The results describe two criteria for the tested samples: preserving source fractional area and agreeing with the reference labels. Cross-sensor differences with native Landsat observations remain unresolved.

Retaining the continuous 30 m ISA fractions would allow users to examine cells with low impervious fractions and apply alternative thresholds.

Changes in the manuscript.

Sections 3.1 and 3.2 describe the aggregation procedure and the quantified threshold sensitivity. Section 5.3 discusses the distinction between source-area preservation and reference-label agreement.

Revised manuscript text in context

Section 3.1 · p. 8, lines 155–161 · full classification paragraph with RC1.4 and RC1.5 combined

ISA classification employed Random Forest ensembles (100 trees) implemented on Google Earth Engine. For the Sentinel-2 epochs (2020 and 2025), the area-weighted fraction of each 30 m cell classified as impervious at 10 m was calculated before thresholding. Cells with an impervious fraction of at least 0.50 were retained in the binary ISA mask on the 30 m grid. Pool-adjacent-violators algorithm (PAVA) isotonic regression was applied to impose a non-decreasing ISA time series at each pixel (Ayer et al., 1955; Barlow, 1980). For the 2015, 2020 and 2025 sensitivity test, we compared no PAVA, equal weights (1:1:1) and double weight on the first epoch (2:1:1). With 1 denoting ISA, equal weighting changed 110 to 111 and 101 to 001, whereas double weighting of the first epoch changed 101 to 111. Thus, the epoch adjusted by PAVA depends on the input sequence and weighting.

Section 3.2 · p. 8, lines 167–173 · full harmonization paragraph

UBA were derived from the extracted ISA maps using a standardized spatial workflow. AOIs were defined using GADM v3.6 administrative units for cities outside China; Chinese cities employed officially designated urban core boundaries. The 30 m fractions derived from 10 m ISA were used to compare thresholds of 0.25, 0.50 and 0.75. In the 2025 test, relative to 0.50, a threshold of 0.25 increased median ISA and UBA by 27.01% and 26.15%, respectively. A threshold of 0.75 decreased these areas by 24.88% and 25.99%, respectively. The 0.50 threshold most closely preserved source fractional area in most tested cities, with a median absolute discrepancy of 2.35%.

Section 5.3 · p. 22 · insert after the paragraph on the sensor transition

A common 30 m grid does not eliminate differences between native Landsat observations and thresholded Sentinel-2 fractions. In the exploratory comparison with recorded UBA reference labels, agreement was 89.59% at 0.25 and 87.92% at 0.50. Preservation of source fractional area and agreement with UBA reference labels therefore favoured different thresholds. The fraction comparison does not independently assess native Landsat accuracy. Retaining continuous 30 m ISA fractions would enable users to examine low-fraction cover and apply alternative thresholds.

Comment RC1.6 — Added value over existing products

> *"As a final note, it is unclear to me why this dataset is more relevant for estimating LCRPGR than the multiple other time series of impervious, or built, or urban land that currently exist in the literature, which already do not agree on the degree of urbanization globally and regionally (Chakraborty et al. 2024). Just because new datasets can be developed does not mean that they need to be developed."*

Response

The reviewer is right, and the original manuscript overstates its case. Correlation with existing products does not establish superiority, and product-to-product disagreement — documented comprehensively by Chakraborty et al. (2024), whom we now cite — cannot be read as evidence in favour of whichever product is newest. We have removed comparative-superiority language throughout (including "better performance" and claims of "improved capture" framed as accuracy statements in Sections 4.2.3, 5.2 and the Conclusions), and we now state explicitly that no independent application benchmark demonstrating better LCRPGR performance than existing products has been carried out.

What GUBAD is for

We have reframed the contribution as complementarity rather than competition. The evidence supports four specific, verifiable claims. First, coverage and definition: GUBAD is, to our knowledge, the only product that applies a UN urban-agglomeration-consistent, population-aware delineation at 30 m resolution to 1,611 cities across six five-year epochs under a single processing chain, with city-level population and boundary harmonization to the WUP city list. Second, readiness for indicator work: GUBAD is organized per city and ships with polygon boundaries, attribute tables and pre-computed SDG 11.3.1 quantities (LCR, PGR, LCRPGR), whereas wall-to-wall global layers require substantial non-trivial processing — city matching, boundary harmonization, contiguity and population screening — before they can support city-scale indicator estimation; that processing chain is exactly where definitional inconsistencies enter, and GUBAD makes it explicit, uniform and auditable. Third, diagnostic value under disagreement: precisely because existing products disagree (Chakraborty et al., 2024), an additional estimate built on a transparent, physically based and independently documented chain has diagnostic value for the community — disagreements can only be diagnosed if the methodological choices behind each estimate are inspectable, and GUBAD archives all intermediate rasters and releases its code for that purpose. Fourth, conceptual separation: GUBAD separates the physical layer (ISA) from the functional layer (UBA), so users who disagree with our ISA-to-UBA translation can apply their own, which boundary-oriented or statistics-based products do not permit.

Changes in the manuscript

Sections 4.2.3 and 5.2 have been rewritten to present inter-product comparisons as consistency diagnostics rather than accuracy evidence; the Introduction now cites Chakraborty et al. (2024) and motivates GUBAD through definitional transparency, city-scale readiness and auditability; the Conclusions no longer claim superior performance. Chakraborty et al. (2024) has been added to the reference list.

Revised manuscript text in context

Abstract · p. 1, lines 11–28 · full revised abstract

Abstract. Urban land-use efficiency (LUE), defined under Sustainable Development Goal (SDG) 11.3.1, is widely used to evaluate the coordination between urban expansion and population growth. However, its global assessment remains constrained by inconsistent definitions of urban boundaries and the lack of long-term, 30 m spatial resolution urban built-up area (UBA) datasets. Here we present the Global Urban Built-up Area Dataset (GUBAD), a multi-epoch dataset at 30 m spatial resolution covering 1,611 cities worldwide with populations exceeding 300,000 inhabitants from 2000 to 2025. GUBAD applied standardized UN-defined urban agglomeration boundaries using a spatial morphological framework integrating impervious surface data with ISA-based metropolitan-context screening. Impervious surface areas (ISA) were extracted using Random Forest classification of Landsat and Sentinel-1/2 imagery, combined with isotonic regression-based temporal correction. The pooled ISA classification assessment yielded a mean overall accuracy of 95.18% and Kappa coefficient of 0.90 across all epochs. Comparisons with existing urban datasets describe differences associated with mapped classes, spatial definitions and observation inputs. We further demonstrate the utility of GUBAD for estimating SDG indicator 11.3.1, revealing spatiotemporal variations in the relationship between land consumption rate and population growth rate (LCRPGR) at the global city scale. GUBAD provides city-level footprints and population-linked indicators for reproducible monitoring of mapped urban expansion and SDG 11.3.1 assessment. GUBAD is available at Zenodo under CC BY 4.0 (Gunasekera et al., 2026) with DOI: https://doi.org/10.5281/zenodo.20051123. The dataset is released as version 1.0 with planned updates every 5 years.

Introduction · p. 2, lines 63–74 · dataset positioning

While GHSL-SMOD R2023A provides global settlement layers at 1 km resolution with temporal coverage (1975–2030), GUB delineates urban extents using a different mapping target (Li et al., 2020; Schiavina et al., 2023). Differences in urban definitions, mapped classes and spatial scale can lead to substantial disagreement among estimates of urban land (Chakraborty et al., 2024). GUBAD complements these products by providing 30 m footprints for 1,611 cities across six epochs and linking them to city identifiers, population records and SDG 11.3.1 indicators. The workflow separates ISA extraction from UBA delineation, making the footprint definition and its effect on the area entering the land consumption rate explicit.

Section 4.2.3 · p. 14, lines 279–291 · first comparison paragraph

GUBAD was compared with the GHSL-SMOD, GUB, MGUP (approximately 1,500 matched cities each), and AUE (approximately 150 cities) datasets at overlapping temporal snapshots (Fig. 9). R^2 ranged from 0.855 to 0.873 for GHSL-SMOD (2000–2025), from 0.788 to 0.843 for GUB (2000–2018), and from 0.742 to 0.768 for MGUP (2001–2018). The AUE comparison yielded $R^2 \approx 0.76$ for the 2000 and 2015 epochs. All reference datasets were evaluated within the GUBAD-derived city AOIs. The statistics describe inter-product agreement for the matched samples. Differences may arise from the mapped classes, definition of urban extent, observation date and spatial resolution.

Section 4.2.3 · p. 14, lines 292–294 · final comparison paragraph

The visual comparisons illustrate differences in the delineation of fragmented and heterogeneous urban structures, particularly in rapidly urbanizing regions. Some products delineate larger urban extents in the examples shown. Differences in mapped classes, boundary definitions and spatial resolution may contribute to these patterns. A common reference assessment would be required to compare delineation accuracy across products.

Figure 9 · p. 15 · full caption

Figure 9. Comparison of GUBAD with global remote sensing urban products. Scatterplots compare GUBAD UBA (km²; y-axis) against (a–b) AUE, (c–h) GHSL-SMOD, (i–m) GUB, and (n–r) MGUP estimates (km²; x-axis) across benchmark years. Blue dashed line: 1:1 identity line; solid orange line: linear regression fit. Insets report R^2, RMSE (km²), MAE (km²), and bias (km²).

Section 5.2 · p. 21, lines 412–438 · replace the first three paragraphs with the two paragraphs below

Global products represent urban land using different mapped classes, boundary definitions, resolutions and observation years (Chakraborty et al., 2024). The comparisons in Sect. 4.2 describe agreement among products. In the illustrated coastal and peri-urban examples, MGUP and DEGURBA delineate larger urban extents, while AUE shows smoother boundaries for some fragmented settlements. The mapped targets and delineation procedures may contribute to these differences.

GUBAD complements existing datasets by providing 30 m urban footprints and linked attributes for a common set of cities and epochs. The separation of ISA extraction and UBA delineation makes the footprint definition explicit. Linked population records and precomputed indicators support city-level SDG 11.3.1 analysis. An independent assessment using a common footprint definition and reference information would be needed to compare LCRPGR performance across products.

Section 7 · p. 23, lines 494–503 · full conclusions paragraph

This study presents the Global Urban Built-up Area Dataset (GUBAD), a multitemporal dataset at 30 m spatial resolution covering 1,611 cities worldwide with populations exceeding 300,000 inhabitants from 2000 to 2025. GUBAD was developed using multi-source satellite observations and a consistent machine-learning framework. The footprints and associated population records support city-level estimates of land consumption, population growth and LCRPGR. The distinction between ISA classification and UBA delineation provides an explicit footprint definition for SDG indicator 11.3.1 analysis. Interpretation of temporal change should account for the transition in observation inputs, the monotonicity constraints and the scope of the validation sample.

References · p. 24 · insert after Cao et al.

Chakraborty, T. C., Venter, Z. S., Demuzere, M., Zhan, W., Gao, J., Zhao, L., and Qian, Y.: Large disagreements in estimates of urban land across scales and their implications, Nat. Commun., 15, 9165, https://doi.org/10.1038/s41467-024-52241-5, 2024.

Minor comment — Dispersion and size-dependent bias in Figure 8

> *"Figure 8: The scatterplots comparing GUBAD with WUP show considerable dispersion and systematic bias for high-density metropolitan areas (R^2 dropping to ~0.54–0.60 in later epochs). A more thorough breakdown of why these discrepancies scale with city size is needed."*

Response

We thank the reviewer for pushing us to look behind the scatterplots. We decomposed the GUBAD–WUP comparison for the 1,100 matched cities — GUBAD built-up areas against urban area statistics from the UN World Urbanization Prospects, 2018 revision (UN-DESA, 2018; https://population.un.org/wup/, 2018 revision data files), the same source that provides the city list and population figures used throughout the manuscript. In the comparison table, columns W2000–W2025 hold the WUP urban areas and columns A2000–A2025 the corresponding GUBAD areas for the six epochs. The decomposition follows three axes: the systematic offset, its scaling with city size, and the driver of the declining R^2. The full per-epoch statistics are given in Table R2-1 and the size-stratified results in Table R2-2; both now appear in Section 4.2.2.

Table R2-1. Per-epoch statistics of the GUBAD–WUP comparison for the 1,100 matched cities (GUBAD UBA regressed on WUP urban area; bias defined as GUBAD minus WUP).

Epoch	R^2	Slope	RMSE (km²)	MAE (km²)	Mean bias (km²)	Median GUBAD/WUP
2000	0.655	2.151	74.9	46.0	+45.8	2.71
2005	0.650	2.364	92.9	57.5	+57.3	2.93
2010	0.633	2.411	109.3	68.6	+68.4	3.11
2015	0.625	2.522	128.5	80.7	+80.5	3.30
2020	0.598	2.705	153.3	95.6	+95.4	3.62
2025	0.544	2.815	189.8	115.6	+115.3	3.93

(i) A systematic, growing offset whose origin is definitional. GUBAD areas exceed the WUP figures by a median factor of 2.71 in 2000, rising to 3.93 in 2025, with the mean bias widening from +45.8 km² to +115.3 km². The dominant cause is the definition behind the WUP area figures: they reproduce national statistical reporting, which for many countries corresponds to the administrative city proper or a narrowly drawn agglomeration core, whereas GUBAD maps the full contiguous impervious footprint under the UN urban-agglomeration concept, including the continuously built fringe. The offset is not universal — for Seoul, for instance, the WUP figure exceeds the GUBAD footprint (ratio ≈ 0.80 in 2025) — which is itself diagnostic: the cross-country heterogeneity of national reporting definitions, not a uniform measurement error, drives much of the dispersion. We have corrected a related error in the previous manuscript text: the statement that statistical datasets "generally include non-built-up areas" while GUBAD is "more conservative" does not hold for the WUP comparison, where GUBAD is systematically the larger estimate; Section 4.2.2 now describes the direction and definitional origin of the offset correctly.

(ii) How the discrepancy scales with city size. The absolute deviation grows approximately linearly with reported city size (regression of |GUBAD − WUP| on WUP area: slope ≈ 1.85 km² per km², R^2 =

0.35), so the largest metropolitan areas dominate the visible dispersion and pull the regression line — this is the "systematic bias for high-density metropolitan areas" the reviewer observes, and it is primarily an absolute-scale effect. In relative terms the pattern is actually reversed: the median GUBAD/WUP ratio in 2025 is highest for the smallest cities and declines with size (4.78 in the smallest quintile versus 3.12 in the largest; Table R2-2). In other words, the divergence is proportionally largest for small cities — where an administrative core under-bounds the physical fabric most severely in relative terms — but it is the largest cities, with their big absolute deviations, that dominate the scatterplot and depress the R^2.

Table R2-2. GUBAD–WUP comparison stratified by city-size quintile (quintiles defined on the WUP 2025 urban area; ratios are GUBAD divided by WUP).

City-size quintile (WUP 2025)	n	Median WUP area (km²)	Median GUBAD area (km²)	Median ratio 2000	Median ratio 2025
Q1 (smallest)	220	8.19	36.81	2.63	4.78
Q2	220	14.34	59.41	2.86	4.15
Q3	220	20.87	85.08	2.77	3.95
Q4	220	33.32	136.46	2.72	3.91
Q5 (largest)	220	76.16	298.63	2.54	3.12

(iii) Why R^2 declined in later epochs. The decline from 0.655 (2000) to 0.544 (2025) is driven disproportionately by China. Excluding the 333 Chinese cities, the R^2 is essentially stable across the record (0.634 in 2000 versus 0.635 in 2025). Chinese cities expanded physically far faster than their reported statistical urban areas, which in the WUP compilation reflect national definitions and infrequent administrative updates, so the two series drift apart precisely where physical expansion was fastest. Regional medians tell the same story in the opposite direction: the 2025 median GUBAD/WUP ratio is highest where national reporting areas are narrowest relative to the physical agglomeration — Europe (5.80) and North America (4.72) — and lowest in Africa (2.68), with Asia (3.91) pulled upward by its rapidly expanding cities (median ratio growth 2000→2025 of 1.52, against 1.20 in North America and 1.18 in Africa). The increasing dispersion in later epochs therefore reflects a widening gap between rapidly evolving physical urban fabric and static or slowly updated statistical delineations, rather than a degradation of the GUBAD estimates in recent epochs — a reading consistent with the stable ex-China R^2.

Changes in the manuscript

Section 4.2.2 has been expanded with the per-epoch statistics (Table R2-1), the size-class breakdown (Table R2-2), and the regional and ex-China diagnostics; the interpretation of Figure 8 now distinguishes absolute-scale dispersion from relative deviation and attributes the decline in R^2 to the widening gap between physical urban fabric and slowly updated statistical delineations rather than to a degradation of the GUBAD estimates in recent epochs. The corrected statement on the direction of the GUBAD–WUP offset replaces the previous wording.

Revised manuscript text in context

Section 4.2.2 · pp. 13–14, lines 270–277 · opening paragraph

The derived UBA estimates were compared with a statistical urban-area series (Fig. 8). Of 1,611 cities, 1,100 cities were successfully matched across the six epochs. The mean signed difference (GUBAD minus the comparison area) increased from +45.8 km² in 2000 to +115.3 km² in 2025, and the median area ratio increased from 2.71 to 3.93 (Table 5). The offset varied among cities. For Seoul, the 2025 ratio was approximately 0.80. Differences in footprint definitions and reporting boundaries may contribute to the observed offsets.

Section 4.2.2 · pp. 13–14 · replace the remaining two interpretation paragraphs

Absolute differences increased with reported comparison area (slope approximately 1.85 km² per km², $R^2 = 0.35$). Relative differences were larger in the smallest 2025 area quintile, where the median ratio was 4.78, compared with 3.12 in the largest quintile (Table 6). These strata represent urban area. Population density was not evaluated separately.

The pooled R^2 decreased from 0.655 in 2000 to 0.544 in 2025. When the 333 matched Chinese cities were excluded, the corresponding R^2 values were 0.634 and 0.635, indicating sensitivity to the composition of the matched sample. Regional median ratios in 2025 were 5.80 in Europe, 4.72 in North America, 3.91 in Asia and 2.68 in Africa. The comparison does not isolate the contributions of footprint definitions, physical expansion and reporting practices to these differences.

Figure 8 · p. 14 · full caption

Figure 8. Comparison of GUBAD UBA (km²; y-axis) with statistical urban-area estimates (km²; x-axis) for matched cities (n = 1,100). The comparison fields are labelled WUP in the panels. Blue dashed line: 1:1 identity line; solid orange line: linear regression fit. Insets report R^2, RMSE (km²), MAE (km²), and bias (km²). Panel bias is comparison area minus GUBAD. Table 5 reports the reverse sign convention.

Table 5. Comparison statistics for 1,100 matched cities across six epochs. Two near-zero comparison values in 2010 were excluded from ratio calculations.

Epoch	R^2	Slope	RMSE (km²)	MAE (km²)	Mean difference (km²)	Median area ratio
2000	0.655	2.151	74.9	46.0	+45.8	2.71
2005	0.650	2.364	92.9	57.5	+57.3	2.93
2010	0.633	2.411	109.3	68.6	+68.4	3.11
2015	0.625	2.522	128.5	80.7	+80.5	3.30
2020	0.598	2.705	153.3	95.6	+95.4	3.62
2025	0.544	2.815	189.8	115.6	+115.3	3.93

Table 6. Comparison by quintile of reported 2025 statistical urban area. Ratios are GUBAD divided by the comparison area.

2025 area quintile	n	Median comparison area (km²)	Median GUBAD area (km²)	Median ratio, 2000	Median ratio, 2025

Q1 (smallest)	220	8.19	36.81	2.63	4.78
Q2	220	14.34	59.41	2.86	4.15
Q3	220	20.87	85.08	2.77	3.95
Q4	220	33.32	136.46	2.72	3.91
Q5 (largest)	220	76.16	298.63	2.54	3.12

We hope the reviewer finds the revised manuscript and these responses satisfactory, and we remain happy to provide any additional material required.

Sincerely,

Dinoo Gunasekera, Zhongchang Sun, Wenjie Du, Jian Gao, Yunpeng Zhang, Yuyu Zhou, Robert Ndugwa, and Huadong Guo

On behalf of all authors