https://doi.org/10.5194/essd-2026-381
Preprint. Discussion started: 16 September 2026

Open Access Earth System Science Data Discussions

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

Dinoo Gunasekera[1,2,3], Zhongchang Sun[1,2,3], Wenjie Du[1,2,3], Jian Gao[4], Yunpeng Zhang[1,2,3], Yuyu Zhou[5], Robert Ndugwa[6], Huadong Guo[1,2,3]

[1]University of Chinese Academy of Sciences, 100094, Beijing, China
[2]Aerospace Information Research Institute, Chinese Academy of Sciences, 100094, Beijing, China
[3]International Research Center of Big Data for Sustainable Development Goals, 100094, Beijing, China
[4]School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
[5]Department of Geography, The University of Hong Kong, 999077, Hong Kong, China
[6]UN-Habitat, Nairobi, Kenya

Correspondence to: Zhongchang Sun (sunzc@aircas.ac.cn), Wenjie Du (duwj1031@126.com)

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 population constraints. Impervious surface areas (ISA) were extracted using Random Forest classification of Landsat and Sentinel-1/2 imagery, combined with isotonic regression-based temporal correction. Validation results yielded a mean overall accuracy of 95.18% and Kappa coefficient of 0.90 across all epochs. GUBAD shows strong agreement with GHSL-SMOD, DEGURBA, and UN-Habitat products while providing enhanced spatial detail. Unlike global urban boundary datasets that delineate administrative-functional city extents, GUBAD explicitly targets the physically built-up fabric derived from impervious surfaces under population-density and contiguity constraints. 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 a temporally consistent, city-scale UBA dataset based on standardized UN urban agglomeration definitions, enabling reproducible monitoring of urban expansion and SDG 11.3.1 assessment globally. 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.

1 Introduction

Global urbanization continues to accelerate, with urban populations reaching approximately 4 billion in 2018 and projected to rise to 6.4 billion by 2050 (United Nations Department of Economic and Social Affairs [UN-DESA], 2018, 2025). This unprecedented growth concentrates resource consumption and environmental pressures in urban areas, necessitating precise monitoring tools for sustainable development planning. Urban population growth has driven land consumption, transforming natural and agricultural landscapes into urban built-up areas and reshaping socio-economic structures (Seto et al., 2012; Sun et al., 2020). These transformations are linked to escalated environmental pressures and urban sustainability challenges, including land degradation, increased energy demand, and urban-heat-island effects (Shahfahad et al., 2022).

To monitor sustainable urban growth, the United Nations (UN) introduced SDG 11 (“Sustainable Cities and Communities"), with indicator 11.3.1 tracking the ratio of the Land Consumption Rate to Population Growth Rate (LCRPGR) to assess coordination

https://doi.org/10.5194/essd-2026-381
Preprint. Discussion started: 16 September 2026

between urban expansion and demographic growth (Estoque et al., 2021; UN-Habitat, 2020). Accurate delineation of UBA provides the spatial foundation for calculating this indicator.
However, UBA delineation remains challenging due to heterogeneous national definitions, ranging from administrative boundaries to population thresholds, that compromise cross-regional comparability (Dijkstra and Poelman, 2014). To improve consistency, international frameworks increasingly emphasize spatially explicit definitions of urban areas. The UN 'urban agglomeration' definition (contiguous high-density settlements with strong socio-economic ties to an urban core) provides a standardized spatial framework for global comparison (UN-DESA, 2018).

Advances in Earth observation technologies have made satellite remote sensing the primary approach for mapping UBA at regional and global scales. Multi-sensor observations from optical, radar, and nighttime light satellites enable the detection of urban impervious surfaces and monitoring of urban expansion over time. Optical sensors such as Landsat, Sentinel-2 and MODIS provide long-term multispectral data for urban land-cover mapping (Liu et al., 2020; Wang et al., 2023; Zhang et al., 2020). Radar sensors, including Sentinel-1 SAR, provide complementary structural information and allow urban mapping under cloud-covered conditions (Cao et al., 2018; Sun et al., 2022). Nighttime light data from DMSP-OLS and VIIRS have also been widely used to delineate urban extents based on human activity patterns (Shi et al., 2023; Zhao et al., 2022). These multi-source observations have supported the development of several global urban datasets, including the Global Human Settlement Layer (GHSL-SMOD R2023A) (Schiavina et al., 2023), Global Artificial Impervious Area (GAIA) (Li et al., 2020), and other Landsat-based global urban land products (Liu et al., 2020; Zhang et al., 2020). Further advancing these efforts, the High-resolution Global Impervious Surface Area (Hi-GISA) product integrates Sentinel-1 and Sentinel-2 data to generate 10 m resolution global impervious surface maps (Sun et al., 2025). This synergistic approach helps overcome challenges like cloud cover in low-latitude regions and improves the extraction of mixed pixels in low-density areas (Sun et al., 2025).

Existing global urban datasets exhibit three critical limitations: (1) inconsistent urban definitions (ranging from morphological (GHSL) to administrative (DEGURBA) approaches) that limit cross-product comparability; (2) insufficient representation of medium-sized cities, which will drive much of future urban growth (Mahtta et al., 2022); and (3) temporal inconsistency arising from varying update cycles and classification model changes (Wang et al., 2023).

While GHSL-SMOD R2023A provides global settlement layers at 1 km resolution with temporal coverage (1975–2030), and DEGURBA offers harmonized urban definitions, no existing dataset applies UN urban agglomeration definitions at 30 m city scale with temporally consistent classification across more than 1,600 cities. This gap motivates the development of GUBAD. It is important to distinguish GUBAD from complementary global urban datasets such as Global Urban Boundaries dataset (GUB) (Li et al., 2020): GUB delineates administrative-functional boundaries based on the GAIA data, effectively mapping the outer envelope of cities, whereas GUBAD targets the UBA—the contiguous, population-supportive core. Consequently, boundary-oriented products may include non-built-up hinterlands, compromising the land consumption rate (LCR) calculation required for the SDG 11.3.1. Furthermore, while UN-Habitat and the World Bank provide official statistics for only a subset of cities, GUBAD offers complete temporal coverage to 1,611 cities, addressing a major completeness gap in global sustainable development monitoring. GUBAD resolves this conceptual inconsistency by separating ISA extraction from UBA delineation: ISA captures the physical urban fabric, and UBA derivation applies population-density and contiguity constraints to translate that fabric into functionally defined urban areas aligned with UN urban agglomeration definitions.

This study presents the Global Urban Built-up Area Dataset (GUBAD), providing 30 m resolution UBA maps for 1,611 cities with populations exceeding 300,000 across six epochs (2000, 2005, 2010, 2015, 2020, and 2025). Using Google Earth Engine, we implement a Random Forest classification for Landsat and Sentinel imagery to extract impervious surfaces, subsequently converting these to UBA through a spatial morphological framework constrained by UN urban agglomeration definitions.

https://doi.org/10.5194/essd-2026-381
Preprint. Discussion started: 16 September 2026

Our objectives are: (i) develop a reproducible framework for ISA extraction and UBA delineation using standardized urban definitions; (ii) validate GUBAD against other established global urban products, including GHSL-SMOD, the Atlas of Urban Expansion, and the Global Urban Boundaries dataset; and (iii) demonstrate its applicability through computation of SDG indicator 11.3.1, assessing the relationship between land consumption and population growth. GUBAD advances sustainable urban development research by providing a temporally consistent, definitionally standardized UBA dataset at the city scale.

The study period starts in 2000 for four reasons: (i) 2000 is the earliest year for which global 30 m satellite coverage is sufficiently dense and consistent: from 1999 onwards the Landsat 7 ETM+ sensor complements the Landsat 5 TM archive, whereas earlier global coverage is spatially and temporally heterogeneous because much of the pre-1999 archive relied on direct downlinks to regional receiving stations. (ii) the ancillary datasets used to delineate and constrain urban extent are anchored at 2000: LandScan global population grids are produced annually from 2000 onwards, and the population and built-up epochs of GHSL and GRUMP are defined at 2000, so that 2000 is the earliest year in which globally consistent population constraints can be applied. (iii) 2000 is a benchmark epoch of all four reference products used for validation (GUB, GHSL-SMOD, MGUP, and AUE), which enables direct intercomparison from the very start of the record. (iv) a 2000-2025 record at five-year intervals spans the full Millennium Development Goal era (2000-2015) and the SDG era up to the present, providing a 15-year baseline preceding SDG indicator 11.3.1 reporting and aligning the product with UN reporting cycles.

2 Study areas and datasets

2.1 Study areas

The study focuses on 1,611 cities worldwide with populations exceeding 300,000 inhabitants in 2020, distributed across all inhabited continents (Fig. 1). The complete list with identifiers, coordinates, and population information is presented in Supplementary Table S1. The city list is derived from the United Nations World Urbanization Prospects 2025 revision (United Nations Department of Economic and Social Affairs, 2025) and cross-referenced with Atlas of Urban Expansion (Angel et al., 2011) to achieve consistency in city identification and spatial representation. The global coverage serves as a representative basis to analyze impervious surface dynamics and the patterns of urban expansion in various urban environments.

Cities are categorized based on their population size following conventions (Angel, 2012; Liu et al., 2020; Schiavina et al., 2023; UN-Habitat, 2020) to facilitate the cross-scale analysis of urban form and LUE (Table 1). This classification facilitates cross-regional comparisons of urban form and offers a consistent framework to analyze the land-use efficiency under the baseline framework SDG 11.3.1.

Table 1. City size classification based on population, with typical morphological descriptions and example cities.

Category	Population (millions)	Typical morphological description	Example Cities
Small City	< 0.5	Compact, single-core urban footprint; low-rise density; regional service center	Bern (Switzerland), Heidelberg (Germany), Salt Lake City (USA)
Medium City	0.5 – 1	Moderate metropolitan area with defined suburban periphery; developing transit systems	Lisbon (Portugal), Marseille (France), Vancouver (Canada)
Large City	1 – 5	Major metropolitan or polycentric region; international connectivity; specialized economy	Madrid (Spain), Munich (Germany), Boston (USA)
Very Large City	5 – 10	Extensive continuous built-up zone across multiple jurisdictions; heavy transit dependence	London (UK), Riyadh (Saudi Arabia), Los Angeles (USA)

https://doi.org/10.5194/essd-2026-381
Preprint. Discussion started: 16 September 2026

Open Access Earth System Science Data Discussions

Megacities	≥ 10	High complex multi-core urban systems with global economic influence	Cairo (Egypt), Shanghai (China), Sao Paulo (Brazil)

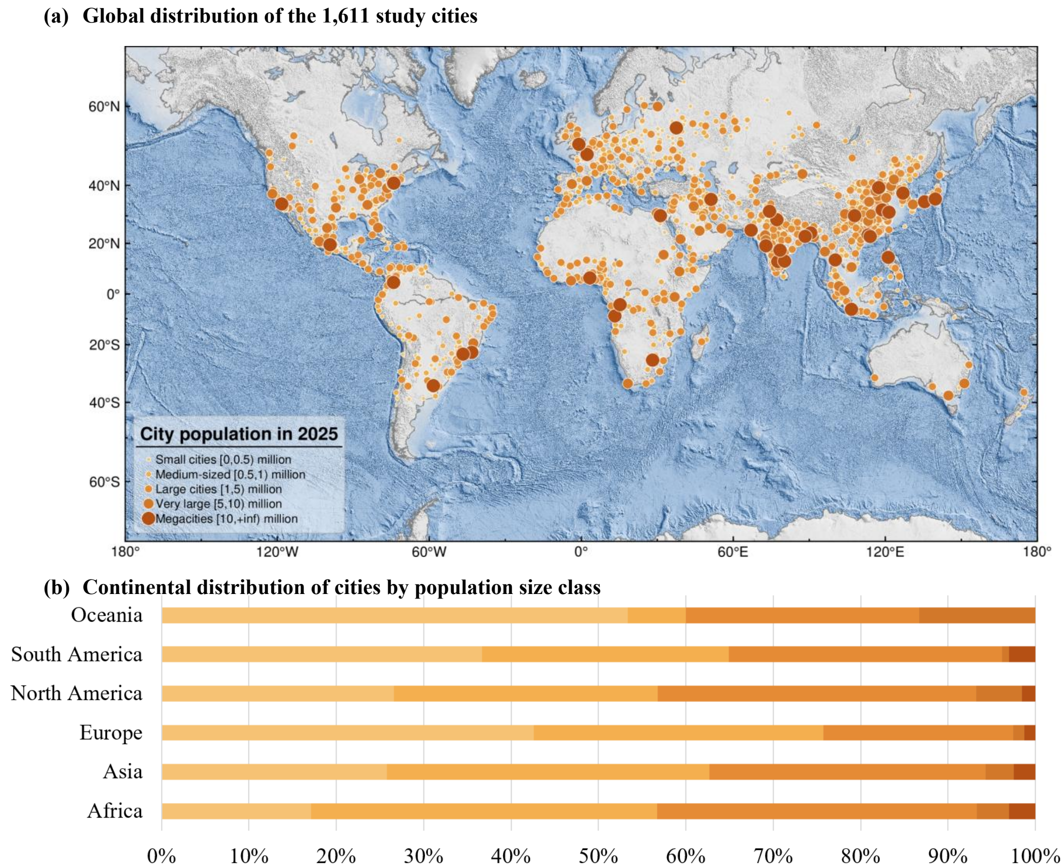

Figure 1. (a) Spatial distribution of the 1,611 study cities with populations exceeding 300,000. Point size and color indicate the city population class in 2025. (b) Composition of the study cities by continent and population size class; bar labels give the number of cities in each continent and its share of the global total

2.2 Data sources

UBA datasets were generated for six epochs (2000, 2005, 2010, 2015, 2020, and 2025). Population data are obtained from UN World Urbanization Prospects 2025 database (WUP) (https://population.un.org/wup/). An initial set of 1,860 cities with population exceeding 300,000 is identified during the time span of 1950–2035. Cities with incomplete spatial coverage and missing population records were filtered out, resulting in 1,611 cities with consistent temporal coverage for 2000–2025.

All analyses are performed on the Google Earth Engine (GEE). The main datasets used for ISA extraction include Landsat 5/7/8 TM/ETM+ (2000–2015), Sentinel-1 synthetic aperture radar (SAR) and Sentinel-2 multispectral imagery (MSI) (2020–2025). Specifically, long-term historical observations are available from Landsat, complemented by the Sentinel-1 and Sentinel-2 datasets that offer simultaneous radar and optical observations to improve the detection accuracy for recent years, especially those regions affected by cloud coverage and complicated urban morphology.

https://doi.org/10.5194/essd-2026-381
Preprint. Discussion started: 16 September 2026

Open Access Earth System Science Data Discussions

The optical images are pre-processed by standard procedures including radiometric calibration, atmospheric correction, and cloud-shadow masking. Terrain normalization is conducted by Copernicus DEM GLO-30 (European Space Agency, 2021). Copernicus Global Land Cover product (CGLS-LC100 Collection 3) is added to support urban-non-urban classification as an auxiliary layer. The detailed pre-processing and classification procedures follow the approaches of Jiang et al. (2021) and Sun et al. (2022).

Administrative and ancillary spatial boundaries are collected from the Dataset of Global Administrative Areas (GADMv3.6) and the Global Rural-Urban Mapping Project (GRUMPv1). These datasets are harmonized with the WUP population database to guarantee consistent city identification from different epochs. The integration of the datasets provides a consistent spatial framework for deriving the UBA from mapped ISA and offers cross-country analysis for the requirement of SDG indicator 11.3.1.

Table 2. List of datasets used for ISA extraction, UBA delineation, and validation.

Type of Data	Data Classification	Spatial Resolution	Data Sources	Time Span	Version/Collection
Administrative areas	Database of Global Administrative Areas (GADMv3.6)	Vector (polygons)	https://www.gadm.org/	2018 (static snapshot)	v3.6
Copernicus Global Land Cover	CGLS-LC100 Collection 3	100 m	https://land.copernicus.eu/global/use-cases	2015–2019	Collection 3
Landsat data	Landsat 5 TM, Landsat 7 ETM+, Landsat 8 OLI	30 m (native)	https://earthexplorer.usgs.gov/	2000–2015	Collection 2 Level-2 (surface reflectance)
Population data	UN World Urbanization Prospects	City-level (tabular)	https://population.un.org/wup/	1950–2035 (projections); 2000–2025 (used)	2024 Revision
Sentinel-1 SAR	GRD VV polarization	10 m (native)	https://dataspace.copernicus.eu/	2020–2025	Level-1 GRD
Sentinel-2 MSI	Surface reflectance (L2A)	10/20/60 m (native)	https://dataspace.copernicus.eu/	2020–2025	Collection-1 Level-2A
Nighttime lights	VIIRS Day/Night Band	375 m	https://eogdata.mines.edu/products/vnl/	2012–present (used: 2020–2025)	v2.1
Digital elevation model	Copernicus DEM GLO-30	30 m	https://spacedata.copernicus.eu/	2021 (static)	GLO-30
Urban extent polygons	Global Rural–Urban Mapping Project (GRUMPv1)	1 km	https://sedac.ciesin.columbia.edu/data/collection/grump-v1	2000 (static)	v1
Built-up surface	GHSL-SMOD R2023A	100 m	https://ghsl.jrc.ec.europa.eu/	1975–2030 (used: 2000–2025)	R2023A

https://doi.org/10.5194/essd-2026-381
Preprint. Discussion started: 16 September 2026

Land cover	ESA WorldCover	10 m	https://worldcover.esa.int/	2020 (static)	v100
Population distribution	LandScan Global	1 km	https://landscan.ornl.gov/	2000–2025 (annual)	2023 Release

3. Methodology

GUBAD employs a two-stage framework (Fig. 2): (1) multi-source ISA extraction (2000–2025), and (2) spatial-morphological conversion to UBA using standardized UN urban agglomeration constraints. Consecutive steps are detailed below.

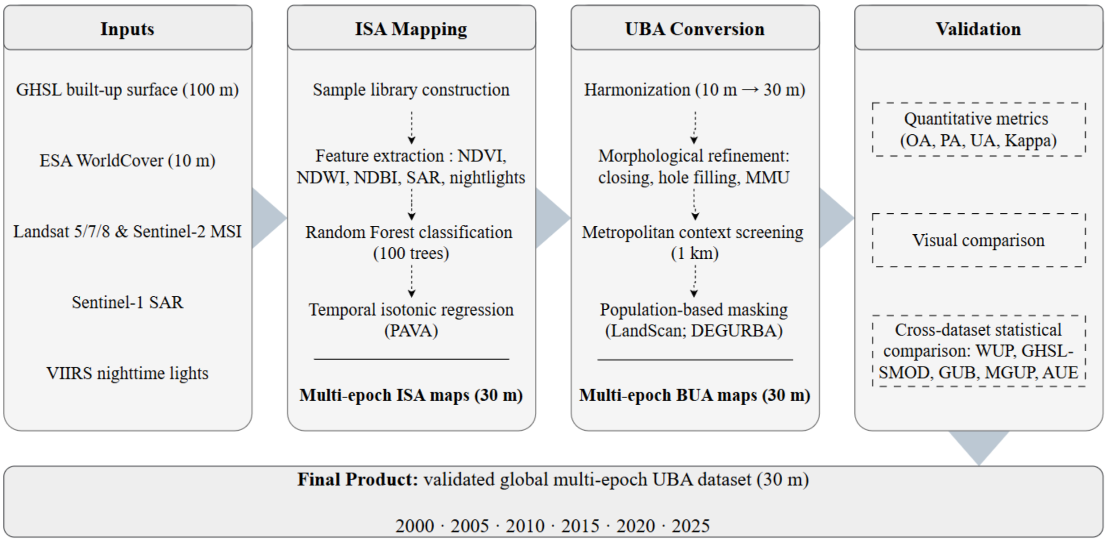

Figure 2. GUBAD production workflow integrating multi-source remote sensing data and spatial-morphological conversion to UBA.

3.1 Impervious surface area extraction

ISA extraction from 2000 to 2025 was conducted through a systematic framework comprising four main stages (Fig. 3): sample label library construction, training dataset generation, Random Forest (RF) classification, and temporal isotonic regression.

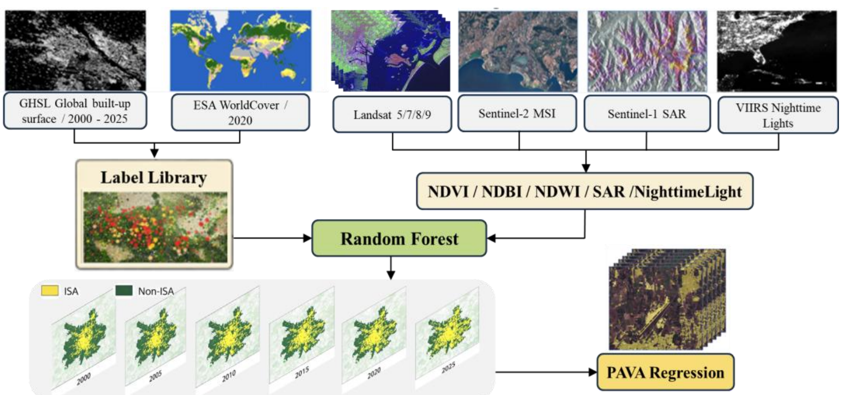

Figure 3. Four-stage methodological framework for ISA extraction: (1) sample label library construction, (2) training dataset generation, (3) Random Forest classification, and (4) temporal isotonic regression.

https://doi.org/10.5194/essd-2026-381
Preprint. Discussion started: 16 September 2026

Sensor and feature configuration varied by epoch to balance historical coverage with data availability (Table 3). For 2000–2015, Landsat 5/7/8 TM/ETM+ optical imagery provided NDVI, NDWI, and NDBI spectral indices. For 2020–2025, Sentinel-2 MSI supplied optical indices at 10 m native resolution, supplemented by Sentinel-1 SAR VV backscatter for structural information under cloud-covered conditions, and VIIRS nighttime lights for human activity patterns. DMSP-OLS nighttime lights were not used in this study due to coarse spatial resolution (2.7 km) and calibration inconsistencies across sensors, which would compromise cross-temporal comparability; VIIRS (375 m resolution, 2012-present) was selected as the sole nighttime light source for 2020–2025 epochs.

Table 3. Multi-sensor configuration and input features for ISA extraction by epoch (2000–2025).

Epoch	Optical Sensor	SAR Sensor	Nighttime Light	Spectral Indices	Reference Labels
2000	Landsat 5/7 TM	—	—	NDVI, NDWI, NDBI	GHSL-Global built-up surface
2005	Landsat 5/7 TM/ETM+	—	—	NDVI, NDWI, NDBI	GHSL-Global built-up surface
2010	Landsat 5/7/8 TM/ETM+/OLI	—	—	NDVI, NDWI, NDBI	GHSL-Global built-up surface
2015	Landsat 7/8 ETM+/OLI	—	—	NDVI, NDWI, NDBI	GHSL-Global built-up surface
2020	Sentinel-2 MSI (10 m)	Sentinel-1 (VV)	VIIRS	NDVI, NDWI, NDBI, VV, Lights	GHSL-Global built-up surface + ESA WorldCover
2025	Sentinel-2 MSI (10 m)	Sentinel-1 (VV)	VIIRS	NDVI, NDWI, NDBI, VV, Lights	GHSL-Global built-up surface

Sample labels were generated for each city by leveraging GHSL-Global built-up surface (100 m): observed epochs 2000, 2005, 2010, 2015, 2020; the 2025 epoch was derived from linear projection of 2020 trends validated against historical growth rates. ESA WorldCover v100 (10 m, 2020) provided complementary high-resolution training labels. High-confidence positive (impervious) and negative (pervious) sampling locations were extracted from these reference products, with epoch-specific label sources: GHSL-Global built-up surface alone for 2000–2015 and 2025; GHSL-Global built-up surface cross-validated with ESA WorldCover for 2020 to improve spatial precision. Each city was assigned a standardized set of 1,000 sample labels (250 positive, 750 negative) to ensure balanced spatial representation. For cities with complex morphology ($n \approx 150$, 9% of the sample), manual visual interpretation improved label quality. Experiments confirmed that sample sizes beyond 1,000 per city yielded marginal accuracy gains relative to computational cost (Sun et al., 2022).

From this sample library, the training dataset was generated by extracting multi-source remote sensing features, with temporal consistency maintained by epoch-specific feature computation. The feature set included three spectral indices derived from Landsat / Sentinel-2 MSI data: the Normalized Difference Building Index (NDBI), Normalized Difference Vegetation Index (NDVI), and Normalized Difference Water Index (NDWI). NDBI enhances built-up detection, NDVI suppresses vegetation, and NDWI masks

https://doi.org/10.5194/essd-2026-381
Preprint. Discussion started: 16 September 2026

water bodies (Jiang et al., 2021). SAR backscatter distinguishes urban structure, while VIIRS nighttime lights capture human activity patterns. This multi-source feature set captured impervious and pervious characteristics across all target cities.

ISA classification employed Random Forest ensembles (100 trees) implemented on Google Earth Engine. Native 10 m Sentinel-2 data (2020–2025) were harmonized to 30 m grid cells via area-weighted aggregation (threshold: 0.5) for consistency with Landsat epochs (2000–2015). To suppress spurious temporal fluctuations and ensure logical consistency, Pool-Adjacent-Violators Algorithm (PAVA) isotonic regression was applied to ISA time series. This corrects anomalous temporal 'spikes' while preserving the generally non-decreasing trend consistent with global urbanization patterns (Ayer et al., 1955; Barlow, 1980). Applied at the pixel level, this constraint primarily suppresses noise-induced inconsistencies rather than enforcing artificial growth, acknowledging that localized urban shrinkage is typically limited in spatial extent relative to overall expansion.

3.2 Conversion from impervious surface area to urban built-up area

In this study, urban built-up area (UBA) is defined as a spatially contiguous extent of impervious surfaces that is further constrained by population density and settlement continuity (Fig. 4). This definition represents a hybrid approach that combines morphological characteristics (physical built-up structure) with functional relevance (population presence), aligning with the concept of urban agglomerations used by the United Nations.

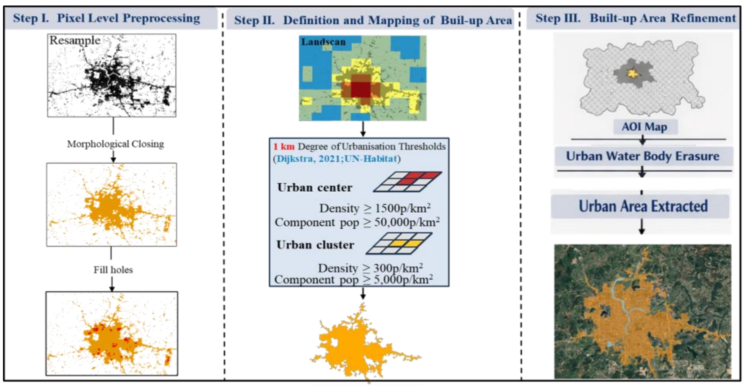

Figure 4. Spatial-morphological conversion workflow from ISA to UBA, including resolution harmonization, morphological refinement, metropolitan context screening, and population-constrained masking.

UBA were derived from the extracted ISA maps using a globally standardized workflow consistent with the urban agglomeration concepts adopted in the World Urbanization Prospects and subsequent UN-Habitat guidelines (UN-DESA, 2018, 2025; UN-Habitat, 2020). AOIs were defined using GADM v3.6 administrative units for global cities; Chinese cities employed officially designated urban core boundaries. To ensure cross-temporal comparability, native 10 m ISA (2020–2025) were harmonized to 30 m grid cells via area-weighted aggregation (threshold 0.5), generating binary impervious-surface masks consistent with 30 m Landsat epochs (2000–2015). This 0.5 threshold follows common practice in binary aggregation of high-resolution land-cover data, representing a balance between omission and commission errors.

Within each AOI, the region of interest was defined as the intersection between the administrative boundary and the buffered envelope of non-zero 30 m ISA, and all operations were implemented in EPSG:4326 (WGS 84, geographic coordinate system) for global consistency, with metric calculations performed on UTM-projected raster (local zone, WGS 84 datum). Area measurements were validated against Mollweide equal-area projection (EPSG:54009) to ensure <2% areal distortion across latitudes. The binary ISA masks were first refined using morphological operations to preserve local settlement continuity while suppressing noise. Specifically, a circular morphological closing kernel with an approximate ground radius of 45 m was applied to seal narrow fissures and intra-block gaps. Interior holes were filled only when they were fully enclosed by built-up pixels, did not touch the ROI

https://doi.org/10.5194/essd-2026-381
Preprint. Discussion started: 16 September 2026

boundary, and were smaller than a conservative area cap of 1,024 pixels at 30 m resolution, equivalent to approximately 0.92 km². To prevent artificial expansion along the boundary of the processing window, holes intersecting a 90 m boundary-protection ring were excluded from filling. Connected components were then labelled using an 8-neighbour rule, and isolated patches smaller than 57,600 m², corresponding to 64 pixels at 30 m resolution, were removed as below the minimum mapping unit.

To further distinguish genuine urban built-up land from scattered non-urban impervious patches, a metropolitan context mask was derived from the same ISA data on a 1 km lattice. The 30 m ISA was aggregated to 1 km mean ISA intensity, and cells with mean ISA ≥0.18 were used to define the broader metropolitan envelope, while cells with mean ISA ≥0.50 were treated as high-intensity built-up cores. Instead of clipping the 30 m built-up patches directly by the 1 km grid, the workflow used a natural-edge keep rule: for binary raster B (30 m ISA) and mask M (1 km context), output O = {connected components $c \in B \mid c \cap M \neq \emptyset$ }. This preserves component perimeter geometry without resampling artifacts.

A population-constrained urban-center mask was further constructed using LandScan population data on the same 1 km grid. Population density thresholds (≥300 persons km^{-2} for urban cells; ≥5,000 persons per contiguous component) follow the DEGURBA urban-center criteria (Dijkstra et al., 2021) to ensure alignment with internationally standardized urban definitions. Contiguity was evaluated using an 8-neighbour rule. To maintain settlement continuity, ISA-based bridging was then applied by including 1 km cells with mean ISA ≥0.50; when no such bridging cells were available, a fallback threshold of 0.25 was used. The resulting urban-center mask was dilated by one 1 km cell to avoid small pinholes and was then used through the same natural-edge keep rule to retain intersecting 30 m ISA components. This population-constrained step prevented isolated impervious facilities, such as transport platforms, airports, and logistics yards, from being promoted to UBA unless they were embedded within a population-supported settlement structure. Final UBA masks were vectorized, topologically corrected, and subjected to the same minimum-area filtering across all cities and epochs. All intermediate rasters, including binary ISA, closed ISA, filled ISA, 1 km ISA context masks, population masks, and final diagnostic layers, were archived to support auditability and reproducibility.

3.3 Accuracy validation

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

Four standard accuracy metrics were computed, which comprised the overall accuracy (OA), producer's accuracy (PA), and user's accuracy (UA) along with the Kappa coefficient. These metrics were obtained from confusion matrices and are widely accessible in land cover and urban mapping validation studies (Foody, 2010; Sun et al., 2019).

Validation was performed using stratified random sampling (n=10 per city: 5 UBA, 5 non-UBA) following Foody (2010) and Sun et al. (2019). The per-city accuracy estimations have 95% confidence intervals of ±12–18% (Clopper-Pearson exact method), however, the aggregate accuracy across 1,611 cities is precise to ±0.5% due to the large sample size. The reference labels were acquired from visual interpretation of concurrent high-resolution satellite imagery where each individual point was manually examined to determine UBA/non-UBA classification. Furthermore, the comparison of classified results obtained with reference labels led to confusion matrices for each respective year along with the OA, PA, UA and Kappa statistics to evaluate the scale of classification performance. While the per-city sample size is limited, the large number of cities (n=1,611) allows the robust global accuracy estimation.

The visual comparison analysis was carried out by comparison with various widely used global built-up area datasets. Representative cities from diverse geographic regions were chosen to be visually compared in terms of spatial patterns against a series of existing reference datasets: Global Urban Boundaries dataset (GUB), Degree of Urbanization (DEGURBA), GHS-SMOD settlement layer,

https://doi.org/10.5194/essd-2026-381
Preprint. Discussion started: 16 September 2026

220 Atlas of Urban Expansion (AUE), and MGUP urban populations dataset. Consistencies in the agreement of the urban extents and morphological characteristics were assessed using spatial overlays and high-resolution satellite imagery. This approach is standard in global urban mapping where plausibility of new spatially constrained urban areas and comparison of geographic extent boundary consistency of the newly generated built-up datasets is evaluated (Schiavina et al., 2023; Wang et al., 2023).

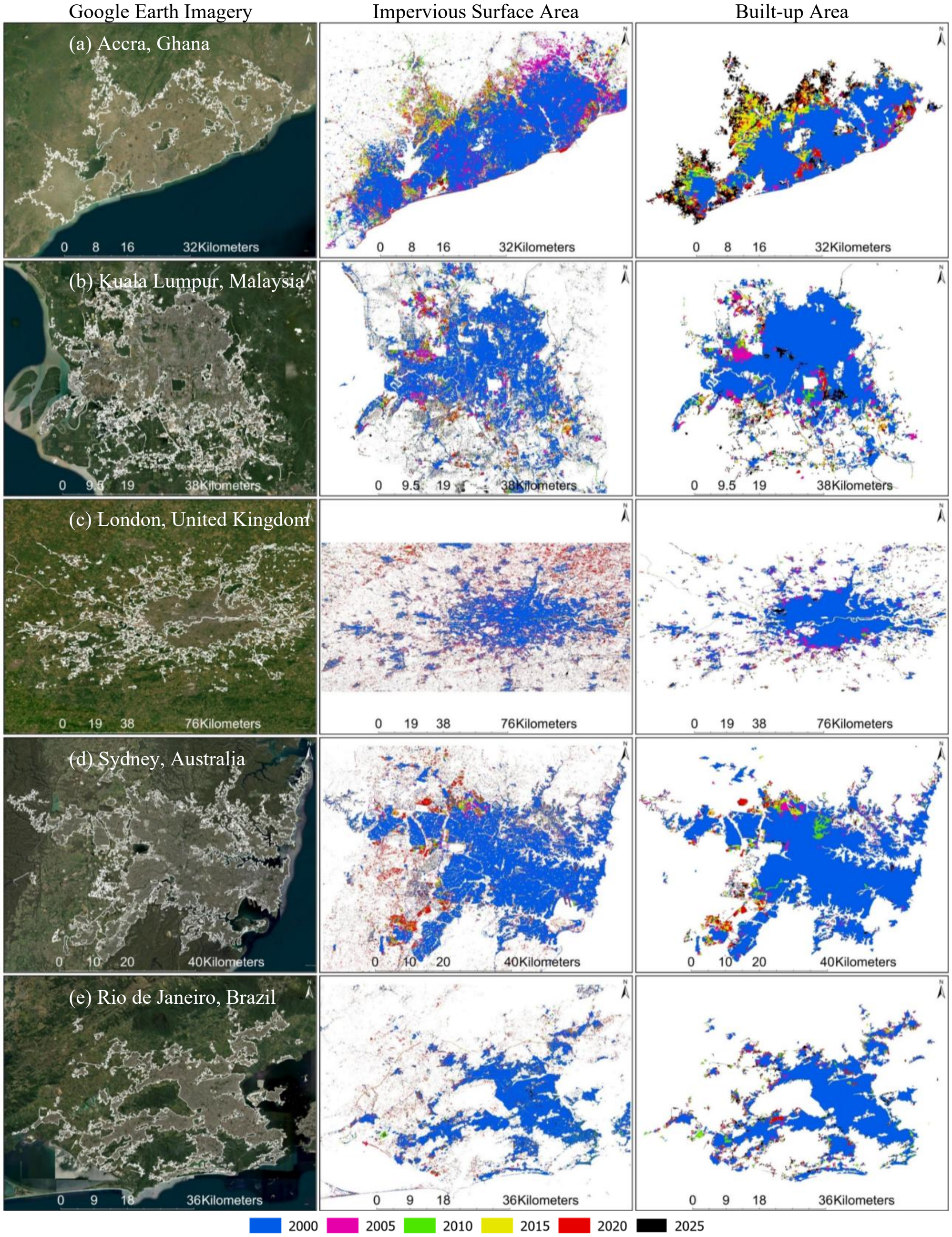

Figure 5. Visual comparison of ISA and UBA spatial expansion against Google Earth reference imagery for five representative cities, 2000–2025: (a) Accra, (b) Kuala Lumpur, (c) London, (d) Sydney, and (e) Rio de Janeiro. White boundaries denote GUBAD-derived UBA extents. Map data © Google. Imagery © Google, Maxar Technologies (2000-2025).

https://doi.org/10.5194/essd-2026-381
Preprint. Discussion started: 16 September 2026

Scatterplot-based statistical comparison was performed to examine quantitative agreement between the derived UBA areas and existing global urban datasets. For each city and temporal epoch, built-up area estimates were compared with UN-Habitat, GUB, and GHS-SMOD datasets. Scatterplots examined relationships between datasets, identifying correlations and potential systematic differences, then compared these datasets to evaluate the relationship between them and identify correlations as possible systematic differences that may exist. Scatterplot analysis is standard in large-scale urban mapping for evaluating consistency and reliability of urban extent datasets from different sources and methodologies (Dijkstra et al., 2021; Sun et al., 2019). Further, the validation approach is advantageous as it addresses global coverage first, rather than dense local sampling, in terms of having a consistent global evaluation; further studies can be constructed for higher-density validation datasets for selected regions to further evaluate local accuracy assessments.

4 Results

4.1 Validation of ISA

The classification results from the ISA products are shown in Fig. 6 and Table 4 below. Highly consistent overall classification performance is indicated as a tight spatial agreement between reference data and classification with respect to overall classification accuracies.

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. These values indicate a strongly performing classification framework which had balanced omission or commission errors. In the individual years, the OA had a narrow range from 94.62%–96.57% with consistent performance of the overall used model, which was stable across time. Similarly, the PA for the impervious class is above 90% throughout all years indicating low omission error. The UA is above 88% indicating a small commission error. The inter-dataset Kappa coefficients indicate a substantial agreement, being higher than chance (0.89–0.93). This consistency in agreement shows a robustness to varying levels of input data sources, especially due to the transition from Landsat-based to Sentinel-based observations that occur in later epochs.

Table 4. ISA classification accuracy metrics by epoch (2000–2025), computed from the stratified random sampling (n = 10 per city, 1,611 cities).

	Predicted Values							
Actual Values	2000	Impervious	Pervious	UA	2015	Impervious	Pervious	UA
	Impervious	7217	388	94.90%	Impervious	7237	368	95.16%
	Pervious	412	7193	94.58%	Pervious	357	7248	95.31%
	PA	94.60%	94.88%		PA	95.30%	95.17%	
	OA	94.74%	Kappa	0.89	OA	95.23%	Kappa	0.90
	2005	Impervious	Pervious	UA	2020	Impervious	Pervious	UA
	Impervious	7220	385	94.94%	Impervious	7213	392	94.85%
	Pervious	361	7244	95.25%	Pervious	397	7208	94.78%
	PA	95.24%	94.95%		PA	94.78%	94.84%	
	OA	95.10%	Kappa	0.90	OA	94.81%	Kappa	0.90

https://doi.org/10.5194/essd-2026-381
Preprint. Discussion started: 16 September 2026

2010	Impervious	Pervious	UA	2025	Impervious	Pervious	UA
Impervious	7203	402	94.71%	Impervious	7286	374	95.12%
Pervious	416	7189	94.53%	Pervious	151	7509	98.03%
PA	94.54%	94.70%		PA	97.97%	95.26%	
OA	94.62%	Kappa	0.89	OA	96.57%	Kappa	0.93

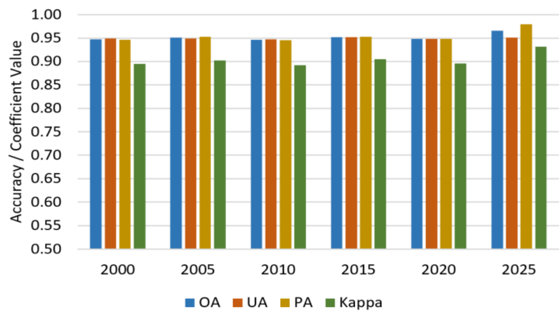

Figure 6. ISA classification accuracy by epoch (2000–2025), derived from stratified random sampling (n = 10 per city, 1,611 cities).

4.2 The UBA dataset and validation

The GUBAD provides multitemporal UBA information on 1,611 cities (six epochs from 2000–2025) combining raster-based impervious surface maps with vector-based urban built-up area polygons to allow consistent analysis at varying spatial scales.

4.2.1 Visual comparison

Visual inspection of representative cities confirms GUBAD is able to represent urban spatial patterns with high detail and coherence (Fig. 7) such as:

- Compact/Morphologically structured: Paris (France), Shanghai (China)— subtle fine-scale features that are well-preserved with no excessive smoothening.
- Rapidly expanding/Heterogeneous: Lagos (Nigeria), Kolkata (India)— dispersed and irregular growth patterns are effectively expressed.
- Coastal/Polycentric: New York (USA), São Paulo (Brazil)—minimal incorporation of non-urban land or water-bodies.

There is strong spatial agreement between the GUBAD-derived UBA with regards to observed urban extents in diverse geographic contexts through these representative cities. These spatial patterns show the GUBAD's ability to express a diversity of urban morphologies, from compact historic cores, to fragmented rapid-growth peripheries in different climate and developmental conditions.

https://doi.org/10.5194/essd-2026-381
Preprint. Discussion started: 16 September 2026

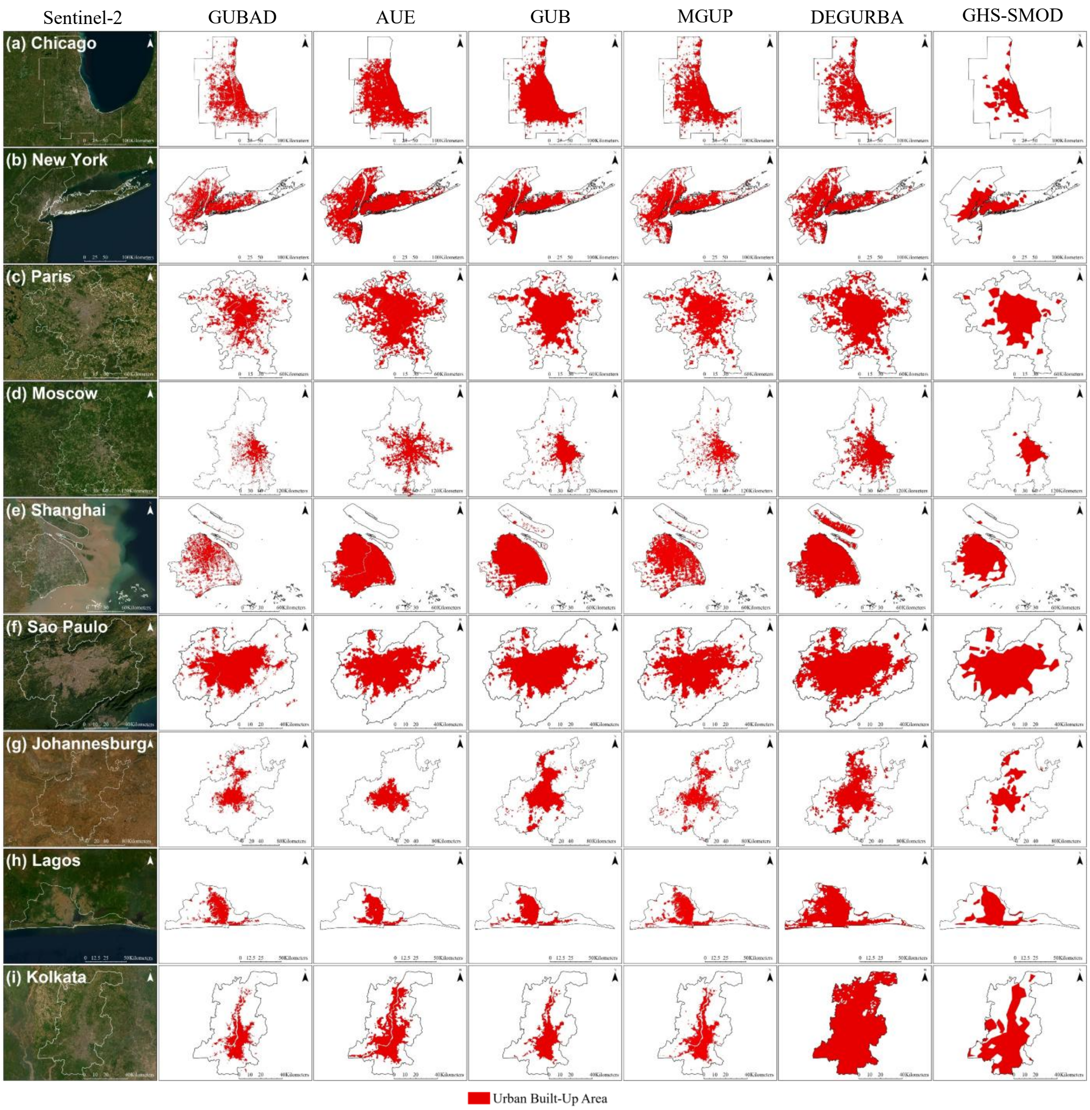

Figure 7. Visual comparison of GUBAD UBA against reference datasets for nine representative cities: Chicago, New York, Paris, Moscow, Shanghai, São Paulo, Johannesburg, Lagos, and Kolkata. Contains modified GHS-SMOD data © European Commission, Joint Research Centre, licensed under CC-BY 4.0. Contains modified Copernicus Sentinel data (2025).

4.2.2 Comparison with statistical datasets

The derived UBA estimates were evaluated against the existing UN-Habitat urban area statistics (World Urbanization Prospects, WUP) (Fig. 8). Of 1,611 cities, 1,100 cities were successfully matched after aligning spatial definitions and temporal coverage.

Scatter-plots show a clear positive linear relationship between GUBAD estimates and the reference datasets for multiple benchmark years. The majority of observations are closely around the 1:1 line demonstrating a high level of agreement in urban area estimates.

These deviations of a subset of cities are probably caused by the difference in urban definitions and their boundary delineation methodologies. Statistical datasets are generally including non-built-up areas (administrative/demographic definitions) for statistics,

https://doi.org/10.5194/essd-2026-381
Preprint. Discussion started: 16 September 2026

while GUBAD provides more conservative and physically representative estimates from impervious surfaces. Although some cities deviate, the overall consistency indicates GUBAD is presenting reliable urban extents estimates at the city level.

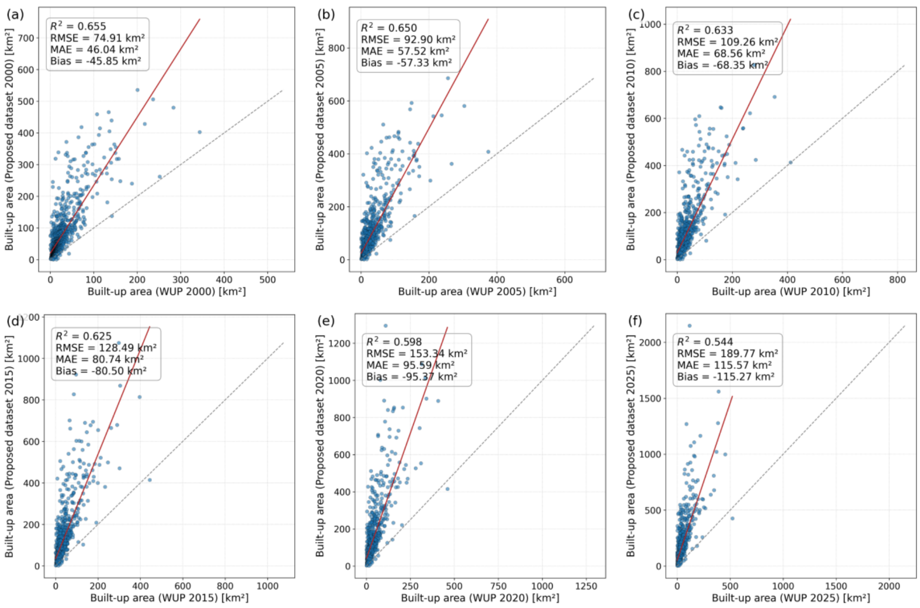

Figure 8. Validation of GUBAD UBA estimates against UN-Habitat World Urbanization Prospects (WUP) urban area statistics. Scatterplots compare GUBAD UBA (km²; y-axis) with WUP estimates (km²; x-axis) for matched cities (n = 1,100). Blue dashed line: 1:1 identity line; solid orange line: linear regression fit. Insets report R², RMSE (km²), MAE (km²), and bias (km²).

4.2.3 Comparison with remote sensing datasets

GUBAD was benchmarked against the GHSL-SMOD, GUB, MGUP ($n \approx 1{,}500$ each), and AUE ($n \approx 150$ cities) datasets multiple temporal snapshots (Fig. 9). In terms of statistical analysis there is high agreement among these datasets with R^2 values ranging from 0.855 to 0.873 for GHSL-SMOD (2000–2025) and 0.788 to 0.843 for GUB (2000–2018), showing a strong alignment and consistency with global settlement standards. A similar strong statistical correlation was assessed for MGUP (2001–2018), with R^2 values falling within a stable range of 0.742–0.768, validating the consistency of GUBAD's larger-scale urban delineations against another global impervious surface layer. Although the smaller AUE sample (2000, 2015) showed slightly lower correlation ($R^2 \approx 0.76$), point density concentrated along the 1:1 line confirms minimal systematic bias. These results confirm that GUBAD provides a statistically robust representation of built-up area dynamics, maintaining strong agreement with existing remote sensing-derived urban products. The relatively higher agreement with GHSL-SMOD compared to MGUP further reflects methodological similarities in morphology-based urban delineation, whereas population-driven datasets tend to exhibit systematic overestimation in low-density regions. To ensure comparability across products with heterogeneous urban definitions, all reference datasets were evaluated within GUBAD-derived city boundaries (AOIs). This allows the classification consistency to be separated in terms of differences in boundary delineation providing a conservative assessment on the accuracy of GUBAD's ISA-to-UBA conversion.

GUBAD produces more spatially detailed and conservative estimates than GUB or MGUP, which can potentially overestimate urban extents due to coarse resolution or population-based proxy. Conversely, GUBAD is better in representing fragmented and heterogeneous urban structures in rapidly urbanizing regions having fine-scale spatial variability as essential.

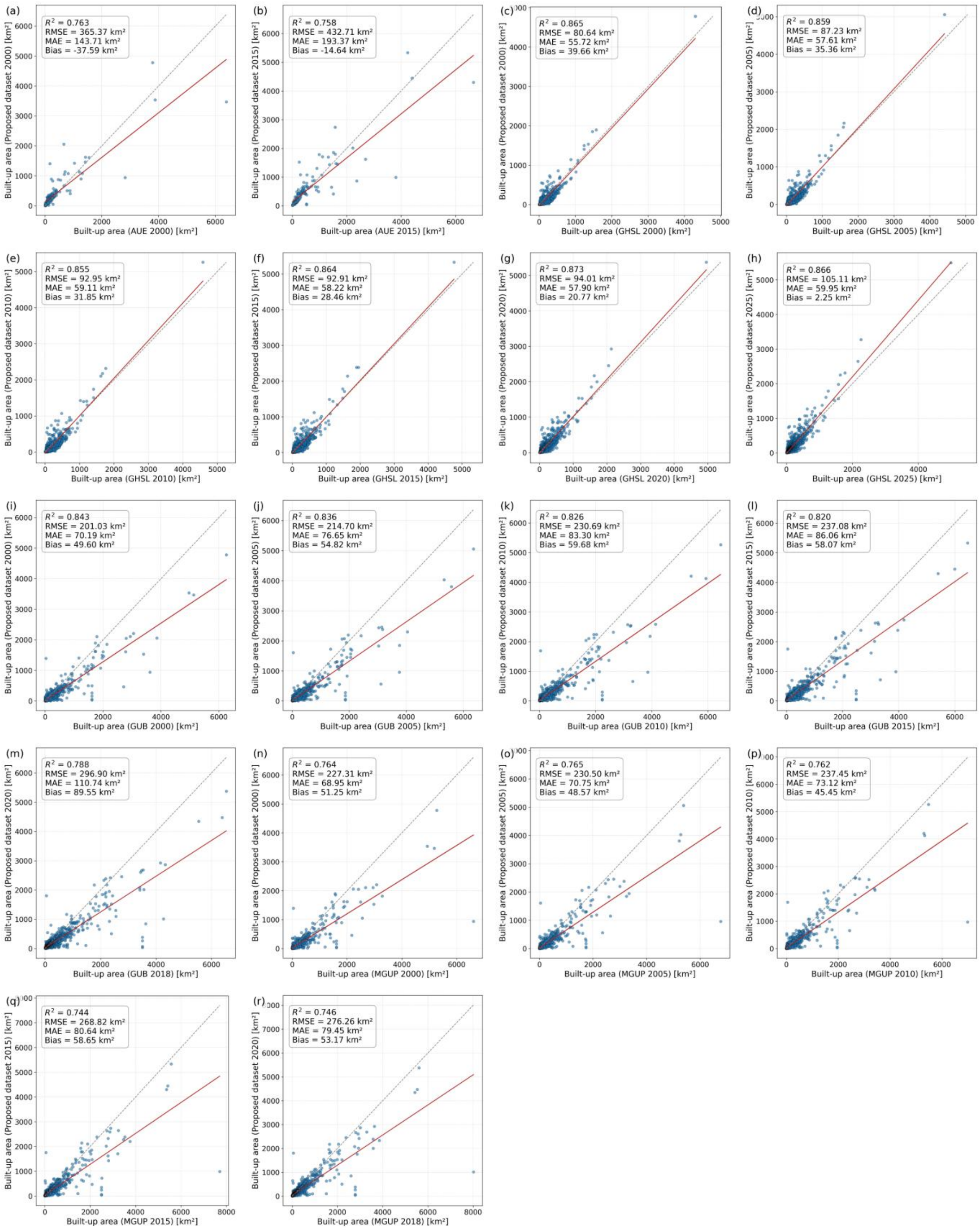

Figure 9. Cross-dataset validation of GUBAD against global remote sensing urban products. Scatterplots compare GUBAD UBA (km²; y-axis) against (a–b) AUE, (c–h) GHSL-SMOD, (i–m) GUB, and (n–r) MGUP estimates (km²; x-axis) across benchmark years. Blue dashed line: 1:1 identity line; solid orange line: linear regression fit. Insets report R², RMSE (km²), MAE (km²), and bias (km²).

https://doi.org/10.5194/essd-2026-381
Preprint. Discussion started: 16 September 2026

4.3 Change of GUBAD

4.3.1 Continent-level Changes

Global UBA expanded substantially between 2000 and 2025 (Fig. 10). Asia recorded the largest absolute increase, with UBA expanding from 64,797 km^2 to 153,165 km^2, representing an increase of over 88,000 km^2. North America experienced the second-largest growth, with UBA increasing from 69,550 km^2 to 103,296 km^2. Together, these two regions accounted for more than 70% of total global urban growth.

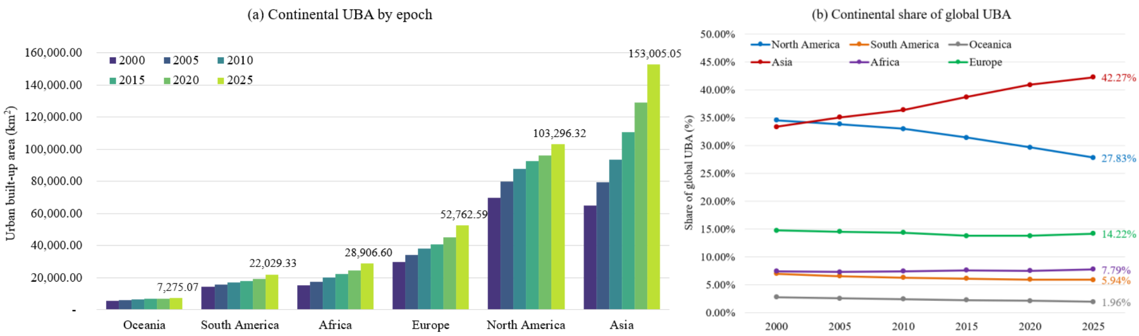

Figure 10. (a) Total UBA (km2) by continent for the six GUBAD epochs, (b) Continental share (%) of the global UBA of the study cities, 2000-2025

Although absolute increases in other continents were smaller, relative growth rates were notable. UBA in Africa approximately doubled between 2000 and 2025, while South America, Europe, and Oceania exhibited increases of around 1.5 times over the study period. By 2025, the total global UBA reached 367,818.04 km^2, nearly doubling its extent in 2000.

In terms of annual growth rates, North America exhibited the highest mean increase (2,782 km^2 yr^{-1}), followed by Asia (2,591 km^2 yr^{-1}), Europe (1,198 km^2 yr^{-1}), Africa (600 km^2 yr^{-1}), South America (566 km^2 yr^{-1}), and Oceania (227 km^2 yr^{-1}). These trends are closely aligned with patterns of urban population growth. For instance, the urban population reported by UN WUP for corresponding agglomerations within UBA in Asia increased from approximately 829 million to 1.5 billion during the study period, while North America's urban population grew from 223 million to 322 million.

In relative terms, the global share of UBA shifted notably over time. The proportional contribution of Asia increased from 32.50% to 41.60%, while Africa also experienced a slight increase. In contrast, the shares of North America, Europe, South America, and Oceania declined. These patterns highlight a clear redistribution of global urban growth, with increasing dominance of rapidly urbanizing regions.

Asia's dominance in absolute growth reflects concentrated urbanization in China and India, while North America's substantial expansion, despite slower population growth, indicates continued sprawl in lower-density metropolitan regions. The doubling of African UBA, though from a smaller base, signals emerging urban transitions that will likely accelerate through 2050.

4.3.2 Country-level changes

At the country level, UBA expansion was highly concentrated among a limited number of countries (Fig. 11). Throughout the study period (2000–2025), the United States and China consistently exhibited the largest urban extents. In 2000, the United States had the highest UBA (56,705 km²), followed by China (23,346 km²) and Japan (9,458 km²). China recorded the fastest absolute growth over the study period, with a linear mean annual increase of 2,365.3 km² yr⁻¹ calculated as $(UBA_{2025} - UBA_{2000})/25\ years$. The United

https://doi.org/10.5194/essd-2026-381
Preprint. Discussion started: 16 September 2026

States also exhibited substantial growth, increasing by 26,525 km² to reach 83,230 km² (linear mean: 1,061.02 km² yr⁻¹). Japan's UBA growth was more moderate, rising to 12,541 km² (linear mean: 123.3 km² yr⁻¹).

Among emerging economies, India, Brazil, and Mexico demonstrated steady and continuous UBA expansion. India's UBA increased from 7,925 km² in 2000 to 14,769 km² in 2025 (linear mean: 273.8 km² yr⁻¹). Brazil exhibited a comparable growth pattern, expanding from 8,212 km² to 12,050 km² (linear mean: 153.5 km² yr⁻¹). Mexico followed a similar trajectory, with UBA increasing from 7,432 km² to 11,837 km² (linear mean: 176.2 km² yr⁻¹).

In Europe, Russia and the United Kingdom showed moderate but persistent UBA growth. Russia's UBA nearly doubled from 5,602 km² in 2000 to 10,482 km² in 2025 (linear mean: 195.2 km² yr⁻¹). The United Kingdom's UBA increased from 4,678 km² to 7,148 km² over the same period (linear mean: 98.8 km² yr⁻¹). In Oceania, Australia's UBA expanded steadily from 5,102 km² to 6,606 km² (linear mean: 60.1 km² yr⁻¹). Canada recorded the slowest growth among the top ten countries, with UBA increasing from 3,836 km² to 5,683 km² (linear mean: 73.9 km² yr⁻¹). Overall, the country-level results highlighted pronounced heterogeneity in urban expansion dynamics, with rapid growth concentrated in China and the United States, steady expansion in emerging economies such as India and Brazil, and comparatively slower growth in developed regions. These patterns reflect differing stages of urbanization and regional development trajectories captured by the GUBAD dataset.

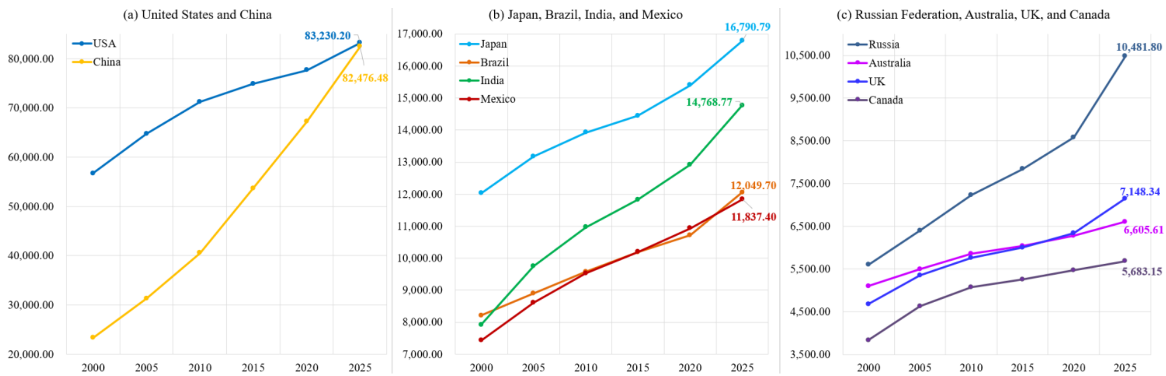

Figure 11. Urban built-up area trajectories (2000-2025) of the ten countries with the largest cumulative UBA: (a) United States and China; (b) Japan, Brazil, India, and Mexico; (c) Russian Federation, Australia, United Kingdom, and Canada

5 Discussion

5.1 The GUBAD dataset: Enabling consistent city-level assessment of SDG 11.3.1 and SDG 15.1.1

To demonstrate the applicability of GUBAD for sustainable urban assessment, the dataset was used to evaluate SDG indicator 11.3.1 for cities worldwide. City-level population data were obtained from the United Nations WUP 2018 and 2025 revisions, while urban land area was derived from the GUBAD dataset.

The calculation of LCRPGR follows the methodology recommended by the United Nations and is defined by Eq. (1)–(3). The Population Growth Rate (PGR) is calculated as:

$$\mathrm{PGR} = \frac{\mathrm{LN}(Pop_{t+n}/Pop_t)}{(y)} \quad (1)$$

where, Pop_t and Pop_{t+n} represent the total population of a city at the initial and final years, respectively, and (y) denotes the time interval (years) between the two periods.

The Land Consumption Rate (LCR) is calculated as:

https://doi.org/10.5194/essd-2026-381
Preprint. Discussion started: 16 September 2026

$$\text{LCR} = \frac{V_{present} - V_{past}}{V_{past}} \times \frac{1}{t} \tag{2}$$

where, $V_{present}$ and V_{past} represented the total UBA of a city at the initial and final years, respectively, measured in square kilometers. t represents the time interval (years) between the two periods.

The SDG 11.3.1 indicator is then expressed as:

$$\text{LCRPGR} = \frac{\text{LCR}}{\text{PGR}} \tag{3}$$

Category thresholds follow UN-Habitat (2020) recommended discretization for SDG 11.3.1 reporting. With LCR > 0, cities were classified into five categories:

Table 5. SDG 11.3.1 (LCRPGR) classification categories following UN-Habitat (2020) discretization. Categories apply to cities with positive land consumption rate (LCR > 0).

Category	Definition
$\text{LCRPGR} \leq -1$	population decline exceeds land consumption
$-1 < \text{LCRPGR} \leq 0$	land consumption continues despite population decline
$0 < \text{LCRPGR} \leq 1$	population growth exceeds land consumption
$1 < \text{LCRPGR} \leq 2$	land consumption 1-2 times population growth
$\text{LCRPGR} > 2$	land consumption exceeds population growth by more than twofold

The distribution of cities across LCRPGR categories reveals clear temporal dynamics (Fig. 12). During the early period (2000–2005), many cities were characterized by moderate to high land consumption relative to population growth, with 32% of cities in the range 0 < LCRPGR ≤ 1 and 27% in the range 1 < LCRPGR < 2. This indicates that, for many cities, land expansion was either comparable to or exceeded population growth.

Cities in Asia and Africa exhibit higher median LCRPGR values compared to European and North American cities, reflecting more rapid land consumption relative to population growth in rapidly urbanizing regions.

https://doi.org/10.5194/essd-2026-381
Preprint. Discussion started: 16 September 2026

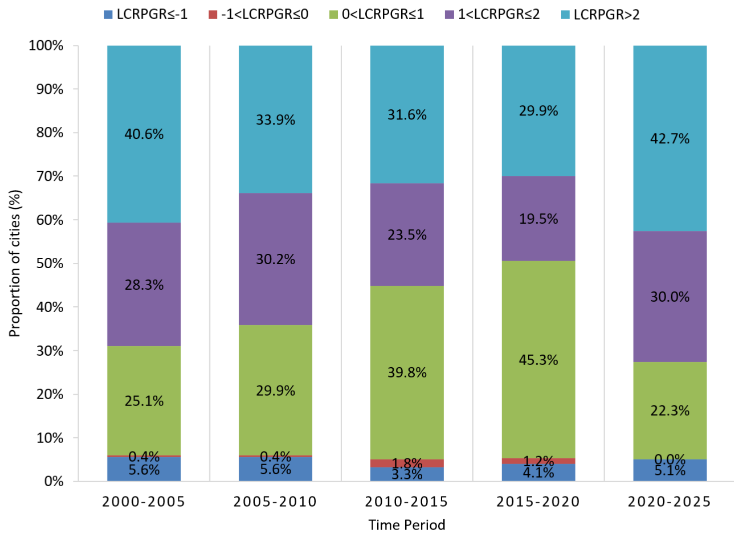

Figure 12. Temporal dynamics of SDG 11.3.1 (LCRPGR) categories across 1,611 cities for five time periods (2000–2005, 2005–2010, 2010–2015, 2015–2020, 2020–2025). Categories follow UN-Habitat (2020) discretization.

From 2005 to 2020, a gradual shift toward more balanced urban growth patterns was observed. The proportion of cities with 0< LCRPGR < 1 increased steadily, reaching approximately 55% during 2015–2020, suggesting improved coordination between land consumption and population growth in many regions.

The increase in LCRPGR during 2020–2025 indicates a renewed divergence between built-up area expansion and population growth in many cities. The drivers of this divergence may include post-2020 changes in demographic growth, infrastructure investment, peri-urban development, and methodological effects associated with recent high-resolution observations; however, attribution of these drivers is beyond the scope of this data descriptor. This proportion of cities with 0 < LCRPGR < 1 declined sharply to 16%, while cities with higher LCRPGR values (1 < LCRPGR < 2 and LCRPGR> 2) increased to 41% and 38%, respectively. This shift indicates a renewed acceleration of land consumption relative to population growth, suggesting increasing pressure on land resources and declining land-use efficiency in many cities. This pattern may reflect post-2020 urban expansion driven by infrastructure development and land speculation, particularly in rapidly developing regions, where land consumption increasingly outpaces population growth.

For all time periods, cities with negative LCRPGR values (LCRPGR<0) have a relatively small percentage (<12%), demonstrating that the population shrinkage was limited in most cases. In general, these results show a shift of urban expansion pattern, from the improvement of land-use efficiency in the previous years to a more recent pattern of imbalanced urban expansion which highlights the necessity of continuous monitoring using the consistent global datasets such as GUBAD. The spatial distribution of values of LCRPGR and their temporal changes over continents are further illustrated in Fig. 13, highlighting the regional difference in urban land-use efficiency.

The land-use efficiency results also align with independent estimates. The aggregate LCRPGR of the study cities declined from 1.52 in 2000-2005 to approximately 1.0 in 2010-2020 before rising again to 1.46 in 2020-2025. The long-term tendency of the LCRPGR to exceed unity matches the global analyses of UN-Habitat (2020), Schiavina et al. (2022) and Estoque et al. (2021), who found that most urban areas worldwide still consume land faster than they gain population; the near-unity values in 2010-2020 indicate a phase of relative densification, consistent with the stabilization of urban densities reported by Güneralp et al. (2020) and the increasing share of land-efficient urban centers documented by Santillan et al. (2025). At the continental aggregate level, Africa's urban population (+117%) grew faster than its built-up footprint (+93%) between 2000 and 2025, implying an aggregate LCRPGR

https://doi.org/10.5194/essd-2026-381
Preprint. Discussion started: 16 September 2026

below 1 and hence increasing average densities, even though the median city-level LCRPGR remains comparatively high in Africa; a similar densification signal in many African functional urban areas has been reported by Schiavina et al. (2022, 2023). The decline of the mean population density of the study cities, from about 7,800 to about 7,000 inhabitants per km^2 between 2000 and 2025, mirrors the density decreases documented globally by Angel et al. (2011), Güneralp et al. (2020) and Sun et al. (2020).

The LCRPGR estimates at city level for all 1,611 cities are provided in the individual attribute table in the supplementary section. The extreme values; very high positives (LCRPGR>>2) and negative (LCRPGR<0), largely reflect demographic boundary effects: the near-zero or negative growth of the population will increase the denominator of the ratio (Eq. 1), whereas rapid population increase can diminish it. The extremes are mathematical in nature and are part of the ratio-based indicators and need to be seen in combination with the absolute population and the land-consumption trend rather than a unique efficiency score.

Beyond SDG 11.3.1, the GUBAD enables quantitative analysis of peri-urban expansion rates and their correlation with agricultural land loss (SDG 15.1.1). Integration with nighttime light trends (VIIRS/DMSP-OLS) supports energy consumption proxy estimation, while combination with global climate reanalysis datasets (ERA5) facilitates urban heat island intensity mapping. The dataset's 30 m resolution and 25-year temporal depth make it suitable for machine-learning-based urban growth projection models, providing training data for future scenario simulations (2050–2100).

The operational utility of GUBAD for SDG 11.3.1 assessment lies in its provision of physically explicit, functionally defined UBA estimates. Because the LCR must be computed from contiguous impervious surface constrained by population presence—not from administrative or morphological boundaries that may absorb non-built-up hinterlands—boundary-based datasets can systematically overestimate land consumption and distort LCRPGR ratios. GUBAD's standardized 30 m UBA masks enable direct, reproducible LCR calculation across all 1,611 cities and six epochs, transforming the dataset from a static land-cover layer into an operational monitoring product capable of tracking urban LUE trajectories over a 25-year span.

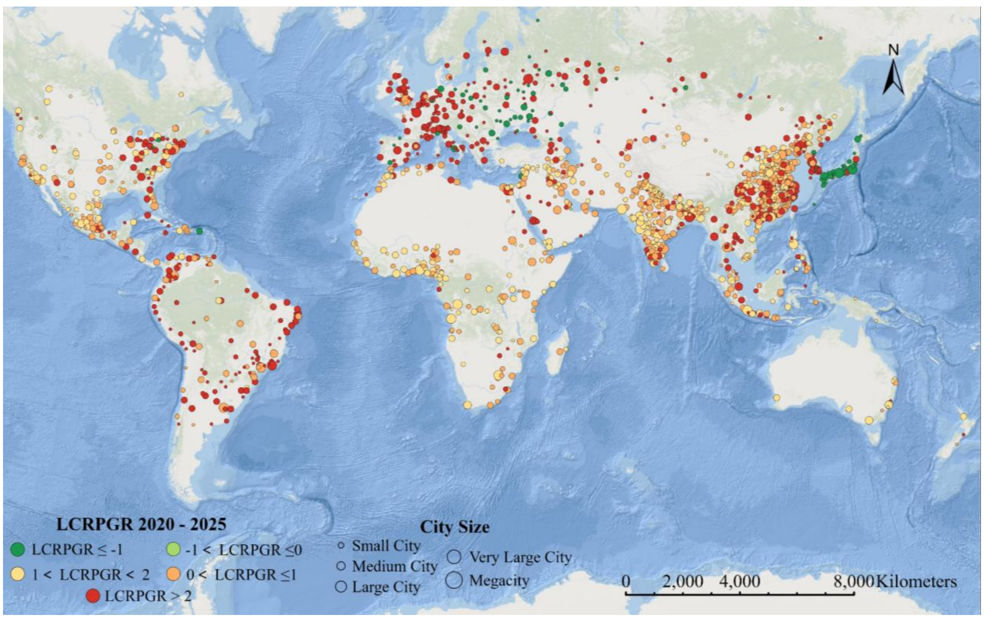

Figure 13. Spatial distribution of LCRPGR values and continental temporal changes across 1,611 cities, 2020–2025.

https://doi.org/10.5194/essd-2026-381
Preprint. Discussion started: 16 September 2026

5.2 Comparison with existing global datasets

Visual comparisons confirm that GUBAD provides a detailed and spatially consistent representation of urban built-up areas compared with existing global datasets. Across diverse urban contexts, GUBAD shows strong agreement with high-resolution imagery while better capturing fine-scale urban morphology, particularly in transitional urban and rural zones where mixed land-use patterns challenge conventional classification. MGUP and DEGURBA overestimate the urban extents in coastal and peri-urban regions, while AUE outputs smoother boundaries while underestimating the fragmented structure. The above differences underscore the strengths of GUBAD: the capability of representing both dense cores and heterogeneous expansions. A limitation lies in uniform parameterization for all cities that might not account for the region-by-region differences in characteristics and suggests the implementation opportunities for an adaptive approach.

GUBAD occupies a unique position that bridges the divide between global urban boundary datasets and raw impervious surface products. While the conceptual distinction from boundary-oriented datasets such as GUB has been established above, the empirical comparisons in Section 4.2 confirm the practical consequences. GUBAD's population-constrained UBA masks consistently yield more conservative area estimates than boundary-based approaches in peri-urban zones, avoiding the overestimation of land consumption that compromises SDG 11.3.1 calculations. Conversely, relative to high-resolution impervious surface layers such as Hi-GISA and GAIA, GUBAD translates physical urban fabric into policy interpretable UBAs aligned with UN urban agglomeration criteria. This hybrid design—multi-sensor ISA extraction coupled with spatial-morphological constraints incorporating LandScan population density, contiguity rules, and metropolitan context masks—explains the strong statistical agreement with GHSL-SMOD and the improved capture of fragmented, heterogeneous urban structures relative to MGUP and DEGURBA, which tend to overestimate extents in coastal and peri-urban regions.

GUBAD's better performance stems from its methodological design: multi-source data integration (optical, SAR, nighttime lights), city-specific training samples, Random Forest classification and temporal isotonic regression which attenuate the noise-induced anomalies. The standardized urban agglomeration framework for ISA-to-UBA-conversion enables consistent boundary delineation in diverse morphologies. The presented results further illustrate the importance of merging standardized urban definitions and high-resolution Earth observation data to enhance the comparability and reproducibility of global urban monitoring. Combining spatial morphology, population constraints and temporally consistent classification approaches enables GUBAD to better capture the heterogeneous urban growth patterns, such as fragmented peri-urban expansion, rapidly evolving settlement structures that are frequently under-represented in existing global products. This provides a strong analytical foundation to monitor the urban expansion dynamics and enables more reliable analysis of SDG indicator 11.3.1 across diverse urban contexts globally.

Beyond these differences in spatial representation, the temporal evolution captured by GUBAD is consistent with, and extends, previous global assessments. The 85% increase in the urban built-up area of the 1,611 study cities between 2000 and 2025 (an average rate of about 2.5% per year) falls within the range of expansion rates reported in the global meta-analysis of Seto et al. (2011) and in the remote-sensing-based estimates of Van Vliet (2019), and the tendency of urban land to grow faster than urban population in most world regions reproduces the central finding of that meta-analysis and of Angel et al. (2011) and Mahtta et al. (2022). The deceleration of the land consumption rate after 2010 visible in GUBAD is in line with Güneralp et al. (2020), who reported declining urban expansion rates at the global scale after 2000. Regionally, the more than threefold growth of urban built-up area in China—which overtook the United States as the country with the largest urban footprint after 2010—corroborates analyses based on GUB and GHSL data that identified East Asia as the hotspot of global urban expansion since 2000 (Jiang et al., 2021; Mahtta et al., 2022; Seto et al., 2011; Wang et al., 2020), while Africa's position as the continent with the fastest relative growth (+93% between 2000 and 2025) is consistent with the high African expansion rates reported by Schiavina et al. (2023) and Seto et al. (2011).

5.3 Limitations of the GUBAD dataset

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 monotonic urban growth assumption might neglect the localized urban shrinkage, redevelopment, or land-use conversion, particularly in post-industrial regions of North America and Europe. Additionally, anthropogenic but permeable surfaces (e.g., golf courses, parks with impervious infrastructure) introduce subjectivity in boundary delineation. Asymmetries between epochs: 2000—2015 used only optical data (Landsat), whereas 2020—2025 incorporated both SAR (Sentinel-1) and nighttime lights (VIIRS), may produce a systematic difference in detection; however, isotonic regression (PAVA) mitigates the temporal inconsistency.

GUBAD demonstrates strong potential for global urban studies and policy applications. It supports SDG 11.3.1 estimation and urban land-use efficiency analysis across cities and regions. Integration with socioeconomic and environmental data enables analysis of urban expansion, climate impacts, and peri-urban dynamics.

Future work should address: (1) incorporating higher-resolution data and deep learning approaches; (2) extending coverage to smaller cities (<300,000 population); and (3) developing spatially explicit uncertainty metrics that propagate classification errors, particularly from sensor resolution differences and uniform threshold application, into derived UBA estimates for enhanced local-scale decision support.

6 Code and data availability

The primary satellite observations used in this study include Landsat 5/7/8 TM/ETM+ imagery for the 2000–2015 period and Sentinel-1 SAR and Sentinel-2 MSI for the 2020–2025 period. Landsat data are available from the United States Geological Survey and were accessed through the Google Earth Engine data catalogue. Sentinel-1 and Sentinel-2 data are available from the Copernicus Data Space Ecosystem (https://dataspace.copernicus.eu, last access: 28 April 2026) under the Copernicus open data policy. Ancillary datasets used for sample generation, urban-boundary harmonization, and population-constrained UBA delineation include GHSL-SMOD, ESA WorldCover, the Copernicus Global Land Cover product, GADM v3.6, GRUMPv1, LandScan population data, and the United Nations World Urbanization Prospects database. These datasets are available from their respective data providers, as listed in Table 2.

The Global Urban Built-up Area Dataset (GUBAD) produced in this study is publicly available at Zenodo, https://doi.org/10.5281/zenodo.20051123 (Gunasekera et al., 2025). GUBAD provides 30 m urban built-up area maps for 1,611 global cities with populations exceeding 300,000 across six epochs: 2000, 2005, 2010, 2015, 2020, and 2025. The released dataset includes: (1) vectorized UBA polygons (GeoPackage format, .gpkg, EPSG:4326/WGS84, UTF-8 encoding, deflate compression) after topological correction and minimum-area filtering; (2) a city-level attribute table (CSV, RFC 4180, UTF-8) containing city identifiers, country names, ISO3 codes, continents, population records, UBA area estimates (km^2), and pre-computed SDG 11.3.1 indicators (LCR, PGR, LCRPGR) for each city and epoch, enabling immediate use by policymakers and researchers without requiring additional geospatial processing; and (3) ISO 19115-compliant metadata (XML) describing file naming conventions, spatial reference systems, data formats, classification codes, quality-control flags, and data lineage.

The city list and supplementary table linking UN World Urbanization Prospects records with spatial city identifiers are also provided to support reproducibility. The scripts used for impervious surface extraction, ISA-to-UBA conversion, accuracy assessment, and SDG 11.3.1 calculations are available at https://github.com/dinoora/GUBADv1.0 and linked through the Zenodo dataset repository. Future updates of GUBAD will be released as independent versioned datasets, accompanied by updated metadata and technical documentation.

7 Conclusions

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 dataset provides a temporally consistent and spatially detailed representation of global urban built-up areas. Validation and cross-dataset comparisons demonstrate that GUBAD achieves high accuracy while offering improved spatial detail and broad coverage compared to existing global urban products. These characteristics enable reliable assessment of urban expansion and support applications such as the evaluation of SDG indicator 11.3.1 and other urban sustainability analyses. By addressing key limitations in existing global urban datasets, particularly inconsistencies in urban definitions and temporal instability, GUBAD provides a more reliable basis for long-term monitoring of urban expansion and a standardized, scalable, and policy-relevant framework for global urban monitoring. This supports more consistent assessment of urban expansion dynamics and strengthening the evidence base for sustainable urban planning and SDG 11 implementation.

Author contributions

DG, DWJ and SZC designed the study; GJ and DWJ discussed the method and analysis sections; DG and DWJ implemented the study and drafted the manuscript; SZC and RN provided the resources. YPZ, YZ, and HG reviewed and revised the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

The authors acknowledge the use of data and resources provided by the Google Earth Engine platform and thank the United Nations and other open data providers for making global datasets publicly accessible. The authors also appreciate the valuable support and contributions from colleagues and collaborators involved in this study.

Financial support

This work was supported in part by the National Natural Science Foundation of China (No. 42361144884, 42171291, and 42471363), and CAS-A*STAR Collaborative Research Project (No. 313GJHZ2025038MI).

References

Angel, S.: Atlas of urban expansion, Lincoln Institute of Land Policy, Cambridge, Mass, 2012.

Angel, S., Parent, J., Civco, D. L., Blei, A., and Potere, D.: The dimensions of global urban expansion: Estimates and projections for all countries, 2000–2050, Elsevier, 75, 53–107, https://doi.org/10.1016/j.progress.2011.04.001, 2011.

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., and Silverman, E.: An Empirical Distribution Function for Sampling with Incomplete Information, Ann. Math. Stat., 26, 641–647, https://doi.org/10.1214/aoms/1177728423, 1955.

https://doi.org/10.5194/essd-2026-381
Preprint. Discussion started: 16 September 2026

Barlow, R. E. (Ed.): Statistical inference under order restrictions: the theory and application of isotonic regression, Repr., Wiley, Chichester, 388 pp., 1980.

Cao, H., Zhang, H., Wang, C., and Zhang, B.: Operational Built-Up Areas Extraction for Cities in China Using Sentinel-1 SAR Data, Remote Sens., 10, 874, https://doi.org/10.3390/rs10060874, 2018.

Dijkstra, L. and Poelman, H.: A harmonised definition of cities and rural areas: the new degree of urbanisation, 2014.

Dijkstra, L., Florczyk, A. J., Freire, S., Kemper, T., Melchiorri, M., Pesaresi, M., and Schiavina, M.: Applying the Degree of Urbanisation to the globe: A new harmonised definition reveals a different picture of global urbanisation, J. Urban Econ., 125, 103312, https://doi.org/10.1016/j.jue.2020.103312, 2021.

Estoque, R. C., Ooba, M., Togawa, T., Hijioka, Y., and Murayama, Y.: Monitoring global land-use efficiency in the context of the UN 2030 Agenda for Sustainable Development, Habitat Int., 115, 102403, https://doi.org/10.1016/j.habitatint.2021.102403, 2021.

European Space Agency: Copernicus DEM: Global Digital Elevation Model, https://spacedata.copernicus.eu/collections/copernicus-digital-elevation-model, 2021.

Foody, G. M.: Assessing the accuracy of land cover change with imperfect ground reference data, Remote Sens. Environ., 114, 2271–2285, https://doi.org/10.1016/j.rse.2010.05.003, 2010.

Gunasekera, D., Sun, Z., Du, W., Gao, J., Zhang, Y., Zhou, Y., Ndugwa, R., & Guo, H. (2026). Global Urban Built-up Area Dataset (GUBAD) for Cities with Populations Exceeding 300,000 (2000–2025) (1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.20051123

Güneralp, B., Reba, M., Hales, B. U., Wentz, E. A., and Seto, K. C.: Trends in urban land expansion, density, and land transitions from 1970 to 2010: a global synthesis, Environ. Res. Lett., 15, 044015, https://doi.org/10.1088/1748-9326/ab6669, 2020.

Jiang, H., Sun, Z., Guo, H., Weng, Q., Du, W., Xing, Q., and Cai, G.: An assessment of urbanization sustainability in China between 1990 and 2015 using land use efficiency indicators, Npj Urban Sustain., 1, 34, https://doi.org/10.1038/s42949-021-00032-y, 2021.

Li, X., Gong, P., Zhou, Y., Wang, J., Bai, Y., Chen, B., Hu, T., Xiao, Y., Xu, B., Yang, J., Liu, X., Cai, W., Huang, H., Wu, T., Wang, X., Lin, P., Li, X., Chen, J., He, C., Li, X., Yu, L., Clinton, N., and Zhu, Z.: Mapping global urban boundaries from the global artificial impervious area (GAIA) data, Environ. Res. Lett., 15, 094044, https://doi.org/10.1088/1748-9326/ab9be3, 2020.

Liu, F., Wang, S., Xu, Y., Ying, Q., Yang, F., and Qin, Y.: Accuracy assessment of Global Human Settlement Layer (GHSL) built-up products over China, PLOS ONE, 15, e0233164, https://doi.org/10.1371/journal.pone.0233164, 2020.

Mahtta, R., Michail, F., Güneralp, B., Mahendra, A., Reba, M., Wentz, E. A., and Seto, K. C.: Urban land expansion: the role of population and economic growth for 300+ cities, Npj Urban Sustain., 2, 5, https://doi.org/10.1038/s42949-022-00048-y, 2022.

Santillan, J., Dorozynski, M., and Heipke, C.: Enhancing Land Use Efficiency Assessment Through Built-Up Area–Built-Up Volume Trajectories: Integrating Vertical Urban Growth into SDG 11.3.1 Monitoring, ISPRS Int. J. Geo-Inf., 14, 404, https://doi.org/10.3390/ijgi14100404, 2025.

Schiavina, M., Melchiorri, M., Freire, S., Florio, P., Ehrlich, D., Tommasi, P., Pesaresi, M., and Kemper, T.: Land use efficiency of functional urban areas: Global pattern and evolution of development trajectories, Habitat Int., 123, 102543, https://doi.org/10.1016/j.habitatint.2022.102543, 2022.

Schiavina, M., Melchiorri, M., and Pesaresi, M.: GHS-SMOD R2023A - GHS settlement layers, application of the Degree of Urbanisation methodology (stage I) to GHS-POP R2023A and GHS-BUILT-S R2023A, multitemporal (1975-2030), https://doi.org/10.2905/A0DF7A6F-49DE-46EA-9BDE-563437A6E2BA, 2023.

Seto, K. C., Fragkias, M., Güneralp, B., and Reilly, M. K.: A Meta-Analysis of Global Urban Land Expansion, PLoS ONE, 6, e23777, https://doi.org/10.1371/journal.pone.0023777, 2011.

Seto, K. C., Güneralp, B., and Hutyra, L. R.: Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools, Proc. Natl. Acad. Sci., 109, 16083–16088, https://doi.org/10.1073/pnas.1211658109, 2012.

Shahfahad, Naikoo, M. W., Mallick, J., and Rahman, A.: Land use/land cover change and its impact on surface urban heat island and urban thermal comfort in a metropolitan city, Sci. Direct, 41, 101052, https://doi.org/10.1016/j.uclim.2021.101052, 2022.

Shi, K., Wu, Y., Liu, S., Chen, Z., Huang, C., and Cui, Y.: Mapping and evaluating global urban entities (2000–2020): A novel perspective to delineate urban entities based on consistent nighttime light data, GIScience Remote Sens., 60, 2161199, https://doi.org/10.1080/15481603.2022.2161199, 2023.

Sun, L., Chen, J., Li, Q., and Huang, D.: Dramatic uneven urbanization of large cities throughout the world in recent decades, Nat. Commun., 11, 5366, https://doi.org/10.1038/s41467-020-19158-1, 2020.

Sun, Z., Xu, R., Du, W., Wang, L., and Lu, D.: High-Resolution Urban Land Mapping in China from Sentinel 1A/2 Imagery Based on Google Earth Engine, Remote Sens., 11, 752, https://doi.org/10.3390/rs11070752, 2019.

Sun, Z., Du, W., Jiang, H., Weng, Q., Guo, H., Han, Y., Xing, Q., and Ma, Y.: Global 10-m impervious surface area mapping: A big earth data based extraction and updating approach, Int. J. Appl. Earth Obs. Geoinformation, 109, 102800, https://doi.org/10.1016/j.jag.2022.102800, 2022.

Sun, Z., Du, W., Li, S., Yang, M., Ouyang, X., and Guo, H.: A rapidly updated mapping method for high-resolution global impervious surface area (Hi-GISA) products, Sci. Bull., 70, 1071–1075, https://doi.org/10.1016/j.scib.2025.01.039, 2025.

UN-Habitat (Ed.): The value of sustainable urbanization, UN-Habitat, Nairobi, Kenya, 377 pp., 2020.

United Nations Department of Economic and Social Affairs: World Urbanization Prospects 2018: Highlights, United Nations, 1 p., 2018.

United Nations Department of Economic and Social Affairs: World Urbanization Prospects 2025: Summary of Results, United Nations, 2025.

Van Vliet, J.: Direct and indirect loss of natural area from urban expansion, Nat. Sustain., 2, 755–763, https://doi.org/10.1038/s41893-019-0340-0, 2019.

Wang, H., Liu, Y., Wang, Y., Yao, Y., and Wang, C.: Land cover change in global drylands: A review, Sci. Total Environ., 863, 160943, https://doi.org/10.1016/j.scitotenv.2022.160943, 2023a.

Wang, Y., Huang, C., Feng, Y., Zhao, M., and Gu, J.: Using Earth Observation for Monitoring SDG 11.3.1-Ratio of Land Consumption Rate to Population Growth Rate in Mainland China, Remote Sens., 12, 357, https://doi.org/10.3390/rs12030357, 2020.

Wang, Y., Sun, Y., Cao, X., Wang, Y., Zhang, W., and Cheng, X.: A review of regional and Global scale Land Use/Land Cover (LULC) mapping products generated from satellite remote sensing, ISPRS J. Photogramm. Remote Sens., 206, 311–334, https://doi.org/10.1016/j.isprsjprs.2023.11.014, 2023b.

Zhang, X., Liu, L., Wu, C., Chen, X., Gao, Y., Xie, S., and Zhang, B.: Development of a global 30 m impervious surface map using multisource and multitemporal remote sensing datasets with the Google Earth Engine platform, Earth Syst. Sci. Data, 12, 1625–1648, https://doi.org/10.5194/essd-12-1625-2020, 2020a.

Zhang, Y., Wang, H., Xie, P., Rao, Y., and He, Q.: Revisiting Spatiotemporal Changes in Global Urban Expansion during 1995 to 2015, Complexity, 2020, 1–11, https://doi.org/10.1155/2020/6139158, 2020b.

Zhao, M., Cheng, C., Zhou, Y., Li, X., Shen, S., and Song, C.: A global dataset of annual urban extents (1992–2020) from harmonized nighttime lights, Earth Syst. Sci. Data, 14, 517–534, https://doi.org/10.5194/essd-14-517-2022, 2022.