Uncertainty Quantification and Monte Carlo Simulations to Enhance SDG 11.3.1 Monitoring

Jojene R. Santillan [1, 2], Mareike Dorozynski [1], Christian Heipke [1]

[1] Institute of Photogrammetry and GeoInformation, Leibniz University Hannover, Nienburger Str. 1, 30167 Hannover, Germany – {santillan, dorozynski, heipke} @ipi.uni-hannover.de
[2] Caraga Center for Geo-Informatics & Department of Geodetic Engineering, College of Engineering and Geosciences, Caraga State University, Ampayon, 8600 Butuan City, Philippines – jrsantillan@carsu.edu.ph

Keywords: Uncertainty quantification, Accuracy assessment, Monte Carlo simulation, Sustainable development goals, SDG 11.3.1.

Abstract

As urbanization accelerates globally, efficient land use is essential for sustainable development. Sustainable Development Goal (SDG) 11 includes Target 11.3, which promotes sustainable urbanization and is assessed through Indicator 11.3.1, measuring land use efficiency (LUE) via the ratio of Land Consumption Rate (LCR) to Population Growth Rate (PGR), known as LCRPGR. While this metric is valuable for guiding urban planning, it could be affected by variability in Earth Observation (EO) data products, especially differences in built-up area definitions and classification errors that lead to data quality issues. This paper presents a comprehensive approach to improving SDG 11.3.1 monitoring by (1) quantifying the impact of EO data variability on LCR and LCRPGR estimates, highlighting the importance of standardized built-up area definitions; (2) adapting a bias-adjustment methodology to correct for over- and underestimations in EO-derived built-up area estimates, thus enhancing accuracy; and (3) incorporating Monte Carlo simulations to quantify uncertainties in LCR and LCRPGR due to classification errors. The findings indicate that definitions and adjustments significantly influence the SDG 11.3.1 metrics. Monte Carlo simulations provide essential insights into the confidence intervals of LCR and LCRPGR values, revealing the degree of uncertainty tied to EO data accuracy. This study supports more reliable urban planning and policy formulation by ensuring LCR and LCRPGR values reflect actual urban dynamics in a better way, enabling robust, equitable comparisons across cities, countries, and SDG regions.

1. Introduction

Urbanization is one of the defining trends of the 21st century, with more than half of the global population currently residing in urban areas—a proportion expected to increase in the coming decades (United Nations, 2019). As cities expand, ensuring efficient land use becomes critical to sustainable development. In this context, Sustainable Development Goal 11 includes Target 11.3, which promotes sustainable urbanization through optimized land use. The key indicator for monitoring progress, SDG 11.3.1, measures the ratio of the land consumption rate to the population growth rate, commonly referred to as LCRPGR (UN Statistics Division, 2021). This metric provides valuable insights into the alignment between urban expansion and population dynamics, serving as a foundational tool for policymakers in designing sustainable urban growth strategies (Li et al., 2021).

To ensure accurate SDG 11.3.1 monitoring, the United Nations recommends an integrated methodology that leverages Earth observation data, geospatial analysis, and demographic information from censuses and surveys (UN Statistics Division, 2021). These data sources collectively provide essential spatial and temporal insights into built-up areas and population distributions, which are crucial for calculating LCR and PGR and, ultimately, for assessing land-use efficiency through the LCRPGR metric. Over the years, research on SDG 11.3.1 monitoring has primarily relied on publicly available global EO data products to estimate built-up area coverage and LCR calculations, with the Global Human Settlement Layers as a popular choice (Santillan and Heipke, 2024). Additionally, various classification techniques, including traditional machine learning and more advanced deep learning approaches, have been applied to extract built-up areas from satellite imagery (e.g., Li et al., 2020; Ghazaryan et al., 2021). To improve accuracy, some studies have employed workflows that integrated multiple data sources (e.g., Cardenas-Ritzert et al., 2024; Jiang et al., 2021), highlighting the ongoing efforts to refine urban expansion analysis to aid SDG 11.3.1 monitoring.

Despite advancements in SDG 11.3.1 monitoring, a key challenge remains the lack of standardized built-up area definitions. While the official SDG 11.3.1 metadata defines built-up areas as "all areas occupied by buildings" (UN Statistics Division, 2021), many studies use broader definitions for their classifications, such as impervious surfaces, artificial surfaces, or urban land (e.g., Cai et al., 2020; Jiang et al., 2021; Li et al., 2021; Huang et al., 2024). These inconsistencies lead to variations in LCR and LCRPGR estimates, affecting comparability across cities, countries, and SDG regions, and potentially undermining the reliability of SDG 11.3.1 assessments. Evidence from recent studies highlights substantial discrepancies in built-up area estimates across EO datasets, attributed to differences in definitions, spatial resolutions, and data processing methodologies (Chakraborty et al., 2024). Given the significant impact of these discrepancies on LCR and LCRPGR calculations, systematic evaluations are needed to assess their influence on SDG 11.3.1 indicators.

Another critical limitation is the absence of uncertainty quantification in LCR, PGR, and LCRPGR estimations. While the SDG 11.3.1 metadata emphasizes data quality assurance, it lacks standardized methodologies for incorporating uncertainty into these computations. This oversight represents a significant gap, particularly given that LCRPGR values are used to benchmark and compare urban sustainability trends at local, national, and global scales (Estoque et al., 2021; Schiavina et al., 2022). Neglecting uncertainties from classification errors, data inconsistencies, built-up area definitions, and population estimates can result to misleading interpretations of urban sustainability progress. Addressing this issue requires standardized classification approaches and rigorous uncertainty quantification frameworks that integrate both built-up area and population data uncertainties. These improvements would enhance the reliability and comparability of SDG 11.3.1 assessments, providing more actionable insights for policymakers and urban planners.

In this paper, we present a comprehensive approach to enhance SDG 11.3.1 monitoring by analyzing EO data variability, implementing bias-adjusted built-up area estimation, and integrating uncertainty quantification. Key contributions include:

- **Impact of EO data variability on LCR and LCRPGR estimates**: We demonstrate how differences in built-up area definitions and classification errors across EO datasets affect LCR and LCRPGR estimates, underscoring the need for consistency in LUE assessments.
- **Bias-adjusted built-up area estimates:** Using a bias-adjustment methodology (Olofsson et al., 2013, 2014), we correct classification errors, ensuring EO-derived built-up area estimates align with the SDG 11.3.1 definition.
- **Uncertainty quantification in SDG 11.3.1 metrics**: We integrate Monte Carlo simulations to propagate uncertainties in built-up area estimates into LCR and LCRPGR calculations, providing LUE metrics that explicitly account for classification uncertainties.

2. Related Work

2.1 Accuracy Assessment, Class Area Estimation, and Uncertainty Quantification Approaches

Land-use/land cover (LULC) maps and other EO-derived products inherently contain multiple sources of uncertainty, with one of the most critical factors being the accuracy of the classifier's predictions (Valle et al., 2023). A map is accurate when it provides an unbiased representation of land cover in the area it depicts (Foody, 2002). However, classification errors are inevitable, making direct area calculations based solely on classified pixels—the so-called "pixel counting" approach—prone to inaccuracies (Stehman, 2013). A standard practice to quantify classification errors and assess a LULC map's reliability involves performing accuracy assessments by comparing classified outputs against those from high-quality reference data, such as ground-truth observations, high-resolution aerial imagery, or finer-resolution satellite data (Olofsson et al., 2014). These comparisons are systematically organized in an error (or "confusion") matrix, from which key accuracy metrics—including User's Accuracy, Producer's Accuracy, and Overall Accuracy—are derived. These metrics provide essential insights into classification performance, allowing the refinement of classification models and adjustments to derived estimates (Foody, 2002).

Since classification errors affect the reliability of built-up area estimates, selecting an appropriate method for area estimation and uncertainty quantification is essential to ensure accurate and unbiased calculations of metrics that depend on these estimates, such as LCR and LCRPGR. In the context of land cover maps, area estimation is a special case of mean estimation, where the mean is calculated for a binary variable that assigns a value of 1 to pixels classified as belonging to the class of interest and 0 to all other pixels (Lu et al., 2024). This mean value represents the proportion of the target class within the study area. Three primary approaches (Table 1) can be used for area estimation: the classical estimator, the post-stratified estimator, and prediction-powered inference (PPI) (Lu et al., 2024).

The classical estimator relies exclusively on ground truth samples to estimate class proportions within the study area. This method assumes that sample proportions are an unbiased representation of the total area, scaling up observed proportions to generate class area estimates (Lu et al., 2024). Its main advantages are simplicity, ease of implementation, and independence from classified land cover maps. When reference data are abundant and well-distributed, the classical estimator can yield reliable estimates with well-defined confidence intervals. However, its effectiveness is limited in reference-scarce settings because it does not leverage valuable spatial information in classified maps.

In contrast, the post-stratified estimator (Olofsson et al., 2013) improves upon the classical approach by integrating ground truth data with the classified land cover map while explicitly correcting for classification errors. This method stratifies samples based on map classes and applies an adjustment using the confusion matrix to account for omission and commission errors. The result is a bias-adjusted class area estimate with confidence intervals reflecting sampling variability and classification uncertainty. Recommended for land cover change analysis (Olofsson et al., 2014), the post-stratified estimator has been widely applied for estimating built-up area and other land cover classes from EO-derived maps (e.g., Liu et al., 2018; Gong et al., 2020).

A more recent approach, PPI (Angelopoulos et al., 2023), integrates large-scale, potentially unreliable predictions with limited but highly trusted ground truth data to compute statistically valid confidence intervals. In the context of land cover class area estimation, PPI improves upon the classical estimator by leveraging a small reference dataset to calibrate and

Method	Class area estimate, $\hat{A}_j$	Standard Deviation, $S(\hat{A}_j)$	Reference
Classical	$A_{tot} \cdot \frac{1}{n}\sum_{i=1}^{n} Y_i$	$A_{tot} \cdot \sqrt{\frac{1}{n}Var(Y_i)}$	Lu et al. (2024)
Post-stratified	$A_{tot} \cdot \sum_{c=1}^{K} W_c \frac{n_{cj}}{n_{c.}}$	$A_{tot} \cdot \sqrt{\sum_{c=1}^{K} W_c^2 \frac{\frac{n_{cj}}{n_{c\cdot}}\left(1-\frac{n_{cj}}{n_{c\cdot}}\right)}{n_{c\cdot}-1}}$	Olofsson et al. (2013)
Prediction-powered inference (PPI)	$A_{tot} \cdot \left(\frac{1}{N}\sum_{p=1}^{N} \hat{Y}_p' - \frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}_i' - Y_i\right)\right)$	$A_{tot} \cdot \sqrt{\frac{1}{N}Var(\hat{Y}_p') + \frac{1}{n}Var(\hat{Y}_i' - Y_i)}$	Angelopoulos et al. (2023), Lu et al. (2024)

Table 1. Land cover class area estimation and uncertainty quantification methods, along with their corresponding formulations. For each method, the 95% confidence interval is computed as $\left(\hat{A}_j - 1.96 \cdot S(\hat{A}_j),\ \hat{A}_j + 1.96 \cdot S(\hat{A}_j)\right)$. The variables are defined as follows: A_{tot} is the total mapped area, Y_i and $\hat{Y}_i'$ $(i = 1, \dots, n)$ are the reference (ground truth) and predicted (classified) labels for sample i, n is the total number of reference samples, $\hat{Y}_p'$ $(p = 1, \dots, N)$ is the predicted label for pixel p in the classified map, N is the total number of pixels in the classified map, W_c $(c = 1, \dots, K)$ is the proportion of pixels mapped as class c, $n_{c.}$ is the number of reference samples in mapped class c, n_{cj} $(c, j = 1, \dots, K)$ is the number of reference samples labelled as class j within mapped class c, K is the number of classes, and $Var(\cdot)$ is the variance operator.

correct bias in a larger predicted dataset (Lu et al., 2024). Unlike the post-stratified estimator, which explicitly corrects for class-specific misclassification, PPI applies a global bias correction to align overall predicted class proportions with ground truth data. Its key advantage lies in its flexibility—it does not require a predefined map product or stratification scheme and often produces narrower confidence intervals, making it particularly suitable when precision is a priority (Lu et al., 2024).

While each method has distinct advantages, the post-stratified estimator (Olofsson et al., 2013) is particularly well-suited for SDG 11.3.1 monitoring due to its balance of accuracy, efficiency, and explicit error correction. The classical estimator, despite its simplicity, is heavily dependent on extensive ground truth data and disregards valuable classified map information, limiting its usefulness in regions with limited reference samples (Lu et al., 2024). Conversely, PPI depends on the accuracy of the classified map, making it most effective when classification closely aligns with the ground truth and when the number of predicted pixels significantly exceeds the number of reference samples (Angelopoulos et al., 2023). Given the inherent variability in built-up area classification accuracy, the post-stratified estimator provides a robust alternative, correcting for known biases to ensure unbiased estimates. This makes it particularly valuable for LCR and LCRPGR calculations, where accurate error-adjusted estimates are essential for reliable SDG 11.3.1 monitoring.

2.2 Uncertainty Propagation in SDG 11.3.1 Monitoring: The Role of Monte Carlo Simulations

Several methodologies are commonly employed for uncertainty propagation, including Monte Carlo simulations (MCS), sensitivity analysis, and analytical approaches. Among these, MCS is widely recognized for its ability to generate probabilistic estimates by repeatedly sampling from the error distributions of input data (Alkhatib et al., 2009). This method is particularly valued for its adaptability and robustness in modeling complex systems, particularly in cases where analytical methods struggle with non-Gaussian error distributions or correlated inputs (Albert, 2020). Although MCS has been extensively applied in environmental modeling and geospatial analysis (e.g., Mustafa et al., 2018), its potential remains largely unexplored within policy-driven monitoring frameworks, such as the one of SDG 11.3.1.

Given its demonstrated effectiveness in environmental and geospatial research, MCS presents a compelling approach for uncertainty propagation in SDG 11.3.1 monitoring. However, its application in this context remains limited. A key advantage of MCS is its ability to account for dependencies among variables, which is particularly critical for propagating built-up area uncertainties across multiple periods. In contrast, analytical methods, such as the Special Law of Propagation of Variances, assume variable independence (Ghilani, 2017)—an unrealistic assumption in multitemporal analysis of built-up area estimates. While the General Law of Propagation of Variances does accommodate correlations, it necessitates explicit covariance estimates (Ghilani, 2017), which are often difficult to obtain. Moreover, the inherently nonlinear nature of SDG 11.3.1 indicators, which involve ratios and rates of change, poses additional challenges for analytical uncertainty propagation techniques.

Given these limitations, a Monte Carlo-based approach offers a more flexible and robust framework for quantifying the impact of built-up area uncertainties on LCR and LCRPGR estimates. By systematically incorporating uncertainty into SDG 11.3.1 assessments, MCS can enhance the reliability of monitoring outcomes, thereby improving the validity of land-use efficiency evaluations and supporting more informed policy decisions. Expanding the application of MCS in this domain represents a crucial step toward addressing the methodological gaps in SDG 11.3.1 uncertainty quantification and ensuring the robustness of sustainability assessments.

3. Methodology for Evaluating the Impact of EO Data Variability and Uncertainty on LCR and LCRPGR Estimation

3.1 Quantifying the Impact of Built-Up Area Definitions on LCR and LCRPGR

To quantify how different built-up area definitions impact SDG 11.3.1 metrics, we analyzed five global EO data products: GHS-BUILT-S, Global Annual Urban Dynamics (GAUD), Global Impervious Surface Area 2.0 (GISA 2.0), Global 30m Impervious Surface Dynamic Dataset (GISD30), and WSF Evolution (WSF-Evo). Table 2 provides an overview of these datasets, highlighting their key characteristics, including built-up area definitions. Except for GHS-BUILT-S, which represents built-up areas as a continuous variable, all other datasets are binary built-up area maps, where built-up pixels are coded as one and non-built-up pixels as 0. Using GHS-BUILT-S as the baseline, we assessed temporal differences in built-up area, LCR, and LCRPGR across datasets.

The Philippines was chosen as the study site (Figure 1) due to its variable urbanization patterns and diverse land cover, shaped by its archipelagic nature (Santillan and Heipke, 2024). This variability provides an ideal setting to evaluate how the different class definitions in EO data products influence LCR and LCRPGR estimates. For GHS-BUILT-S, total built-up area was calculated by summing all pixel values within the country boundary using zonal statistics. For binary EO datasets, built-up area was determined by counting pixels classified as built-up and multiplying this count by the ground area represented by each pixel. The analysis utilized an administrative boundary shapefile from the United Nations Office for the Coordination of Humanitarian Affairs (OCHA), accessed via the Humanitarian Data Exchange (https://data.humdata.org/dataset/cod-ab-phl, accessed 8 April 2024).

To ensure temporal consistency, the study focused on the 1985–2015 period, covered by most datasets. The SDG 11.3.1 metrics were calculated using the following formulae provided by the UN Statistics Division (2021), with notations slightly modified for consistency:

$$LCR = \frac{BU_{t_2} - BU_{t_1}}{BU_{t_1}} \cdot \frac{1}{\Delta t} \quad (1)$$

$$PGR = \frac{ln\left(\frac{Pop_{t_2}}{Pop_{t_1}}\right)}{\Delta t} \quad (2)$$

$$LCRPGR = \frac{LCR}{PGR} \quad (3)$$

where *BU* and *Pop* are the total built-up area and population for the previous (t_1) and current (t_2) years, respectively, and Δt represents the number of years between them (set to 5 years in our case). Population data were sourced from the Philippine Statistics Authority (PSA) website (https://www.psa.gov.ph/, accessed 13 September 2024). Since no census survey was conducted in 2005, the population data for that year was linearly interpolated using the annual growth rate provided by PSA for 2000–2007. To assess the impact of built-up area differences on LCR and LCRPGR, we compared each dataset's LCR and LCRPGR values against the GHS-BUILT-S baseline using

EO Data Product	Pixel size (m)	Temporal Coverage	Built-up Class Name / Definition	Reference
Global Human Settlement – BUILT-S (GHS-BUILT-S)	100	1975-2020 at 5-year intervals	‘Built-up Area’ (areas occupied by buildings and “any roofed structure erected above ground for any use”)	European Commission (2023)
Global 30 m Impervious Surface Dynamic Dataset (GISD30)	30	1985-2020 at 5-year intervals	Impervious Surface	Zhang et al. (2022)
Global Impervious Surface Area (GISA) 2.0	30	1972-2019 at 1-year intervals	Impervious Surface Area	Huang et al. (2022)
Global Annual Urban Dynamics (GAUD)	30	1985-2015 at 1-year intervals	Urban Areas	Liu et al. (2020)
World Settlement Footprints Evolution (WSF Evo)	30	1985-2015 at 5-year intervals	Settlements	Marconcini et al. (2021)

Table 2. List of global EO data products included in the analysis.

relative differences (%), which quantify the percentage deviation from the baseline.

3.2 Unbiased Built-up Area Estimation and Uncertainty Quantification

In this section, we outline the workflow for unbiased built-up area estimation and uncertainty quantification to improve the reliability of SDG 11.3.1 metrics. The process adheres to good practice recommendations (Olofsson et al., 2014) and consists of five key steps: (1) stratified random sample selection, (2) sample labeling using reference (ground truth) data, (3) accuracy assessment using a confusion matrix approach, (4) post-stratified class area estimation and uncertainty quantification, following the methods described in Olofsson et al. (2013), and (5) Monte Carlo simulations to propagate uncertainties in built-up area estimates into LCR and LCRPGR calculations.

3.2.1 Study Area and Datasets: The study area is the City of Manila (Figure 1), a highly urbanized region within Metropolitan Manila, Philippines, with a total area of approximately 42 km². The analysis period spans 2000 to 2015 in 5-year intervals, aligned with the availability of most EO data products for the study area. This timeframe is long enough to capture trends in LCR and LCRPGR, enabling an assessment of how data uncertainties impact these metrics over time. The analysis focuses on binary built-up area maps derived from the 30m spatial resolution EO data products GAUD, GISA 2.0, GISD30, and WSF-Evo. Including multiple EO datasets enables an evaluation of the consistency of unbiased area estimation and uncertainty quantification. However, GHS-BUILT-S was excluded, as the methods applied in this study are not suitable for continuous datasets. For each analysis year, subsets of the EO data products covering the study area were extracted (Figure 2), all standardized to the UTM 51 WGS 84 coordinate reference system. Pixel counting was used to determine built-up and non-built-up areas, and the proportion of each class (W_c) was calculated for use in subsequent steps. Table 3 presents the proportion of the built-up areas in each map. Each map consists of a total of 46,645 pixels (N).

3.2.2 Stratified Random Sampling: We employed a stratified random sampling design with the map classes as strata and pixels as spatial units. To ensure statistically valid and robust analysis, we used the following recommended formula to determine the required number of ground truth samples (n) for each data product (Olofsson et al., 2014):

$$n = \frac{(\sum W_c S_c)^2}{[S(\hat{O})]^2 + \frac{\sum W_c S_c^2}{N}} \tag{4}$$

where N = number of pixels in the land cover map within the study area, $S(\hat{O})$ represents the targeted standard deviation of the estimated overall accuracy $\hat{O}$, W_c is the mapped ("classified") proportion of class c, S_c is the class c standard deviation; $S_c = \sqrt{U_c(1-U_c)}$, where U_c is the conjectured User's accuracy of class c (Olofsson et al., 2014). We set $S(\hat{O})$ to 0.01 to ensure high level of precision and statistical confidence in the accuracy estimations. For U_c, we used class-specific User's accuracies reported in previous studies that assessed the accuracy of the data products (Marconcini et al., 2020; Huang et al., 2022; Zhang et al., 2022). For all the considered years, the assigned U_c values for the 'built-up' class range from 0.59 (WSF Evo) to 0.85 (GAUD), while for the 'non-built-up' class, they range from 0.91 (WSF Evo) to 0.96 (GISD30). The low U_c assigned to WSF-Evo resulted in higher n than the other data products (Table 4). From the calculated value of n, the sample allocation for each class was determined, using proportional allocation with the mapped class proportions (Table 3) serving as multipliers.

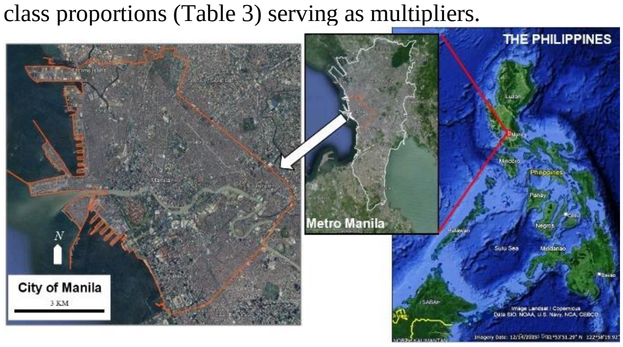

Figure 1. The study site. Image and maps data © 2024 Google, Airbus, Landsat / Copernicus, SIO, NOAA, US Navy, NGA, GEBCO.

3.2.3 Reference Data and Sample Labelling Process: High-resolution historical satellite images from Google Earth Pro, with a ground sampling distance of approximately m, served as the primary reference for labeling samples (Table 4). Before labeling, we assessed the co-registration accuracy using 14 spatially distributed reference points at road intersections visible in each image. Reference coordinates were obtained from a high-quality road network dataset from the Earthquake Impact Reduction Study for Metro Manila (MMEIRS, 2004). The analysis yielded a total RMSE of 3.00 to 7.68 meters, indicating that while the images were not perfectly co-registered, they remained suitable for validating coarser-resolution EO datasets. Nevertheless, we improved geolocation accuracy in sample labeling by adjusting the sample coordinates rather than images, because Google Earth Pro does not support geometric transformations. Using the same 14 reference points per image, we derived affine transformation coefficients to align sample coordinates with Google Earth imagery. This transformation, validated with another (independent) set of 14 reference points, reduced the total RMSE to under 2m. Finally, 30m × 30m polygons were generated around each transformed sample point, overlaid on Google Earth imagery, and labeled into three categories: Buildings, Roads and Other Impervious Surfaces, and Non-Impervious Surfaces (e.g., trees, grass, barren, and water) based on the dominant cover type.

We created two versions of labeled sample datasets for each EO data product. The first version retained each dataset's original class definitions. For instance, in GISA 2.0 ground truth samples, buildings, roads, and other impervious surfaces were grouped under 'ISA' (Impervious Surface Area), while all other samples were labeled as 'Non-ISA.' The second version reclassified the samples to align with the definition of SDG 11.3.1, where only 'Buildings' were labeled as 'Built-up area', and the rest were categorized as 'Non-built-up area.' This reclassification allowed us to assess how well each dataset estimates built-up areas based on the SDG 11.3.1 definition, assuming the absence of datasets designed to follow this classification.

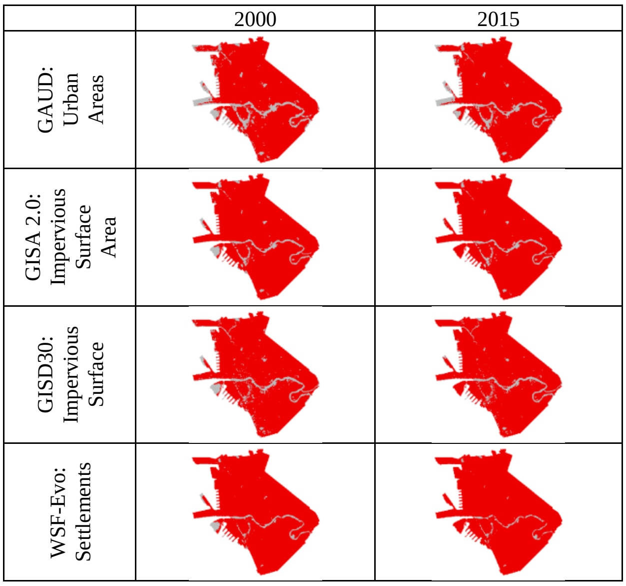

Figure 2. Example subsets of EO data products for Manila, Philippines, highlighting built-up areas (■) based on different definitions for 2000 and 2015.

Year	GAUD	GISA 2.0	GISD30	WSF Evo
2000	0.881	0.937	0.914	0.955
2005	0.890	0.943	0.924	0.964
2010	0.894	0.948	0.933	0.967
2015	0.902	0.951	0.938	0.974

Table 3. Mapped proportions of 'built-up area' by data product and year for Manila, Philippines.

Year	GAUD	GISA 2.0	GISD30	WSF Evo	Reference Google Earth Image Date
2000	1174	1242	1520	2208	2001/06/13
2005	1179	1247	1536	2226	2005/08/23, 2007/02/13*
2010	1181	1250	1551	2231	2010/02/05
2015	1185	1253	1559	2245	2015/09/09, 2015/08/07*

Table 4. Ground truth samples (n) for each data product and year covering Manila, Philippines, with high-resolution imagery dates from Google Earth Pro. Supplementary images (*) were used for cloud-covered samples.

3.2.4 Estimating accuracy, area, and confidence intervals: For each analysis year, error matrices of sample counts (n_{ij}) were generated using each version of the labeled sample dataset for each data product. These matrices were then used to apply the post-stratified method following Olofsson et al. (2013) as formulated in Table 1. The sample count error matrices were first converted into error matrices of estimated area proportions, enabling the direct computation of accuracy metrics—including bias-adjusted Producer's, User's, and Overall Accuracy—and adjusted built-up area estimates that account for classification errors. Standard deviations were calculated for the area estimates, and 95% confidence intervals (CIs) were constructed to quantify the uncertainty.

3.2.5 Monte Carlo simulations: The error-adjusted built-up area estimates and their corresponding standard deviations from previous calculations were used as inputs for Monte Carlo simulations to assess the uncertainty in LCR and LCRPGR estimates. The simulations were conducted for each analysis period (i.e., 2000–2005, 2005–2010, and 2010–2015) with two cases. For Case 1, estimates and standard errors were derived using each dataset's original class definition, while for Case 2, they were based on the SDG 11.3.1 built-up area definition.

The simulations were implemented using Python, generating 100,000 random samples for built-up area estimates within the range of their standard deviations, assuming a normal distribution. The Central Limit Theorem justifies the assumption that built-up area errors follow a normal distribution (Fischer, 2011), because the estimates are derived using a sufficiently large number of validation samples collected through stratified sampling. This assumption simplifies statistical analysis and ensures consistent uncertainty quantification across different datasets and periods.

For each analysis period, paired random samples of built-up area estimates—one for each year—were generated to compute LCR values, which were then used to derive LCRPGR. The PGR was calculated from census-based population data and was assumed to be error-free. The mean, standard deviation, and 95% confidence intervals of LCR and LCRPGR were computed from these simulations and compared across datasets and periods.

4. Results and Discussion

4.1 Impact of Built-Up Area Definitions on LCR and LCRPGR

Figure 3a illustrates trends in built-up areas across the Philippines, derived from multiple global EO data products, each employing distinct definitions of 'built-up area.' All products consistently show an upward trend in the built-up area over time. However, considerable disagreements are evident. The WSF-Evo dataset reports the highest total built-up area (categorized as 'settlements'), followed by GISD30, which captures impervious surfaces. Other data products estimate considerably lower built-up extents.

Compared to GHS-BUILT-S, GAUD and GISA 2.0 exhibit smaller differences in built-up areas, particularly in later years. On average, GAUD and GISA 2.0 differ by 16% and 7%, respectively, making them the closest datasets to GHS-BUILT-S in terms of mapped built-up area. In contrast, GISD30 and WSF-Evo show larger differences, consistently yielding more built-up area than GHS-BUILT-S. WSF-Evo exhibits the largest differences, exceeding 200% in all years, peaking at 221% in 1985 and averaging 205% more built-up area.

When these data products are used for LCR calculations, significant differences are observed in the earlier years, with variations decreasing in more recent periods (2005-2010 and 2010-2015) (Figure 3b). GISA 2.0 estimates indicate substantial land consumption in 1985-1990, while other datasets suggest moderate levels. GAUD and GISA 2.0 report higher LCR values than the other datasets in 1990-1995. LCR estimates from 1995 to 2005 vary widely across datasets, with GHS-BUILT-S and WSF-Evo reporting the lowest values. Beginning in 2005, LCR values across all datasets fall below 3%, indicating a general decline in land consumption rates.

Utilizing these LCR estimates for LCRPGR calculations resulted in significant discrepancies in LCRPGR values (Figure 3c). The patterns observed for LCR are the same for LCRPGR because the PGR values used in the calculation remain the same for all products each year. In 1985-1990, LCRPGR values calculated using GAUD, GISD30, and WSF-Evo were below 1, indicating efficient land use. However, the other EO products suggest inefficiency during the same period. These differences persist in subsequent years. For example, starting from 1990, all data products—except GHS-BUILT-S—indicate inefficient land utilization in the Philippines (LCRPGR >1).

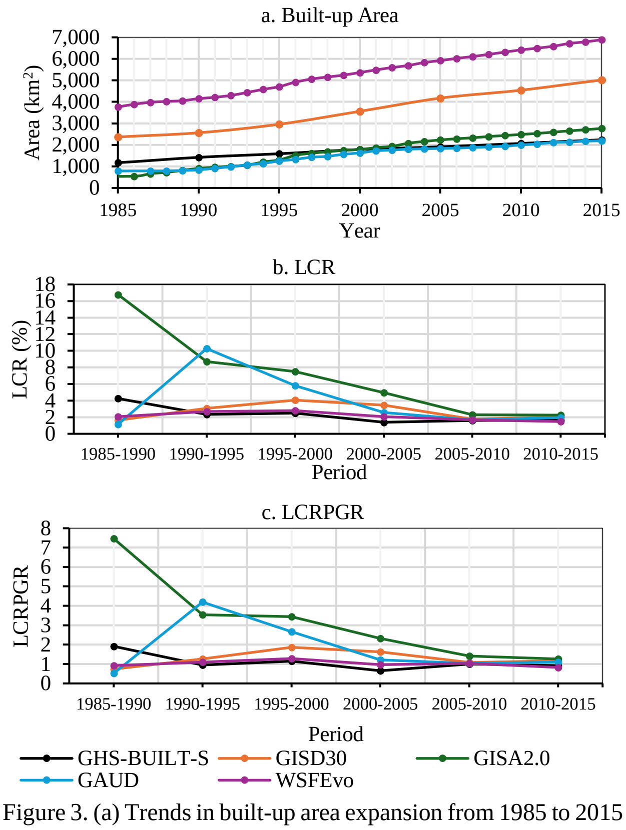

Figure 3. (a) Trends in built-up area expansion from 1985 to 2015 based on various global EO data products. (b) LCR and (c) LCPRGR at five-year intervals, derived from built-up area estimates of selected EO datasets. All graphs represent data for the Philippines.

Compared to GHS-BUILT-S as the baseline, GAUD underestimates both LCR and LCRPGR by up to 73% in 1985-1990 but later overestimates by up to 338% in 1990-1995, averaging 84% higher than the overall values. GISA 2.0 consistently overestimates, with values 294% higher in 1985-1990 and on average 183% higher than GHS-BUILT-S across all periods. GISD30 initially underestimates LCR and LCRPGR by 61% (1985-1990) but later overestimates by 149% (2000-2005), averaging 36% higher than GHS-BUILT-S. In contrast, WSF-Evo remains the closest, with differences primarily within ±50% and on average of just about 3% higher overall. These findings suggest that datasets with broader built-up definitions (e.g., GISA 2.0, GISD30) can more than double LCR and LCRPGR estimates, while those with stricter definitions (GAUD) can underestimate in some years and overestimate in others. However, beyond definitions and total built-up area differences, datasets depicting similar rates or patterns of built-up change can still produce comparable LCR and LCRPGR estimates. Despite WSF-Evo reporting a much larger built-up area than GHS-BUILT-S, their LCR and LCRPGR remain close, suggesting that the rate of change in built-up areas is more robust than the absolute extent.

Although this analysis focused solely on built-up area definitions and used mapped built-up areas without accounting for data accuracy and uncertainties, the findings demonstrate how dataset choice can significantly influence SDG 11.3.1 metric interpretations. The variations in LCR and LCRPGR estimates highlight the potential for divergent conclusions about urban growth and LUE depending on the EO product used. These discrepancies underscore the need for higher consistency in built-up area definition to ensure comparability across geographic contexts, periods, and datasets.

4.2 Accuracy of EO Data Products, Built-up Area Estimates, and Their Impact on LCR and LCRPGR

This section presents the accuracy assessment results of EO data products, their built-up area estimates, and the resulting uncertainties in LCR and LCRPGR calculations using Manila, Philippines, as the study area.

4.2.1 Case 1 - Using the data product's original built-up class definition: Table 5 summarizes the accuracy assessment based on the original class definitions of each dataset. In general, the datasets differ in accuracy, with no single dataset consistently outperforming the others across all metrics and years. While all datasets achieve relatively high Overall Accuracy (OA), exceeding 0.85 in most cases, there are variations in their User's Accuracy (UA) and Producer's Accuracy (PA). WSF-Evo exhibits the highest PA in multiple years, capturing most built-up areas with minimal omission errors. In contrast, GAUD generally has higher UA, indicating better precision in avoiding false positives. GISA 2.0 and GISD30 show consistently strong accuracy metrics across all years, with GISA 2.0 (2015) achieving the highest OA (0.91).

The results in Table 6 reveal that mapped built-up areas consistently overestimate the bias-adjusted built-up areas across all EO data products and periods. WSF-Evo reports the largest mapped built-up areas in all years, peaking at 40.91 km² in 2015, while GISD30 and GAUD estimate smaller extents. However, after bias adjustment, the built-up areas decrease across all datasets, indicating that mapped built-up areas include classification errors that inflate the estimates. Across all years, WSF-Evo exhibits the largest overestimation of built-up area, averaging 4.43 km² (12.29%) above the adjusted estimates. GISA 2.0 and GISD30 also overestimate built-up areas, with average differences of 3.48 km² (9.70%) and 2.86 km² (7.96%), respectively. GAUD has the smallest overestimation, averaging 1.25 km² (3.48%) across all years.

Across all datasets and years, the average overestimation in built-up area estimates is 1.25 km², while the average uncertainty amounts to ±0.67 km². WSF-Evo has the narrowest 95% confidence interval (±0.56 km² on average across years), likely

https://doi.org/10.5194/isprs-annals-X-G-2025-753-2025

due to its larger number of ground truth samples, resulting in more precise estimates than the other datasets. Despite variations in overestimation, the 95% confidence intervals of the bias-adjusted built-up areas overlap across all datasets for each year (Figure 4a), indicating no strong statistical evidence that the adjusted built-up areas differ significantly. A similar pattern emerges when examining each dataset's adjusted built-up areas and confidence intervals over time. For most of the datasets, their confidence intervals overlap across years, indicating no statistically significant change in adjusted built-up areas. However, GISA 2.0 stands out, as its confidence intervals do not overlap between 2000 and 2010, 2000 and 2015, 2005 and 2010, and 2005 and 2015. This lack of overlap suggests a statistically significant increase in built-up area (or, according to the dataset's definition, impervious surface area) as detected by GISA 2.0 over these periods. The observed changes likely reflect actual impervious surface expansion rather than those due to errors in classification.

Accuracy Metric	Year	GAUD	GISA 2.0	GISD30	WSF-Evo
User's Accuracy	2000	**0.91**	0.89	0.91	0.87
	2005	0.89	0.87	**0.91**	0.87
	2010	0.91	**0.92**	0.89	0.89
	2015	**0.92**	**0.92**	0.89	0.89
Producer's Accuracy	2000	0.94	**0.99**	0.96	**0.99**
	2005	0.93	**0.99**	0.97	**0.99**
	2010	0.94	0.99	0.98	**1.00**
	2015	0.95	**0.99**	0.98	**0.99**
Overall Accuracy	2000	0.87	0.88	**0.89**	0.87
	2005	0.85	0.87	**0.88**	0.87
	2010	0.87	**0.91**	0.88	0.90
	2015	0.88	**0.91**	0.89	0.89

Table 5. Accuracy assessment of EO data products for Manila, Philippines, based on their original built-up area definitions. Bold values indicate the highest accuracy for each year.

Type	Year	GAUD	GISA 2.0	GISD30	WSF-Evo
Mapped Area (km^2)	2000	36.97	39.32	38.38	40.08
	2005	37.36	39.59	38.78	40.48
	2010	37.52	39.79	39.19	40.59
	2015	37.87	39.92	39.39	40.91
Bias-adjusted Area with Uncertainties (km^2)	2000	35.54 ± 0.77	35.36 ± 0.75	36.41 ± 0.64	35.46 ± 0.59
	2005	35.65 ± 0.81	34.98 ± 0.79	35.45 ± 0.68	35.65 ± 0.59
	2010	36.54 ± 0.75	37.07 ± 0.66	35.7 ± 0.68	36.43 ± 0.53
	2015	36.99 ± 0.71	37.28 ± 0.65	36.73 ± 0.63	36.8 ± 0.54

Table 6. Summary of mapped and bias-adjusted built-up areas estimated from EO data products for Manila, Philippines, using their original built-up class definitions. Uncertainties are expressed as 95% confidence intervals.

Figure 4b shows the LCR calculated using bias-adjusted built-up area estimates from different EO data products, based on their original built-up area class definitions with 95% confidence intervals derived from Monte Carlo simulations, while the same metric calculated using the mapped built-up area are listed in Table 7. While LCR values derived from mapped built-up areas indicate consistently low land consumption over time (all below 0.22%), this trend changes when bias-adjusted areas are used in the calculations. LCR values can range from as low as about -1% (GISD30, 2000–2005) to as high as about 1.75% (GISA 2.0, 2005–2010), reflecting the impact of built-up area uncertainties on LCR estimates. On average, across datasets and years, the uncertainties in LCR amounted to ±0.54%.

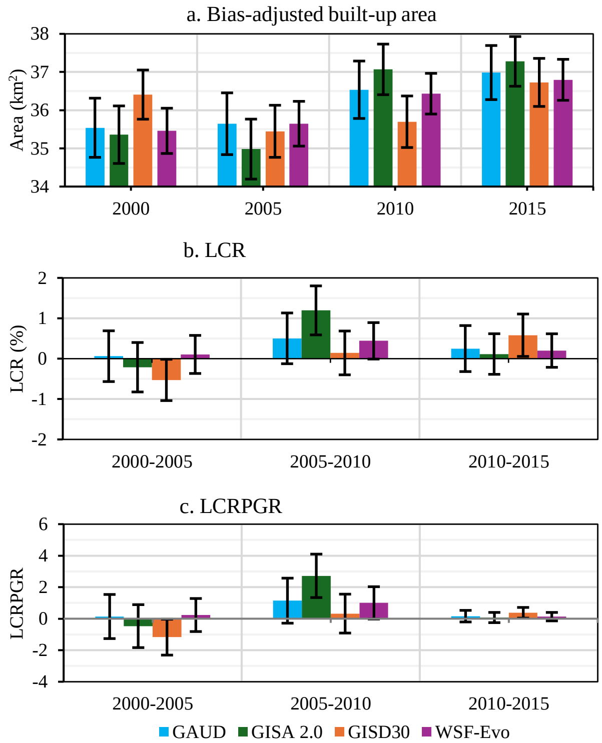

Figure 4. (a) Bias-adjusted built-up area estimates from different EO data products across selected years (2000–2015), based on their original built-up area class definitions. (b) LCR and (c) LCRPGR, derived from bias-adjusted built-up area estimates across selected periods. All graphs represent data for Manila, Philippines. Error bars indicate the 95% confidence intervals.

Metric	Period	GAUD	GISA 2.0	GISD30	WSF-Evo
LCR (%)	2000-2005	0.21	0.14	0.21	0.20
	2005-2010	0.09	0.10	0.21	0.06
	2010-2015	0.19	0.07	0.10	0.16
PGR (%)	2000-2005	0.44			
	2005-2010	0.44			
	2010-2015	1.49			
LCRPGR	2000-2005	0.48	0.32	0.48	0.44
	2005-2010	0.20	0.22	0.48	0.13
	2010-2015	0.13	0.05	0.07	0.10

Table 7. LCR and LCPGR derived from mapped built-up areas estimated using EO data products for Manila, Philippines. For reference, PGR is included, as calculated from census data.

On the other hand, LCRPGR values (Table 7) computed using mapped built-up areas remain below 1 across different periods and datasets, indicating efficient LUE—where population growth outpaces land consumption. However, this pattern shifts when bias-adjusted built-up areas and their uncertainties are considered (Figure 4c). Except for the 2010–2015 period, where most of the LCRPGR values from mapped and bias-adjusted built-up areas are relatively consistent, the 2000–2005 and 2005–2010 results suggest more significant uncertainty in LUE dynamics. Depending on the EO dataset, LCRPGR estimates can vary widely when uncertainties are considered (average of 0.39 ± 0.94 across datasets and periods). The least uncertain estimates are found for the 2010-2015 period. Moreover, LCR and LCRPGR values derived from bias-adjusted built-up areas are statistically similar across datasets at the 95% confidence level, following from the overlapping confidence intervals of their bias-adjusted built-up area estimates (as depicted in Figure 4a). This suggests that, despite variations in built-up area definitions, the land

https://doi.org/10.5194/isprs-annals-X-G-2025-753-2025

consumption rate can remain comparable across datasets after accounting for classification uncertainties. However, while these LCR and LCRPGR values may be statistically similar across datasets or years, a clear differentiation is necessary when interpreting them, as they are inherently tied to each dataset's built-up area definition. For example, LCR and LCRPGR derived from GISA 2.0 should be explicitly called "impervious surface area LCR and PGR." At the same time, the LCR and LCRPGR from WSF-Evo should be identified as "settlement LCR and LCRPGR" to accurately reflect the differences in the products' built-up class definition.

4.2.2 Case 2 - Using the SDG 11.3.1 built-up area definition: Table 8 summarizes the accuracy assessment of the EO data products based on the SDG 11.3.1 built-up area class definition. The results reveal variability in how well each EO dataset aligns with this definition. WSF-Evo exhibits the highest PA across all years (100% in 2005, 2010, and 2015), consistently capturing most reference built-up areas. However, the PA values for other datasets remain high (96% and above), suggesting that all datasets effectively identify areas occupied by buildings. GAUD generally achieves the highest UA, though the differences between datasets are minor. Regarding OA, GAUD performs best across all years, but other datasets remain comparable, with only a few percentage points lower accuracy. UA and OA are significantly lower when assessed using the SDG 11.3.1 definition compared to each dataset's original built-up class definition. On average, UA decreased by approximately 25%, while OA dropped by about 22% across all datasets and years.

Accuracy Metric	Year	GAUD	GISA 2.0	GISD30	WSF-Evo
User's Accuracy	2000	**0.71**	0.67	0.68	0.65
	2005	0.67	**0.68**	0.67	0.67
	2010	**0.68**	0.66	0.63	0.65
	2015	**0.70**	0.68	0.69	0.67
Producer's Accuracy	2000	0.96	**0.99**	0.97	**0.99**
	2005	0.98	0.99	0.98	**1.00**
	2010	0.97	0.99	0.99	**1.00**
	2015	0.98	0.99	0.99	**1.00**
Overall Accuracy	2000	**0.72**	0.68	0.69	0.66
	2005	**0.70**	0.69	0.69	0.68
	2010	**0.69**	0.68	0.65	0.66
	2015	**0.71**	0.69	0.70	0.67

Table 8. Accuracy assessment of EO data products for Manila, Philippines, evaluated using the SDG 11.3.1 definition of the built-up area as "all areas occupied by buildings." Bold values indicate the highest accuracy for each year.

Built-up area estimates (Figure 5a) from 2000 to 2010 exhibit high variability across datasets, reflecting differences in classification and bias adjustments. However, by 2015, the estimates become more consistent, suggesting better alignment among datasets in capturing built-up areas in the most recent period. On average, across all datasets and years, the built-up area is estimated at 26.71 km², with an average uncertainty of ±0.99 km². This estimate is approximately 12.42 km² lower than the mapped built-up area, indicating an overestimation of about 42% when EO datasets are used without bias adjustment. The average uncertainty is approximately 46% higher than when the datasets were assessed using their original built-up class definitions. The overlapping confidence intervals suggest that uncertainties in built-up area estimates remain substantial, limiting the ability to differentiate between datasets and detect significant trends over time. Despite the high uncertainties, the findings suggest that estimating built-up areas according to the SDG 11.3.1 definition remains feasible across EO datasets.

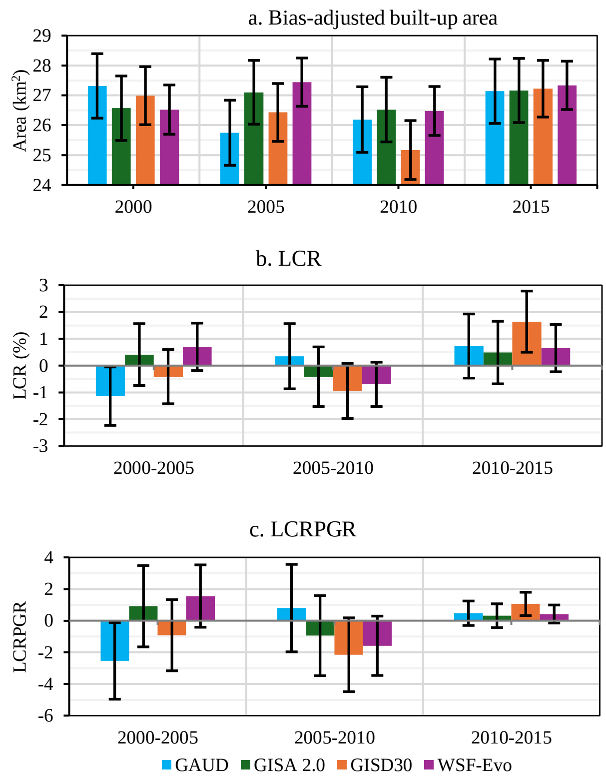

Figure 5. (a) Bias-adjusted built-up area estimates from different EO data products across selected years (2000–2015), based on the SDG 11.3.1 built-up area definition. (b) LCR and (c) LCRPGR, derived from bias-adjusted built-up area estimates across selected periods. All graphs represent data for Manila, Philippines. Error bars indicate the 95% confidence intervals.

The LCR and LCRPGR estimates derived from bias-adjusted built-up area data (Figure 5b & c) reveal notable trends and substantial uncertainties. In the earlier periods (2000–2005 and 2005-2010), most datasets report negative LCR and LCRPGR values, whereas positive estimates appear for 2010–2015. However, in both cases, uncertainties are substantial. LCR uncertainties range from ±0.83% (WSF-Evo, 2005–2010) to ±1.22% (GAUD, 2005–2010). Among the datasets, GAUD exhibits the highest LCR uncertainty (average of ±1.17%), while the other datasets show moderate uncertainty levels, averaging ±0.52%. Across all datasets and years, the average uncertainty is ±0.68%, notably higher than when using the original built-up class definitions. LCRPGR uncertainty is also substantial, with estimates varying by ±1.80 across datasets and years. However, uncertainties are significantly lower for the most recent period (2010–2015), averaging ±0.71 across datasets. This suggests that EO datasets provide more precise LUE estimates for this period than in earlier years—a trend also observed when using the dataset's original built-up class definitions.

Comparing Figure 4 with Figure 5, LCR and LCRPGR temporal trends vary significantly in magnitude and direction depending on the built-up area definition. For example, during 2005–2010, the original class definition produces positive LCR and LCRPGR values, indicating high land consumption rates and inefficient land use practices. However, when using the SDG 11.3.1 built-up definition, the trend shifts in the opposite direction, suggesting a lower rate of land consumption or even contraction of built-up areas. This discrepancy highlights that built-up area development rates differ based on the classification approach, affecting interpretations of urban growth and LUE.

5. Summary, Conclusions and Outlook

This study quantifies the impact of EO data variability on LCR and LCRPGR estimates, demonstrating how inconsistencies in built-up area definitions and dataset accuracy contribute to significant discrepancies in LUE assessment. To mitigate these discrepancies, we implement a bias-adjustment methodology that corrects classification errors, aligning EO-derived built-up area estimates more closely with the official SDG 11.3.1 definition. Additionally, we incorporate uncertainty quantification using Monte Carlo simulations, ensuring that LCR and LCRPGR estimates more accurately reflect the inherent uncertainties in LUE metrics.

The substantial variability observed in LCR and LCRPGR estimates across different datasets and periods underscores the sensitivity of these metrics to built-up area definitions and classification uncertainties. These variations highlight the potential for divergent interpretations of urban expansion trends and LUE assessment, depending on the EO dataset utilized. Such discrepancies can lead to inconsistencies in SDG 11.3.1 monitoring and, consequently, in policy recommendations. The fact that LCR and LCRPGR values can change not only in magnitude but also in direction depending on whether mapped or bias-adjusted built-up areas are used reinforces the critical need for methodological transparency and careful dataset selection in LUE assessments.

To enhance the reliability of SDG 11.3.1 monitoring, a standardized approach to defining and measuring built-up areas is essential. Strengthening adherence to the SDG 11.3.1 definition and promoting robust methodological frameworks will improve the comparability and credibility of LUE assessments. The application of the post-stratified estimator provides a more accurate basis for LCR and LCRPGR calculations by correcting classification errors in EO-derived built-up area estimates. Meanwhile, Monte Carlo-based uncertainty quantification reveals the high sensitivity of these metrics to input data accuracy, reinforcing the need for uncertainty-aware methodologies in urban sustainability evaluations. These methodological improvements can equip policymakers and urban planners with more reliable insights, facilitating more effective urban development strategies and informed decision-making.

Despite these contributions, some limitations warrant further investigation. This study does not explicitly account for potential uncertainties introduced by temporal mismatches between reference imagery and the EO data products being assessed. Such mismatches may influence the accuracy of validation samples and bias adjustments, potentially affecting the robustness of LUE estimates. Additionally, the generalizability of these findings could be enhanced by testing the methodology across different urban contexts and EO data products with various spatial and temporal resolutions. While this study assumes population data to be error-free, this is not always the case. Uncertainties in population data can also significantly impact PGR and LCRPGR estimates, and future studies should incorporate their quantification. Future work could also refine uncertainty quantification by incorporating spatially explicit error models to account for localized (pixel-level) classification uncertainties. Expanding this approach to other regions and urban typologies would also support the creation of standardized, uncertainty-aware SDG 11.3.1 monitoring frameworks, allowing for more equitable and actionable comparisons in urban sustainability assessments.

Acknowledgements

This work was supported by the Philippines' Department of Science and Technology – Science Education Institute (DOST-SEI) through a doctoral scholarship and by Caraga State University through a fellowship, both awarded to J. R. Santillan. We sincerely thank Meriam Makinano-Santillan and Jasper Obedencio for their invaluable assistance in ground truth sample labeling.

References

Albert, D. R., 2020. Monte Carlo uncertainty propagation with the NIST uncertainty machine. *Journal of Chemical Education*, 97(5), 1491–1494. doi.org/10.1021/acs.jchemed.0c00096.

Alkhatib, H., Neumann, I., Kutterer, H., 2009. Uncertainty modeling of random and systematic errors by means of Monte Carlo and fuzzy techniques. *Journal of Applied Geodesy*, 3(2). doi.org/10.1515/JAG.2009.008.

Angelopoulos, A. N., Bates, S., Fannjiang, C., Jordan, M. I., Zrnic, T., 2023. Prediction-powered inference. *Science*, 382(6671), 669–674. doi.org/10.1126/science.adi6000.

Cai, G., Zhang, J., Du, M., Li, C., Peng, S., 2020. Identification of urban land use efficiency by indicator-SDG 11.3.1. *PLoS ONE*, 15(12 December). doi.org/10.1371/journal.pone.0244318.

Cardenas-Ritzert, O. S. E., Vogeler, J. C., Shah Heydari, S., Fekety, P. A., Laituri, M., McHale, M., 2024. Automated geospatial approach for assessing SDG indicator 11.3.1: A multi-level evaluation of urban land use expansion across Africa. *ISPRS International Journal of Geo-Information*, 13(7). doi.org/10.3390/ijgi13070226.

Chakraborty, T., Venter, Z. S., Demuzere, M., Zhan, W., Gao, J., Zhao, L., Qian, Y., 2024. Large disagreements in estimates of urban land across scales and their implications. *Nature Communications*, 15(1), 9165. doi.org/10.1038/s41467-024-52241-5.

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

European Commission, 2023. GHSL Data Package 2023. Publications Office of the European Union. doi.org/10.2760/098587.

Fischer, H., 2011. *A History of Central Limit Theorem from Classical to Modern Probability Theory*. Springer, New York.

Foody, G. M., 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80(1), 185–201. doi.org/10.1016/S0034-4257(01)00295-4.

Ghazaryan, G., Rienow, A., Oldenburg, C., Thonfeld, F., Trampnau, B., Sticksel, S., Jürgens, C., 2021. Monitoring of urban sprawl and densification processes in western Germany in the light of SDG indicator 11.3.1 based on an automated retrospective classification approach. *Remote Sensing*, 13(9). doi.org/10.3390/rs13091694.

Ghilani, C. D., 2017. *Adjustment Computations: Spatial Data*

Analysis. John Wiley Sons, New Jersey.

Gong, Y., Xie, H., Tong, X., Jin, Y., Xv, X., Wang, Q., 2020. Area estimation of multi-temporal global impervious land cover based on stratified random sampling. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B4-2020, 103–108. doi.org/10.5194/isprs-archives-XLIII-B4-2020-103-2020.

Huang, M., Liu, F., Gong, D., Lin, H., Chen, Y., Hu, B., Ge, Y., Xiao, C., 2024. Spatiotemporal evolution of land use efficiency in 357 cities across mainland China from 2000 to 2020 based on SDG 11.3.1. *Science of The Total Environment*, 954, 176157. doi.org/10.1016/j.scitotenv.2024.176157.

Huang, X., Song, Y., Yang, J., Wang, W., Ren, H., Dong, M., Feng, Y., Yin, H., Li, J., 2022. Toward accurate mapping of 30-m time-series global impervious surface area (GISA). *International Journal of Applied Earth Observation and Geoinformation*, 109, 102787. doi.org/10.1016/j.jag.2022.102787.

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

Li, W., El-Askary, H., Lakshmi, V., Piechota, T., Struppa, D., 2020. Earth observation and cloud computing in support of two sustainable development goals for the River Nile watershed countries. *Remote Sensing*, 12(9), 1391. doi.org/10.3390/rs12091391.

Li, C., Cai, G., Du, M., 2021. Big data supported the identification of urban land efficiency in Eurasia by indicator SDG 11.3.1. *ISPRS International Journal of Geo-Information*, 10(2). doi.org/10.3390/ijgi10020064.

Liu, X., Hu, G., Chen, Y., Li, X., Xu, X., Li, S., Pei, F., Wang, S., 2018. High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine platform. *Remote Sensing of Environment*, 209, 227–239. doi.org/10.1016/j.rse.2018.02.055.

Liu, X., Huang, Y., Xu, X., Li, X., Li, X., Ciais, P., Lin, P., Gong, K., Ziegler, A. D., Chen, A., Gong, P., Chen, J., Hu, G., Chen, Y., Wang, S., Wu, Q., Huang, K., Estes, L., Zeng, Z., 2020. High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015. *Nature Sustainability*, 3(7), 564–570. doi.org/10.1038/s41893-020-0521-x.

Lu, K., Bates, S., Wang, S., 2024. Quantifying uncertainty in area and regression coefficient estimation from remote sensing maps. *arXiv:2407.13659v3*. https://arxiv.org/abs/2407.13659v3 (10 October 2024).

Marconcini, M., Metz- Marconcini, A., Esch, T., Gorelick, N., 2021. Understanding current trends in global urbanisation - The World Settlement Footprint Suite. *GI_Forum*, 1, 33–38. doi.org/10.1553/giscience2021_01_s33.

Marconcini, M., Metz-Marconcini, A., Üreyen, S., Palacios-Lopez, D., Hanke, W., Bachofer, F., Zeidler, J., Esch, T., Gorelick, N., Kakarla, A., Paganini, M., Strano, E., 2020. Outlining where humans live, the World Settlement Footprint 2015. *Scientific Data*, 7(1), 242. doi.org/10.1038/s41597-020-00580-5.

MMEIRS, 2004. Earthquake Impact Reduction Study for Metropolitan Manila, Republic of the Philippines. https://openjicareport.jica.go.jp/216/216/216_118_11763745.html (5 October 2024).

Mustafa, A., Saadi, I., Cools, M., Teller, J., 2018. A Time Monte Carlo method for addressing uncertainty in land-use change models. *International Journal of Geographical Information Science*, 32(11), 2317–2333. doi.org/10.1080/13658816.2018.1503275.

Olofsson, P., Foody, G. M., Stehman, S. V., Woodcock, C. E., 2013. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, 129, 122–131. doi.org/10.1016/j.rse.2012.10.031.

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., Wulder, M. A., 2014. Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148, 42–57. doi.org/10.1016/j.rse.2014.02.015.

Santillan, J. R., Heipke, C., 2024. Assessing patterns and trends in urbanization and land use efficiency across the Philippines: A comprehensive analysis using global earth observation data and SDG 11.3.1 indicators. *PFG - Journal of Photogrammetry, Remote Sensing and Geoinformation Science*. doi.org/10.1007/s41064-024-00305-y.

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

Stehman, S. V., 2013. Estimating area from an accuracy assessment error matrix. *Remote Sensing of Environment*, 132, 202–211. doi.org/10.1016/j.rse.2013.01.016.

UN Statistics Division, 2021. SDG indicator metadata (Harmonized metadata template – format version 1.0). Last updated 2021-03-01. https://unstats.un.org/sdgs/metadata/files/Metadata-11-03-01.pdf (15 October 2024).

United Nations, 2019. World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420). United Nations. https://digitallibrary.un.org/record/3833745 (15 October 2024).

Valle, D., Izbicki, R., Leite, R. V., 2023. Quantifying uncertainty in land-use land-cover classification using conformal statistics. *Remote Sensing of Environment*, 295, 113682. doi.org/10.1016/j.rse.2023.113682.

Zhang, X., Liu, L., Zhao, T., Gao, Y., Chen, X., Mi, J., 2022. GISD30: Global 30 m impervious-surface dynamic dataset from 1985 to 2020 using time-series Landsat imagery on the Google Earth Engine platform. *Earth System Science Data*, 14(4), 1831–1856. doi.org/10.5194/essd-14-1831-2022.