ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume X-5/W4-2025
Philippine Geomatics Symposium (PhilGEOS) 2025 "Enhancing Human Quality of Life through Geospatial Technologies",
24–25 November 2025, Quezon City, Philippines

A Deep Learning Framework for Forecasting Built-Up Area and Population to Support Land Use Efficiency Projections under SDG 11.3.1

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: Land Use Efficiency, SDG 11.3.1, Probabilistic Time Series Forecasting, Deep Learning, Urban Growth.

Abstract

Sustainable Development Goal (SDG) 11.3.1 monitors land use efficiency (LUE) through the relationship between the land consumption rate (LCR) and the population growth rate (PGR), yet current applications have largely remained retrospective and offer limited foresight for policy and planning. To address this gap, this study develops a scalable probabilistic deep learning (DL)-based framework that jointly forecasts built-up area (BU) and population (Pop) and propagates these forecasts into LCR, PGR, and their ratio (LCRPGR). The framework is based on a lightweight hybrid model composed of one-dimensional convolutional neural networks (Conv1D) and a recurrent network with Long Short-Term Memory (LSTM) units, trained using quantile loss to produce interval-bounded probabilistic forecasts. Ensemble aggregation was used to improve robustness, with a calibration stage included to ensure reliable coverage of prediction intervals. Applied to 8,478 urban centres using historical built-up area and population data from the Global Human Settlement Layer – Urban Centre Database (GHS-UCDB) 2025 edition, the framework generated probabilistic forecasts for the period 2025-2030, with prediction intervals shown to be well-calibrated on historical data. Results indicate that total built-up area is projected to expand from 108,808 km² in 2020 to 118,123 km² by 2030 (+8.6%), while population is expected to grow from 2.875 to 3.159 billion (+9.9%). At the aggregated level, LCRPGR values suggest a general trend toward efficiency, with PGR slightly outpacing LCR, though uncertainty intervals cross the efficiency threshold. While the implementation addressed the constraints of SDG 11.3.1 forecasting (short time series, limited covariates, and large-scale application), the framework remains flexible. It can be extended with denser time series, richer covariates, and spatiotemporal architectures. This study offers a practical basis for near-term LUE monitoring and risk-aware policy action as 2030 approaches by providing reproducible and uncertainty-aware forecasts.

1. Introduction

Sustainable urbanization has become a global priority due to rapid population growth and urban expansion. SDG Target 11.3 seeks to promote efficient land use (UN-Habitat, 2018), monitored through Indicator 11.3.1. This indicator measures land use efficiency (LUE) as the ratio of the land consumption rate (LCR) to the population growth rate (PGR), commonly referred to as LCRPGR (UN Statistics Division, 2021). Values above one indicate inefficient expansion, while values at or below one reflect more efficient use.

With the 2030 SDG deadline approaching, reliable forecasts of built-up area and population are essential for projecting LCR, PGR, and LCRPGR. Yet forecasting faces several challenges: Earth observation (EO)-derived built-up data and census-based population figures are available only at sparse intervals, producing short time series; forecasts must be generated for thousands of diverse urban areas; and built-up expansion and population growth are interdependent, varying by geography, development stage, and urban form (Schiavina et al., 2022a). Urban trajectories are also shaped by regional spillovers, where growth in one city is linked to neighbouring cities (Wang et al., 2022). Addressing these issues requires a generalized forecasting approach that can jointly model built-up area and population, account for spatial interdependencies, and quantify uncertainties, enabling robust projections of LUE.

Advances in artificial intelligence, particularly in deep learning (DL), provide new opportunities to address the forecasting challenges of SDG 11.3.1. DL models excel at capturing complex, nonlinear, and multivariate dynamics, making them well-suited for analyzing urban processes. Recent reviews (e.g., Benidis et al., 2023; Li and Law, 2024) highlight their strong predictive performance across diverse domains, including finance, energy, health, environmental monitoring, and mobility. Yet, their potential for jointly forecasting built-up area and population, and by extension projecting LCR, PGR, and LCRPGR across thousands of cities worldwide, remains largely underexplored. At the same time, uncertainty quantification through probabilistic techniques has become a central requirement in modern forecasting, ensuring that predictions are accurate and reliable for decision-making (Abdar et al., 2021). However, these techniques have limited application to urban indicators and are not yet integrated into the operational SDG 11.3.1 monitoring framework.

Building on the identified issues and gaps, this paper introduces a framework for jointly forecasting built-up area and population to support SDG 11.3.1-based LUE assessments. The framework integrates a probabilistic DL-based time series forecasting approach with post hoc uncertainty calibration to produce reliable, interval-bounded forecasts that can be propagated into LCR, PGR, and LCRPGR estimates. This is the first systematic attempt to embed such a framework into SDG 11.3.1 monitoring, advancing beyond retrospective analyses and scenario-based approaches toward data-driven and uncertainty-aware forecasts.

As a case example, we apply the proposed framework to global urban centres, projecting and assessing LUE for the period 2025–2030 using historical data from the Global Human Settlement Layer – Urban Centre Database (GHS-UCDB) 2025 edition (European Commission et al., 2024).

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-5-W4-2025-385-2026

2. Related Work

In this section, we review related work on SDG 11.3.1 monitoring and forecasting and recent advances in DL for time series prediction. It identifies key gaps in existing approaches and situates the contributions of this study within that context.

2.1 SDG 11.3.1 Monitoring and LUE Forecasting

Most SDG 11.3.1 applications remain retrospective, using EO data to compare built-up area and population across historical time points to compute LCRPGR and assess LUE (e.g., Estoque et al., 2021; Schiavina et al., 2022b). While useful for trend analysis, this offers little insight into future development. Only a few studies have attempted LUE forecasting, typically through scenario-based projections using Shared Socio-economic Pathways (SSPs) (e.g., Wang et al., 2024) or local policy constraints (e.g., Lu et al., 2022). These methods extrapolate trends with statistical or efficiency models but do not directly learn temporal dynamics from empirical data. These limitations highlight the need for approaches that can jointly forecast built-up area and population, thereby enabling data-driven projections of LCR, PGR, and LCRPGR to support SDG 11.3.1 monitoring.

2.2 Time Series Forecasting With Deep Learning

Time series forecasting methods are either univariate, relying on past values of a single variable, or multivariate, which capture dependencies among variables (Benidis et al., 2023). While univariate models work well for strongly autoregressive series, multivariate approaches are crucial where processes co-evolve, as in energy, finance, traffic, and demography. The joint forecasting of built-up area and population falls into this latter category, given their shared socio-economic and spatial drivers.

Traditionally, time series forecasting has relied on statistical approaches such as autoregressive integrated moving average (ARIMA) and exponential smoothing, as well as classical machine learning approaches such as support vector machine (SVM) and random forest (RF) (Fatima and Rahimi, 2024; Li and Law, 2024; Masini et al., 2023). While useful for broad trend detection, these methods struggle with nonlinear, multivariate, and heteroscedastic dynamics and lack scalability for global applications. They also capture temporal dependencies poorly, limiting their forecasting performance.

Deep learning (DL) models address these limitations by learning flexible, data-driven representations of temporal dynamics that capture nonlinear relationships and heteroscedastic patterns while scaling across thousands of localities (Li and Law, 2024; Sen et al., 2019). Unlike classical models with fixed functional forms, DL models learn hierarchical features directly from data, resulting in global models that generalize across heterogeneous urban contexts.

Different DL architectures have been developed for time series forecasting, each with strengths and trade-offs (Li and Law, 2024). Early Recurrent Neural Networks (RNNs) processed sequences through hidden states but struggled with vanishing gradients, leading to the development of Long Short-Term Memory (LSTM) networks with gating mechanisms (Hochreiter and Schmidhuber, 1997) and the more compact Gated Recurrent Unit (GRU) (Cho et al., 2014). These recurrent foundations underpin models such as DeepAR, which outputs probabilistic forecasts for univariate series (Salinas et al., 2020). Convolutional approaches treat time series as one-dimensional signals, with Conv1D capturing local patterns and Temporal Convolutional Networks (TCNs) extending this with dilated convolutions for longer dependencies (Bai et al., 2018). Hybrid models such as Conv1D/TCN+LSTM or Conv1D/TCN+GRU combine the strengths of convolutional layers to detect local temporal patterns, and of the recurrent units to capture broader temporal dynamics. Transformer-based models replace recurrence with attention to capture dependencies across time steps (Vaswani et al., 2017), with variants like the Temporal Fusion Transformer (TFT) adding interpretability and support for heterogeneous inputs (Lim et al., 2021). Graph Neural Networks (GNNs) extend forecasting to spatiotemporal settings, jointly modeling temporal dynamics and spatial interdependencies (Jin et al., 2024).

DL architectures have shown strong performance across finance, energy, health, traffic, and mobility (Benidis et al., 2023; Li and Law, 2024). However, urban applications remain limited and primarily target a single variable. Examples include LSTM-based models for small-area and gridded population forecasting (Grossman et al., 2023; Geiß et al., 2023) and the Population Fusion Transformer for subnational population forecasting (Alghanmi et al., 2024). Beyond their limited scope, another challenge is that most existing models are complex, trained on dense, covariate-rich data, and tailored to local contexts. Their scalability and suitability for the short, sparse, and heterogeneous series typical of built-up and population data remain uncertain, highlighting the need to adapt and evaluate DL models for urban forecasting within the constraints of SDG 11.3.1.

2.3 Probabilistic Forecasting and the Need for Calibration

DL models can be designed or adapted to produce predictive distributions by modifying the loss function or output structure. For example, DeepAR estimates Gaussian mean and variance via negative log-likelihood (Salinas et al., 2020), while the TFT predicts quantiles directly (Lim et al., 2021). Other approaches include deep ensembles (Lakshminarayanan et al., 2017), Monte Carlo dropout (Gal and Ghahramani, 2016), and likelihood-based methods (Sluijterman et al., 2024).

A challenge is that prediction intervals (PIs) from these probabilistic DL models can be miscalibrated. Post hoc methods such as conformal prediction (CP) correct this, offering distribution-free coverage guarantees under exchangeability (Angelopoulos and Bates, 2023), though this assumption is often violated in nonstationary urban growth. Conformalized Quantile Regression (CQR) improves local adaptation (Romano et al., 2019), but still relies on the exchangeability assumption, while Ensemble Bootstrap Prediction Intervals (EnbPI) (Xu and Xie, 2023) relax this assumption for temporally dependent data but remain centered on point forecasts.

Ensemble Conformalized Quantile Regression (EnCQR) addresses these issues by combining ensembles of quantile regressors with conformalization using asymmetric residuals (Jensen et al., 2024). This approach yields sharper, more correct intervals under nonstationary and temporally dependent data. These properties make EnCQR particularly well-suited for DL models that already output quantiles. They are highly relevant for generating reliable built-up area and population forecasts that propagate into LCR, PGR, and LCRPGR projections for SDG 11.3.1 monitoring.

2.4 Summary of Research Gaps and Opportunities

Three main gaps emerge in SDG 11.3.1 forecasting. First, most applications are retrospective or scenario-based, with little direct

modeling of built-up–population dynamics from time series data. Second, existing DL models are designed for dense, feature-rich data, whereas built-up and population series are short, sparse, and covariate-poor, raising concerns about scalability. Third, it remains unclear how global DL models can jointly capture these dynamics across thousands of heterogeneous cities. Advances in probabilistic forecasting and calibration methods, such as conformal prediction, offer opportunities to address these gaps by delivering reliable, well-calibrated prediction intervals under heteroscedasticity, dependence, and non-stationarity conditions.

3. The DL-based Built-up Area and Population Forecasting Framework for LUE Projection

3.1 Overview

We propose a global probabilistic DL framework to jointly forecast built-up area (BU) and population (Pop) while providing calibrated uncertainty estimates for LUE projection. The framework is tailored to SDG 11.3.1 constraints (short time series (≤10 observations), limited covariates, and large-scale application across thousands of cities) by training a single global model that captures local variation and long-term co-evolution. Given these conditions, we adopt a simpler hybrid Conv1D+LSTM architecture rather than more complex models like DeepAR (Salinas et al., 2020) and TFT (Lim et al., 2021).

3.2 Model Architecture and Loss Function

The Conv1D+LSTM takes as input a multivariate sequence of length L time steps, where each step contains F features (here, BU and Pop) from N cities/urban centres. The model jointly predicts the next H time steps for both variables, producing quantile-based forecasts.

The first component of the model's architecture is a stack of one-dimensional convolutional layers, each with f_i filters and kernel size k_i. These layers apply dilations d_i to expand the receptive field, which enables the extraction of short-term local patterns across the input sequence. The convolutional outputs are passed through nonlinear activation functions (e.g., Rectified Linear Unit or ReLU) and optionally through dropout layers, which act as a regularization mechanism to reduce overfitting (Goodfellow et al., 2016). The convolutional block is followed by a unidirectional LSTM layer with u hidden units. The LSTM captures longer-range temporal dependencies and provides memory of prior states. The unidirectional design preserves causality, ensuring forecasts are based only on past information (Hochreiter and Schmidhuber, 1997). The LSTM outputs are passed through a multi-layer perceptron (fully connected layers with h hidden units), which maps them into the output space, with L_2 regularization applied to mitigate overfitting. The final output head predicts $H \times V \times Q$, where H is the forecast horizon, V is the number of variables, and Q is the number of quantiles, thereby generating distributional forecasts of future BU and Pop trajectories.

To make the model probabilistic, it is trained using the quantile (pinball) loss (Wang et al., 2019), which evaluates how closely the estimated quantile aligns with the true data distribution, with lower values indicating more accurate estimations (Jensen et al., 2024). The loss is averaged across variables, horizons, and quantiles during training. This enables the network to provide a median forecast and user-specified prediction intervals.

The model explains the BU-Pop relationship by jointly learning their temporal co-evolution: it captures how changes in one variable influence the other over time and embeds these dependencies into the forecast. The quantile outputs then provide expected trends and the range of plausible futures, making the results both predictive and interpretable.

3.3 Ensembling and Post hoc Calibration

Uncertainty is incorporated through a conditional two-stage strategy. In the first stage, multiple versions of the Conv1D+LSTM model are trained independently using different random weight initializations. This step follows the principle of deep ensembles (Lakshminarayanan et al., 2017), where diversity comes from stochastic initialization and optimization rather than from training on disjoint subsets. This adaptation is necessary as the short and sparse sequences in the dataset cannot be meaningfully partitioned. In a multi-city setting, each ensemble member produces quantile forecasts (e.g., at the 2.5th, 50th, and 97.5th levels), resulting in n forecasted quantiles for each city, where n is the number of ensemble members. For each case, these n forecasts are aggregated (e.g., by the median) into a single set of lower, median, and upper quantiles.

The aggregated intervals are next evaluated on held-out calibration data by computing the empirical prediction interval coverage probability (PICP) (Jensen et al., 2024). The raw intervals are retained to preserve sharpness if the PICP is already close to the nominal level (e.g., 95%). If not, a second stage of post hoc calibration is performed. In this stage, deviations between forecasts and observations are expressed as asymmetric residual-based conformity scores, which are used to shift the raw intervals following the Ensemble Conformalized Quantile Regression (EnCQR) method (Jensen et al., 2024). Specifically, the lower and upper bounds are adjusted by empirical quantiles of the conformity scores so that, across calibration cases, the empirical coverage matches the target level. This step corrects for systematic under- or over-coverage when present, while ensuring that the resulting intervals remain valid and adaptive.

3.4 Forecasting Built-up Area and Population for LUE Assessment

The proposed framework produces probabilistic forecasts of BU and Pop that can be directly used to estimate LUE. These forecasts at different epochs are combined to calculate the Land Consumption Rate (LCR), Population Growth Rate (PGR), and their ratio (LCRPGR) following the official UN methodology (UN Statistics Division, 2021):

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

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

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

where *BU* and *Pop* denote the total built-up area and population at the earlier (t_1) and the later epochs (t_2), respectively, and Δt represents the time interval between them.

Uncertainty is incorporated in these indicators by transforming the prediction intervals obtained from the previous step into empirical predictive distributions through linear interpolation, with quantile points treated as anchors of the cumulative distribution function (CDF). These predictive distributions then serve as inputs to a Monte Carlo propagation procedure (Santillan et al., 2025). For each spatial unit and forecast period, repeated random samples of BU_{t_1}, BU_{t_2}, Pop_{t_1}, and Pop_{t_2} are drawn from the empirical distributions. LCR, PGR, and LCRPGR are computed

for each sample using Equations (1)–(3). The resulting simulated distributions of LCRPGR are then summarized into median values and 95% prediction intervals, yielding uncertainty-aware estimates of LUE across time and space.

4. Case Example: Application to Global Urban Centres

To demonstrate its applicability, the framework was applied to global urban centres using the GHS-UCDB 2025 edition (European Commission et al., 2024). A probabilistic DL model was trained on historical data (1975–2020) to forecast built-up area and population for 2025 and 2030, which were then combined to produce LUE projections aligned with SDG 11.3.1 target horizons.

4.1 Data Description

The GHS-UCDB 2025 contains polygons for 11,422 urban centres worldwide (European Commission et al., 2024), defined under the Degree of Urbanisation (DEGURBA) methodology as areas with ≥50,000 persons and ≥1,500 inhabitants/km² of built-up land (European Commission & Statistical Office of the European Union, 2021). Each centre includes attributes such as country, SDG region, World Bank income classification, and multitemporal BU and Pop (1975–2020, 5-year steps), derived from GHS-BUILT-S and GHS-POP at 100 m resolution (European Commission, 2023). These inputs are reliable, with built-up errors averaging ~6% and population accuracy >80% (Pesaresi et al., 2024).

4.2 Data Preparation

This study used the spreadsheet version of the GHS-UCDB 2025, downloaded from the European Commission - Joint Research Centre Data Catalogue portal (https://data.jrc.ec.europa.eu/dataset/1a338be6-7eaf-480c-9664-3a8ade88cbcd, accessed 30 December 2024). A preliminary quality check involved calculating five-year interval annual change rates in built-up area and population from 1975 to 2020. The frequency distributions showed strong skewness and extreme values, which could bias model training and cause overfitting. To address this issue, mild to extreme outliers were removed using a modified Interquartile Range (IQR) method (Barbato et al., 2011). Urban centres with at least one built-up or population change rate outside the acceptance range were excluded, reducing the dataset from 11,422 to 8,478 centres.

After outlier removal, the dataset was split into training, validation, calibration, and test sets. The training and validation sets were used for model fitting and hyperparameter tuning. The calibration set was reserved for raw PICP evaluation and, if required, post hoc EnCQR adjustment, while the test set was used for independent evaluation. Because the dataset contained only 10 time steps, a temporal split was not feasible. Instead, we applied a geographical split based on stratified random sampling by paired UN SDG region and World Bank income group. Strata with fewer than four centres contributed at least one to the training set to ensure representation. About 70% of centres were assigned to training (6,016), and around 10% each to validation (832), calibration (849), and testing (781) (Figure 1).

BU and Pop values in each set were first log-transformed and then normalized separately using z-scores, with the mean and standard deviation for each variable computed from the training set to prevent data leakage. Log transformation was applied because initial trials showed improved training stability and accelerated convergence.

4.3 Model Fitting and Hyperparameter Tuning

The Conv1D+LSTM model (Section 3.2) was implemented in Python (3.10.11) with TensorFlow/Keras and trained using the Adam optimizer (Kingma and Ba, 2015). Hyperparameters were tuned with Optuna (Akiba et al., 2019) using the Hyperband pruning algorithm (Li et al., 2018). A two-stage process was adopted: Stage A explored the full search space (filters, kernel sizes, dilations, LSTM units, dropout, learning rate) across 120 trials (≤200 epochs, early stopping after 10), and Stage B retrained the top 10 configurations more extensively (≤400 epochs, early stopping after 20). Gradient clipping (Pascanu et al., 2013) and an adaptive learning rate scheduler (factor 0.5, minimum 1e-6) were used to stabilize training (Goodfellow et al., 2016). The input–output sequence length was fixed at 8–2, corresponding to 1975–2010 inputs (5-year intervals) and forecasts for 2015 and 2020, yielding four targets ($BU_{2015}, BU_{2020}, Pop_{2015}, Pop_{2020}$). For each target, three quantiles (2.5th, 50th, and 97.5th) were estimated (corresponding to the median and the 95% prediction intervals), giving twelve outputs in total. Model fitting and hyperparameter tuning were conducted using the training set. The validation set was used for monitoring generalization, guiding early stopping, learning-rate scheduling, and selecting the best configuration based on the lowest averaged pinball loss. Among the retrained Stage B candidates, the best-performing model was retained as the final model for subsequent analyses. The search space and final best values are reported in Table 1.

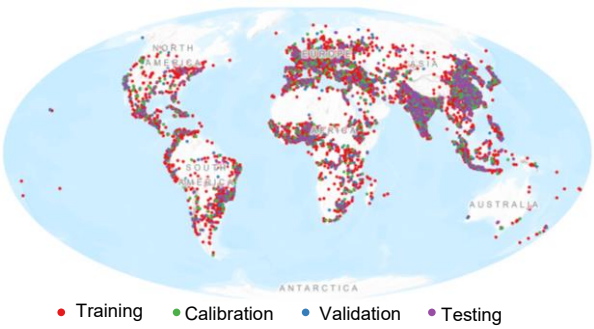

Figure 1. Distribution of urban centres across training, validation, calibration, and test sets, based on stratified splitting by SDG region and World Bank income classification.

Hyperparameter	Search space	Best value
Conv1D filters	{32, 64, 128}	128
Kernel size	{3, 5}	3
Conv1D layers	{1, 2}	2
Dilation (layer 1 / layer 2)	{1} / {1, 2}	1 / 1
LSTM units	{32, 64, 128}	64
Bidirectional	{False} (causal enforced)	False
LSTM dropout	{0.0, 0.1, 0.2}	0
Recurrent dropout	{0.0, 0.1}	0
Post-LSTM dropout	{0.0, 0.1, 0.2}	0
Dense units	{32, 64, 128}	32
L_2 regularization	{1e-6, 1e-5, 1e-4}	1e-6
Learning rate	{1e-3, 5e-4}	1e-3
Batch size	{32, 64}	32
Gradient clip norm	{1.0}	1.0

Table 1. Hyperparameter search space and selected best configuration for the Conv1D+LSTM model.

4.4 Ensembling, Calibration, and Model Performance Evaluation

An ensemble was formed from the best model identified in the previous step, retrained on the combined training and validation sets 30 times with different random initializations. Forecasts from the ensemble were aggregated, with PICP checked on the calibration set; prediction intervals were either retained with PICP close to the 95% target, or adjusted post hoc with EnCQR (Section 3.3).

Model performance was evaluated on the independent test set using whichever intervals were carried forward (raw or calibrated). Based on observed 2015 and 2020 data, two aspects were evaluated: (i) point forecast accuracy, via Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) (Hyndman and Koehler, 2006); and (ii) uncertainty quality, via empirical prediction interval coverage (PICP) and average interval width (PIAW) (Jensen et al., 2024). Each ensemble member produced BU and Pop medians and intervals for all test urban centres, yielding 30 values per year–variable combination. The median and median absolute deviation (MAD) were reported for MAE and MAPE to summarize performance and assess consistency. These statistics were chosen for their robustness to skewness, outliers, and alignment with the quantile-based forecasts (median and 95% prediction intervals). PICP and PIAW were reported for the ensemble intervals both before and, where relevant, after calibration to demonstrate the impact of the adjustment.

4.5 Forecasting Built-up Area and Population for LUE Assessment: 2025-2030

The tuned ensemble model was applied to historical sequences from 1985 to 2020 to generate forecasts of BU and Pop for 2025 and 2030 for the 8,478 urban centres included in the analysis. These forecasts and observed 2020 values were used to estimate LCR, PGR, and LCRPGR for the periods 2020–2025 and 2025–2030. Observations for 2020 were treated as uncertainty-free, while forecast uncertainty for 2025 and 2030 was addressed through the conditional ensembling and post hoc calibration strategy outlined in Section 3.3. If calibration was necessary, the combined 2015 and 2020 calibration data conformity scores were applied to adjust the raw ensemble intervals.

Uncertainty propagation into the LCR, PGR, and LCRPGR metrics followed the Monte Carlo procedure described in Section 3.4. For each urban centre and forecast period, 50,000 random draws of BU_{t_1}, BU_{t_2}, Pop_{t_1}, and Pop_{t_2} were generated from the empirical predictive distributions, a sample size chosen to ensure stable estimates of distribution tails while remaining computationally feasible. The three indicators were recomputed for each draw, and the resulting simulated distributions were summarized into medians and 95% prediction intervals, providing uncertainty-aware estimates of LUE for 2020–2025 and 2025–2030

5. Results and Discussion

5.1 Ensemble Point Forecasting Accuracy

Table 2 and Figure 2 summarize the test evaluation results of the Conv1D+LSTM ensemble for point forecasting of BU and Pop. The ensemble achieved high accuracy for BU in 2015, with a median MAE of 0.10 km² and a MAPE of 0.73%. By 2020, the accuracy declined, where the MAE had nearly tripled to 0.28 km² and the MAPE more than tripled to 2.43%. When both years were combined, intermediate values were obtained, as expected from the lower errors in 2015 and higher errors in 2020. The boxplots (Figure 2) confirm these patterns, showing compact error distributions but a clear upward shift from 2015 to 2020. Population forecasts followed a similar trend. In 2015, the MAE was 6,582 persons and MAPE 1.90%. By 2020, the MAE had almost tripled to 17,735 persons, while the MAPE more than doubled to 4.88%. The combined case yielded intermediate results (MAE = 12,198; MAPE = 3.39%). The distributions remain narrow across ensemble members, but the medians clearly show the decline in accuracy with longer forecast horizons.

Across both variables, forecast accuracy is stronger in the near term, with MAPE values remaining below 5% even at the 2020 horizon. This indicates that the ensemble model produces practically reliable point forecasts despite the increased difficulty of longer-range prediction. The very small deviations across ensemble members (Table 2) and the narrow notches in the boxplots (Figure 2) demonstrate high consistency of the Conv1D+LSTM ensemble, with most variability attributable to forecast horizon rather than model instability.

Forecast Year	BU		Pop	
	MAE (km²)	MAPE (%)	MAE (no. of persons)	MAPE (%)
2015	0.10 ± 0.01	0.73 ± 0.03	6582 ± 251	1.90 ± 0.01
2020	0.28 ± 0.01	2.43 ± 0.02	17735 ± 340	4.88 ± 0.02
2015 & 2020	0.19 ± 0.01	1.58 ± 0.03	12198 ± 276	3.39 ± 0.01

Table 2. Test evaluation results of the Conv1D+LSTM ensemble for point forecasts of built-up area (BU) and population (Pop) in 2015, 2020, and combined 2015–2020. Results are reported as median ± median absolute deviation (MAD).

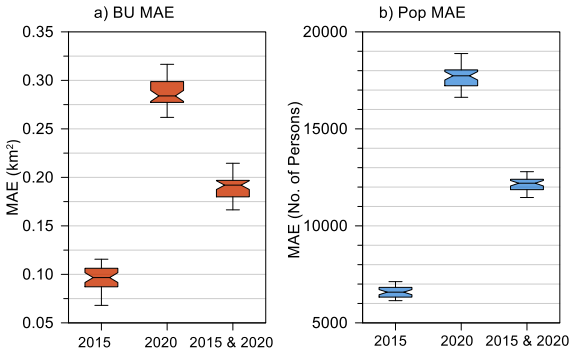

Figure 2. Notched box plots of Mean Absolute Error (MAE) of point forecasts from the 30 Conv1D+LSTM ensemble members on the test set. The middle horizontal line marks the median, notches indicate its 95% confidence interval, and whiskers show the minimum and maximum values.

5.2 Ensemble Model Uncertainty Quality

The Conv1D+LSTM ensemble produced prediction intervals that closely aligned with the nominal 95% coverage level across both calibration and test sets (Tables 3 and 4). For BU, PICP values ranged from 94% to 97%, while for Pop they ranged from 94% to 95%. In other words, more than 94% of the observed BU and Pop values fell within their respective 95% prediction intervals. This demonstrates that the ensemble intervals were well-calibrated, providing reliable uncertainty estimates without systematic under- or over-coverage, and thus did not require post hoc adjustment.

The widths of the prediction intervals were generally stable

across datasets and forecast years. For BU, average widths remained around 0.5–1.5 km², which is narrow relative to the typical size of urban areas and thus provides decision-useful bounds. For Pop, however, average widths were substantially larger, ranging from 39,000 to over 107,000 persons. While these intervals still provided valid coverage, their magnitude is non-negligible, especially for medium-sized urban centres where such differences can represent a sizeable share of the total population. This result suggests that although the uncertainty estimates are reliable, the model is less precise for population forecasts than for built-up area. These patterns are illustrated in Figure 3, which shows example forecasts for Greater Metro Manila, Philippines.

Forecast Year	PICP (%)		PIAW (km²)	
	Cal. Set	Test Set	Cal. Set	Test Set
2015	96	97	0.63	0.55
2020	94	96	1.53	1.50
2015 & 2020	95	96	1.08	1.02

Table 3. Conv1D+LSTM's built-up area (BU) raw prediction interval coverage probability (PICP) and average width (PIAW).

Forecast Year	PICP (%)		PIAW (no. of persons)	
	Cal. Set	Test Set	Cal. Set	Test Set
2015	94	95	39,202	45,399
2020	94	95	95,206	107,925
2015 & 2020	94	95	67,204	76,662

Table 4. Conv1D+LSTM's population (Pop) raw prediction interval coverage probability (PICP) and average width (PIAW).

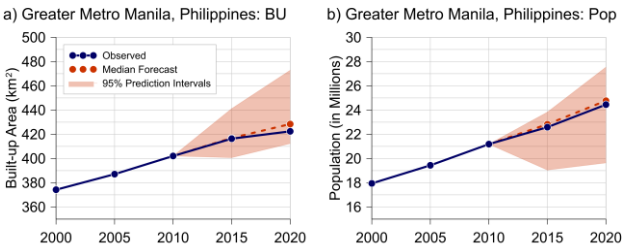

Figure 3. Observed and forecasted built-up area (BU) and population (Pop) for 2015 and 2020, with 95% prediction intervals, shown for Greater Metro Manila, Philippines.

5.3 Forecasts for 2025 and 2030

Figure 4 shows the forecasted totals of BU and Pop across 8,478 urban centres for 2025 and 2030, using 2020 as the baseline. By 2025, BU is projected to reach 113,849 km² (95% prediction interval: 113,584–114,105 km²), representing an increase of about 4.6% from its 2020 value of 108,808 km². By 2030, BU is expected to rise further to 118,123 km² (95% prediction interval: 117,765–118,482 km²), a 3.7% increase relative to 2025. Over the decade, this amounts to a net expansion of about 8.6%, indicating a sustained but moderate pace of urban land growth beyond the 2020 baseline.

The population is likewise projected to increase steadily. From 2.875 billion in 2020, it is forecasted to reach 3.011 billion in 2025 (+4.7%; 95% prediction interval: 2.993–3.027 billion) and then 3.159 billion in 2030 (+4.9% from 2025; 95% prediction interval: 3.130–3.188 billion). This represents a total growth of about 9.9% over the decade.

These forecasts align with broader global projections of continued BU and Pop growth. Li et al. (2019) estimate urban land expansion of 40–67% between 2013 and 2050 (≈1.1–1.8% annually), while Gao and O'Neill (2020) project an 80–490% increase by 2100 (≈0.8–4.9% annually). For our 8,478 urban centres, BU is forecast to grow by 0.86% annually from 2020 to 2030, consistent with the lower range of these estimates. Population is projected to rise 9.9% over the same decade, close to the UN's 8.6% projection (UN Population Division, 2025).

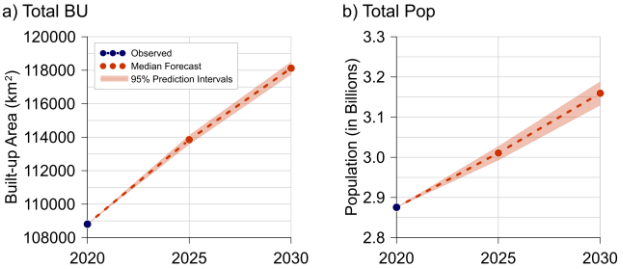

Figure 4. Forecasts of total built-up area (BU) and total population (Pop) in 2025 and 2030, aggregated across the 8,478 urban centres analyzed in the study.

5.4 LUE Projections

Using observed BU and Pop totals for 2015 and 2020 together with the aggregated forecasts, we computed LCR, PGR, and LCRPGR for 2015–2020, 2020–2025, and 2025–2030 (Table 5). While computed per urban center, only the aggregated results are shown due to space constraints. Using 2015–2020 as the baseline, the forecasts provide insights into how LUE may evolve toward 2030. For 2020–2025, the median LCRPGR is 1.01, almost exactly at the efficiency threshold, with the 95% prediction interval (0.89–1.16) spanning both efficient and slightly inefficient outcomes. For 2025–2030, the median value (0.78) suggests efficiency, as population growth is projected to outpace land expansion, though the interval (0.62–1.02) again includes marginal inefficiency.

The results indicate a general trend toward efficiency, driven more by moderating land expansion relative to population growth than by rapid demographic gains. This projection continues the improvement seen in earlier decades, when more than half of the global functional urban areas improved their efficiency between 2000 and 2015 compared to 1990–2000 (Schiavina et al., 2022b). However, because the forecast intervals cross the efficiency threshold in both periods, the possibility of inefficiency cannot be ruled out, underscoring the importance of interpreting these results cautiously and within the broader context of SDG 11.3.1 monitoring.

Period	LCR	PGR	LCRPGR
2015-2020	0.72	0.94	0.76
2020-2025	0.93 (0.88-0.97)	0.92 (0.80-1.03)	1.01 (0.89-1.16)
2025-2030	0.75 (0.67-0.83)	0.97 (0.75-1.18)	0.78 (0.62-1.02)

Table 5. Median LCR, PGR, and LCRPGR aggregated for 8,478 urban centres across three periods. Forecast-based values for 2020–2025 and 2025–2030 include ±95% prediction intervals in parentheses.

6. Conclusions and Outlook

In this study, we advance SDG 11.3.1 monitoring from retrospective analysis toward forward-looking, uncertainty-aware assessment by introducing a scalable framework that jointly forecasts built-up area and population and propagates those forecasts into LCR, PGR, and LCRPGR. The framework uses a lightweight Conv1D+LSTM ensemble trained with quantile loss to generate probabilistic forecasts. Applied to 8,478 urban centres from GHS-UCDB 2025, it produced projections for the period 2025-2030 with prediction intervals shown to be well-

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-5-W4-2025-385-2026 | © Author(s) 2026. CC BY 4.0 License.

calibrated on historical data. The results suggest a general trend toward efficiency, with PGR slightly outpacing LCR, while acknowledging uncertainty near the efficiency threshold. The key contributions are twofold: (i) a reproducible global pipeline tailored to short, sparse urban time series, and (ii) an uncertainty-explicit foundation for evaluating near-term LUE trajectories to support timely and risk-aware policy decisions as 2030 approaches.

The current implementation was designed for the constraints typical of SDG 11.3.1 forecasting: short time series (≤10 observations), limited covariates, and large-scale application across thousands of urban centres. The Conv1D+LSTM ensemble provided a practical balance between accuracy, scalability, and uncertainty quantification within this setting. At the same time, the framework remains flexible and can be extended in several ways. Using denser built-up area or census observations would improve temporal precision. Adding socio-economic or environmental covariates could enhance predictive accuracy. Adopting alternative multi-step architectures may strengthen long-horizon forecasts. Finally, incorporating spatial dependencies through graph neural networks or attention-based models would allow regional dynamics to be captured. With these improvements, the framework can provide even stronger and more policy-relevant support for SDG 11.3.1 monitoring.

Acknowledgements

J.R. Santillan acknowledges support from Philippines' Department of Science and Technology – Science Education Institute (DOST-SEI) through a doctoral scholarship and from Caraga State University through a fellowship.

References

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U. R., Makarenkov, V., Nahavandi, S., 2021. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76, 243–297. doi.org/10.1016/j.inffus.2021.05.008.

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631. doi.org/10.1145/3292500.3330701.

Alghanmi, N., Alotaibi, R., Alshammari, S., Mahmood, A., 2024. Population fusion transformer for subnational population forecasting. *International Journal of Computational Intelligence Systems*, 17(1), 26. doi.org/10.1007/s44196-024-00413-y.

Angelopoulos, A. N., Bates, S., 2023. Conformal prediction: A gentle introduction. *Foundations and Trends® in Machine Learning*, 16(4), 494–591. doi.org/10.1561/2200000101.

Bai, S., Kolter, J. Z., Koltun, V., 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *ArXiv Preprint*, arXiv:1803.01271.

Barbato, G., Barini, E. M., Genta, G., Levi, R., 2011. Features and performance of some outlier detection methods. *Journal of Applied Statistics*, 38(10), 2133–2149.

Benidis, K., Rangapuram, S. S., Flunkert, V., Wang, Y., Maddix, D., Turkmen, C., Gasthaus, J., Bohlke-Schneider, M., Salinas, D., Stella, L., Aubet, F.-X., Callot, L., Januschowski, T., 2023. Deep learning for time series forecasting: Tutorial and literature survey. *ACM Computing Surveys*, 55(6), 1–36.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *ArXiv Preprint*, arXiv:1406.1078.

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.

European Commission, Joint Research Centre, Melchiorri, M., Mari Rivero, I., Florio, P., Schiavina, M., Krasnodebska, K., Politis, P., Uhl, J., Pesaresi, M., Maffenini, L., Sulis, P., Crippa, M., Guizzardi, D., Pisoni, E., Belis, C., Oom, D., Branco, A., Mwaniki, D., … Dijkstra, L., 2024. *Stats in the City the GHSL Urban Centre Database 2025*. https://data.europa.eu/doi/10.2760/3046391.

European Commission, & Statistical Office of the European Union, 2021. *Applying the Degree of Urbanisation — A methodological manual to define cities, towns and rural areas for international comparisons — 2021 edition*. Publications Office of the European Union.

Fatima, S.S.W., Rahimi, A., 2024. A review of time-series forecasting algorithms for industrial manufacturing systems. *Machines*, 12(6), p.380. doi.org/10.3390/machines12060380.

Gal, Y., Ghahramani, Z., 2016. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. *Proceedings of the International Conference on Machine Learning (ICML)*, PMLR (2016), 1050-1059.

Gao, J., O'Neill, B.C., 2020. Mapping global urban land for the 21st century with data-driven simulations and shared socio-economic pathways. *Nature Communications*, 11, 2302.

Geiß, C., Maier, J., So, E., Zhu, Y., 2023. LSTM models for spatiotemporal extrapolation of population data. *2023 Joint Urban Remote Sensing Event (JURSE)*, 1–4. doi.org/10.1109/JURSE57346.2023.10144145.

Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep learning*. MIT Press.

Grossman, I., Wilson, T., Temple, J., 2023. Forecasting small area populations with long short-term memory networks. *Socio-Economic Planning Sciences*, 88, 101658. doi.org/10.1016/j.seps.2023.101658.

Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780.

Hyndman, R. J., Koehler, A. B., 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688.

Jensen, V., Bianchi, F. M., Anfinsen, S. N., 2024. Ensemble conformalized quantile regression for probabilistic time series forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, 35(7), 9014–9025.

Jin, M., Koh, H.Y., Wen, Q., Zambon, D., Alippi, C., Webb, G.I., King, I., Pan, S., 2024. A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12), 10466 – 10485.

Kingma, D. P., Ba, J., 2015. Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations (ICLR)*.

Lakshminarayanan, B., Pritzel, A., Blundell, C., 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in Neural Information Processing Systems* 30 (NIPS 2017).

Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., Talwalkar, A., 2018. Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research*, 18(185), 1-52.

Li, X., Zhou, Y., Eom, J., Yu, S., Asrar, G.R., 2019. Projecting global urban area growth through 2100 based on historical time series data and future shared socio-economic pathways. *Earth's Future*, 7(4), 351-362.

Li, W., Law, K. L. E., 2024. Deep learning models for time series forecasting: A review. *IEEE Access*, 12, 92306–92327. doi.org/10.1109/ACCESS.2024.3422528.

Lim, B., Arık, S. Ö., Loeff, N., Pfister, T., 2021. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4), 1748–1764. doi.org/10.1016/j.ijforecast.2021.03.012.

Masini, R.P., Medeiros, M.C., Mendes, E.F., 2023. Machine learning advances for time series forecasting. *Journal of Economic Surveys*, 37(1), 76-111.

Lu, L., Qureshi, S., Li, Q., Chen, F., Shu, L., 2022. Monitoring and projecting sustainable transitions in urban land use using remote sensing and scenario-based modelling in a coastal megacity. *Ocean & Coastal Management*, 224, 106201.

Pascanu, R., Mikolov, T., Bengio, Y., 2013. On the difficulty of training recurrent neural networks. *Proceedings of the 30th International Conference on Machine Learning*, PMLR 28(3), 1310-1318.

Pesaresi, M., Schiavina, M., Politis, P., Freire, S., Krasnodębska, K., Uhl, J. H., Carioli, A., Corbane, C., Dijkstra, L., Florio, P., Van Den Hoek, J., Kemper, T., 2024. Advances on the global human settlement layer by joint assessment of earth observation and population survey data. *International Journal of Digital Earth*, 17(1).

Romano, Y., Patterson, E., Candes, E., 2019. Conformalized quantile regression. *Advances in Neural Information Processing Systems 30 (NeurIPS 2019)*.

Salinas, D., Flunkert, V., Gasthaus, J., Januschowski, T., 2020. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3), 1181–1191. doi.org/10.1016/j.ijforecast.2019.07.001.

Santillan, J. R., Dorozynski, M., Heipke, C., 2025. Uncertainty quantification and Monte Carlo simulations to enhance SDG 11.3.1 monitoring. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 753–762. doi.org/10.5194/isprs-annals-X-G-2025-753-2025.

Schiavina, M., Melchiorri, M., Corbane, C., Freire, S., Batista e Silva, F., 2022a. Built-up areas are expanding faster than population growth: regional patterns and trajectories in Europe. *Journal of Land Use Science*, 17(1), 591–608.

Schiavina, M., Melchiorri, M., Freire, S., Florio, P., Ehrlich, D., Tommasi, P., Pesaresi, M., Kemper, T., 2022b. Land use efficiency of functional urban areas: Global pattern and evolution of development trajectories. *Habitat International*, 123, 102543.

Sen, R., Yu, H.-F., Dhillon, I., 2019. Think globally, act locally: A deep neural network approach to high-dimensional time series forecasting. *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*.

Sluijterman, L., Cator, E., Heskes, T., 2024. Optimal training of mean variance estimation neural networks. *Neurocomputing*, 597, 127929. doi.org/10.1016/j.neucom.2024.127929.

UN Population Division, 2025. UN Population Division Data Portal. https://population.un.org/dataportal/home (2 June 2025).

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 (2 June 2025).

UN-Habitat, 2018. Metadata on SDGs Indicator 11.3.1 Indicator category: Tier II. https://unhabitat.org/sites/default/files/2020/07/metadata_on_sdg_indicator_11.3.1.pdf (2 June 2025).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. *Advances in Neural Information Processing Systems*, 30 (NIPS 2017), 30.

Wang, C., Liu, S., Feng, T., Liu, Y., Liu, Y., Hu, Z., Yang, L., Niu, Q., Mao, X., 2024. Is the assessment approach of Sustainable Development Goal 11.3.1 justified? Evidence from the drivers of future urban land use efficiency. *Journal of Cleaner Production*, 444, 141147.

Wang, Y., Gan, D., Sun, M., Zhang, N., Lu, Z., Kang, C., 2019. Probabilistic individual load forecasting using pinball loss guided LSTM. *Applied Energy*, 235, 10–20.

Wang, Y., Sun, B., Zhang, T., 2022. Do polycentric urban regions promote functional spillovers and economic performance? Evidence from China. *Regional Studies*, 56(1), 63–74.

Xu, C., Xie, Y., 2023. Conformal prediction for time series. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(10), 11575–11587.