



DGK Ausschuss Geodäsie (DGK)
der Bayerischen Akademie der Wissenschaften

Reihe C

Dissertationen

Heft Nr. 978

Jojene R. Santillan

**Probabilistic and Multidimensional Assessment
of Urban Land-Use Efficiency from Earth Observation Data**

München 2026

Verlag der Bayerischen Akademie der Wissenschaften, München

ISSN 0065-5325

ISBN 978 3 7696 5390 8



Probabilistic and Multidimensional Assessment
of Urban Land-Use Efficiency from Earth Observation Data

Von der Fakultät für Bauingenieurwesen und Geodäsie
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des akademischen Grades
Doktor-Ingenieur (Dr.-Ing.)
genehmigte Dissertation

von

Jojene R. Santillan, M. Sc.

München 2026

Verlag der Bayerischen Akademie der Wissenschaften, München

Adresse des Ausschusses Geodäsie (DGK)
der Bayerischen Akademie der Wissenschaften:



Ausschuss Geodäsie (DGK) der Bayerischen Akademie der Wissenschaften

Alfons-Goppel-Straße 11 • D – 80 539 München
Telefon +49 – 89 – 23 031 1113 • Telefax +49 – 89 – 23 031 - 1283 / - 1100
e-mail post@dgk.badw.de • <http://www.dgk.badw.de>

Prüfungskommission:

Vorsitzender: Prof. Dr.-Ing. habil. Jürgen Müller
Referent: Prof. Dr.-Ing. habil. Christian Heipke
Korreferenten: Prof. i.R. Dr.-Ing. Winrich Voß
Prof. Dr. Hannes Taubenböck

Tag der mündlichen Prüfung: 17.06.2026

:

Diese Dissertation ist auf dem Server des Ausschusses Geodäsie (DGK)
der Bayerischen Akademie der Wissenschaften, München unter <http://dgk.badw.de/>
sowie unter Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik
der Leibniz Universität Hannover (ISSN 0174-1454), Nr. 421,
<<https://repo.uni-hannover.de/handle/123456789/21863>>, Hannover 2026, elektronisch publiziert

© 2026 Ausschuss Geodäsie (DGK) der Bayerischen Akademie der Wissenschaften, München

Alle Rechte vorbehalten. Ohne Genehmigung der Herausgeber ist es auch nicht gestattet,
die Veröffentlichung oder Teile daraus zu vervielfältigen.

Abstract

Urban land-use efficiency (LUE) is central to sustainable development, as urban areas must accommodate population growth while minimizing unnecessary land consumption and environmental consequences. Sustainable Development Goal (SDG) Indicator 11.3.1 was introduced to monitor LUE through the land consumption rate (LCR), the population growth rate (PGR), and their ratio (LCRPGR). However, current implementations remain largely deterministic, retrospective, and two-dimensional. The input variables used for the calculation—built-up area (BUA) and population (Pop)—are typically treated as exact values, thus neglecting measurement uncertainties. At the same time, the indicator itself remains focused on horizontal urban expansion and does not capture the vertical dimension of urban growth. Consequently, prevailing applications of SDG Indicator 11.3.1 remain insufficient for a reliable, policy-relevant assessment of urban LUE.

This dissertation addresses these limitations by developing and evaluating uncertainty-aware, forecast-enabled, and multidimensional approaches to LUE assessment based on SDG Indicator 11.3.1. First, a deterministic sensitivity analysis investigates how alternative datasets for BUA and Pop under retrospective conditions affect LCR, PGR, LCRPGR, and the resulting LUE interpretations. Second, a general uncertainty-propagation framework is developed in which uncertain BUA and Pop inputs are probabilistically represented and propagated through the SDG Indicator 11.3.1 formulae using Monte Carlo simulations. Subsequently, a probabilistic LUE classification is used to express class membership as probabilities. Third, a global deep learning workflow is designed to jointly forecast BUA and Pop for urban centers (UCs) using probabilistic sequence-to-sequence models, deep ensembles, and dependency-aware uncertainty propagation for a forward-looking LUE assessment. Finally, a trajectory framework is introduced that integrates BUA and built-up volume (BUV) to improve the interpretability and robustness of LUE assessment under SDG Indicator 11.3.1, capturing both horizontal and vertical dimensions of urban growth.

Applied to Philippine local government units (LGUs), the sensitivity analysis demonstrated that SDG Indicator 11.3.1 is materially affected by routine choices of BUA and Pop datasets, with BUA representation emerging as the stronger source of deterministic variability. The uncertainty propagation analysis, demonstrated through both a hypothetical example and a real-world case study of Baguio City, Philippines, further showed that explicitly representing BUA and Pop as uncertain inputs can substantially alter the estimated values and interpretation of the indicator. Probabilistic LUE classification improved interpretation by identifying both the dominant class and the degree of support for alternative classes. Across global UCs, the forecasting analysis demonstrated that an uncertainty-aware deep learning workflow can support forward-looking assessment of SDG Indicator 11.3.1 by jointly predicting BUA and Pop while quantifying confidence in projected LUE classes across forecast horizons. The BUA–BUV trajectory framework further demonstrated that integrating horizontal and vertical dimensions of urban growth improves the interpretability of LUE

assessment by revealing structural growth patterns that are not captured by conventional indicator values alone. In particular, it showed that favorable LCRPGR outcomes do not necessarily imply compact or vertically oriented urban development.

Overall, the dissertation demonstrates that a more reliable and policy-relevant assessment of urban LUE requires SDG Indicator 11.3.1 to be calculated and interpreted with explicit consideration of input-data sensitivity and uncertainty, projected change, and the multidimensional nature of urban growth.

Keywords: Urban land-use efficiency, SDG Indicator 11.3.1, uncertainty propagation, probabilistic forecasting, BUA–BUV trajectory framework.

Kurzfassung

Die städtische Flächennutzungseffizienz (urban land-use efficiency, LUE) ist von zentraler Bedeutung für eine nachhaltige Entwicklung, da städtische Gebiete dem Bevölkerungswachstum gerecht werden müssen, ohne dabei unnötigen Flächenverbrauch zu verursachen und damit verbundene Umweltfolgen zu verstärken. Der Indikator 11.3.1 der Ziele für nachhaltige Entwicklung (Sustainable Development Goals, SDGs) wurde eingeführt, um LUE anhand der Flächenverbrauchsrate (land consumption rate, LCR), der Bevölkerungswachstumsrate (population growth rate, PGR) und ihres Verhältnisses (LCRPGR) zu überwachen. Gegenwärtige Realisierungen dieses Indikators sind jedoch weitgehend deterministisch, retrospektiv und zweidimensional. Die für die Berechnung verwendeten Eingangsgrößen bebaute Fläche (built-up area, BUA) und Bevölkerung (Pop) werden in der Regel als exakte Größen behandelt, wodurch Messunsicherheiten unberücksichtigt bleiben. Zugleich konzentriert sich der Indikator auf horizontale städtische Expansion und erfasst nicht die vertikale Dimension städtischen Wachstums. Folglich sind die derzeitigen Anwendungen des SDG-Indikators 11.3.1 für eine verlässliche und politikrelevante Bewertung der städtischen LUE unzureichend.

Die vorliegende Dissertation adressiert diese Einschränkungen, indem sie unsicherheitsbewusste, prognosefähige und multidimensionale Ansätze zur Bewertung der LUE auf Grundlage des SDG-Indikators 11.3.1 entwickelt und evaluiert. Zunächst untersucht eine deterministische Sensitivitätsanalyse, wie sich alternative Datensätze zu BUA und Pop unter retrospektiven Bedingungen auf LCR, PGR, LCRPGR und die daraus resultierenden LUE-Interpretationen auswirken. Darüber hinaus wird ein allgemeiner methodischer Ansatz zur Unsicherheitsfortpflanzung entwickelt, in dem unsichere BUA- und Pop-Eingangsgrößen probabilistisch dargestellt und mithilfe von Monte-Carlo-Simulationen durch die Formeln des SDG-Indikators 11.3.1 fortgepflanzt werden. Anschließend wird eine probabilistische LUE-Klassifizierung angewendet, um die Klassenzugehörigkeit in Form von Wahrscheinlichkeiten auszudrücken. Drittens wird ein globales Deep-Learning-Verfahren für eine zukunftsgerichtete LUE-Bewertung konzipiert: Dabei werden für urbane Zentren (urban centers, UCs) BUA und Pop gemeinsam mithilfe probabilistischer Sequence-to-Sequence-Modelle, Deep Ensembles und einer Unsicherheitsfortpflanzung unter Berücksichtigung von Abhängigkeiten prognostiziert. Schließlich wird ein Trajektorienansatz eingeführt, der BUA und bebautes Volumen (built-up volume, BUV) integriert, um die Interpretierbarkeit und Robustheit einer auf dem SDG-Indikator 11.3.1 basierenden LUE-Bewertung zu verbessern, indem sowohl die horizontalen als auch die vertikale Dimensionen städtischen Wachstums erfasst werden.

Angewandt auf lokale philippinische Verwaltungseinheiten (local government units, LGUs), zeigte die Sensitivitätsanalyse, dass der SDG-Indikator 11.3.1 wesentlich durch routinemäßige Entscheidungen hinsichtlich der verwendeten BUA- und Pop-Datensätze beeinflusst wird, wobei sich die Darstellung der BUA als stärkere Quelle deterministischer Variabilität erweist. Die Analyse der

Unsicherheitsfortpflanzung demonstriert, sowohl anhand eines hypothetischen Beispiels als auch anhand einer realen Fallstudie für Baguio City auf den Philippinen, dass die explizite Modellierung von BUA und Pop als unsichere Eingangsgrößen die geschätzten Werte und die Interpretation des Indikators erheblich verändern kann. Die probabilistische LUE-Klassifikation verbessert die Interpretation, indem sie sowohl die dominante Klasse als auch den Grad der Unterstützung für alternative Klassen identifizierte. Über globale UCs hinweg, zeigt die Prognoseanalyse, dass ein unsicherheitsbewusstes Deep-Learning-Verfahren eine zukunftsgerichtete Bewertung des SDG-Indikators 11.3.1 unterstützen kann, indem BUA und Pop gemeinsam vorhergesagt und zugleich die Konfidenz der prädizierten LUE-Klassen über verschiedene Prognosehorizonte hinweg quantifiziert werden. Der BUA-BUV-Trajektorienansatz zeigt darüber hinaus, dass die Integration horizontaler und vertikaler Dimensionen städtischen Wachstums die Interpretierbarkeit der LUE-Bewertung verbessert, indem strukturelle Wachstumsmuster sichtbar gemacht werden, die durch konventionelle Indikatorwerte allein nicht erfasst werden. Insbesondere wird deutlich, dass positive LCRPGR-Ergebnisse nicht zwangsläufig auf eine kompakte oder vertikal orientierte Stadtentwicklung hinweisen.

Insgesamt zeigt die Dissertation, dass eine verlässlichere und politikrelevantere Bewertung der städtischen LUE es erfordert, den SDG-Indikator 11.3.1 unter expliziter Berücksichtigung von Sensitivität und Unsicherheit der Eingangsdaten, projizierter Veränderungen sowie der multidimensionalen Natur städtischen Wachstums zu berechnen und zu interpretieren.

Schlagwörter: Effizienz der städtischen Flächennutzung, SDG-Indikator 11.3.1, Unsicherheitsfortpflanzung, probabilistische Prognose, BUA–BUV-Trajektorienansatz

Acknowledgements

This work was made possible through the financial and institutional support of the Department of Science and Technology–Science Education Institute (DOST–SEI) of the Philippines, under the Foreign Graduate Scholarship Program in Specialized Priority Fields in Science and Technology, and Caraga State University (CSU), through its doctoral fellowship program. I am sincerely grateful for this support, which enabled me to pursue my doctoral studies at Leibniz University Hannover (LUH).

I would like to express my sincere gratitude to my supervisor, Prof. Dr.-Ing. habil. Christian Heipke, for giving me the opportunity to pursue my doctoral studies at the Institute of Photogrammetry and GeoInformation (IPI), LUH. I am thankful for his trust, guidance, constructive questions, and academic advice throughout this work. Although the topic of this dissertation was not fully within the central scope of his main research areas, he gave me the freedom to develop my ideas, encouraged me to pursue my own research direction, and accompanied the work with patience and openness. For this academic generosity, I am deeply grateful.

I would also like to thank the other members of my doctoral committee: Prof. i.R. Dr.-Ing. Winrich Voß and Prof. Dr. Hannes Taubenböck as co-examiners, and Prof. Dr.-Ing. habil. Jürgen Müller as chair. I appreciate their time, thoughtful questions, and constructive comments, which helped strengthen this dissertation and made the final stage of the doctoral process a meaningful academic exchange. I sincerely value their role in this important milestone.

I am also deeply grateful to Dr. Enrico C. Paringit, my former mentor from the University of the Philippines Diliman (UPD), for his important contribution to my development as a researcher. His mentorship, guidance, and example helped build my research capabilities and shaped the way I approach academic work. I am especially thankful for his endorsement and support for my scholarship application, which played an important role in opening the path toward this doctoral journey. I also sincerely thank Dr. Ariel C. Blanco of UPD for his support and endorsement. My sincere gratitude also goes to Dr. Rowena P. Varela of CSU for her support and endorsement of my scholarship application, which I truly appreciate. I would also like to thank the current CSU President, Dr. Rolyn C. Daguil, and former CSU President, Dr. Anthony M. Penaso, for their support, endorsement, and trust in my applications.

My sincere thanks also go to my colleagues at the IPI. In many different ways, they contributed to this journey through discussions, advice, technical and administrative support, shared experiences, and the collegial environment that made IPI a good place to work. I am grateful for the support and kindness I received during my stay in Hannover. I would like to give special mention to Dr.-Ing. Mareike Dorozynski for her valuable contributions and assistance in co-writing the research papers that formed part of, and contributed to, the contents of this dissertation; to Dr.-Ing. Max

Mehltretter for his assistance with the German translation of the abstract and for facilitating the defense rehearsal; and to Dr.-Ing. Andreas Piter for his assistance with the doctoral examination application process.

I would also like to thank my family, friends, and colleagues in the Philippines for their support, encouragement, and prayers throughout this doctoral work. I am especially grateful to my colleagues and students at CSU, whose encouragement and continued interest in my work helped me remain connected to the academic community and professional responsibilities to which I will return.

Most importantly, I would like to thank my wife, Meriam, for her patience, understanding, and support throughout this doctorate. She shared many of the difficulties that came with this journey, including periods of stress, uncertainty, and time apart. Her encouragement and steady support were invaluable in helping me complete this work.

To all of you, thank you very much.

Contents

Abstract	i
Kurzfassung	iii
Acknowledgements	v
Abbreviations	xiii
1 Introduction	1
1.1 Background and Motivations	1
1.1.1 The Uncertainty of Indicator-Based Assessment	1
1.1.2 The Need for Forward-Looking Assessment	3
1.1.3 The Dimensional Constraints of Planar Metrics	3
1.2 Central Thesis Statement	5
1.3 Objective and Research Questions	6
1.4 Contributions	6
2 Conceptual Foundations and Related Work	8
2.1 Conceptual Foundations	8
2.1.1 LUE and SDG Indicator 11.3.1	8
2.1.2 LUE Interpretation in SDG Indicator 11.3.1 Context	9
2.2 State of the Art in SDG Indicator 11.3.1 Monitoring	14
2.2.1 Earth Observation (EO)-based Data Sources and Implementation Practices	14
2.2.2 Result Interpretation and Reporting Practices	16
2.2.3 Methodological Characteristics and Known Challenges	17
2.2.3.1 Mathematical Behavior and Interpretation of LCRPGR under Pop- ulation Change	17
2.2.3.2 Operational Definitions of Built-up Area in SDG Indicator 11.3.1 Monitoring	19
2.2.3.3 Treatment of Accuracy and Uncertainty in Input Data	20
2.2.3.4 Emerging Sensitivity and Robustness Analyses	21
2.2.3.5 Dimensional and Spatial Scope of Land Consumption Representation	22
2.2.4 Conceptual and Methodological Extensions of LCRPGR	23
2.2.5 Extensions of SDG Indicator 11.3.1 to LUE Forecasting	24
2.3 Synthesis and Summary of Research Gaps	26

3	Sensitivity of SDG Indicator 11.3.1 to Alternative Representations of Built-up Area and Population	28
3.1	Chapter Overview	28
3.2	Sensitivity Analysis Methodology	29
3.3	Experimental Setup and Implementation Details	29
3.3.1	Study Area	29
3.3.2	Spatial Units of Analysis, Temporal Scope, and Analytical Interval	31
3.3.3	Datasets Used	31
3.3.4	Data Preparation	36
3.3.5	Indicator Calculation and Sensitivity Analysis	36
3.3.6	Data Completeness and Indicator Calculability Across Configurations	37
3.4	Results	38
3.4.1	Dataset Differences in BUA and Pop	38
3.4.2	Baseline SDG Indicator 11.3.1 Metrics (GHS-BUILT-S + Census)	40
3.4.3	Sensitivity to BUA Representation	40
3.4.4	Sensitivity to Pop Representation	43
3.4.5	City Examples for Interpreting Deterministic Sensitivity	48
3.5	Discussion	49
3.5.1	Directional Effects of BUA and Pop Substitution	49
3.5.2	Why BUA Substitutions Exert the Stronger Effect	49
3.5.3	Relation to Previous Studies	51
3.5.4	Implications for SDG Indicator 11.3.1 Monitoring and Policy Use	51
3.6	Limitations and Further Work	52
3.7	Chapter Conclusion	53
4	Uncertainty Propagation and Probabilistic LUE Classification for SDG Indicator 11.3.1	54
4.1	Chapter Overview	54
4.2	General Framework for Uncertainty Propagation and Probabilistic LUE Classification	56
4.2.1	Conceptual Representation of Input Uncertainty	56
4.2.2	Indicator Transformation	57
4.2.3	Distribution-Based Uncertainty Propagation	58
4.2.4	Extension to SDG 11.3.1 Secondary Indicators	60
4.2.5	Probabilistic LUE Classification	60
4.3	Uncertainty Quantification of EO-Derived BUA for SDG Indicator 11.3.1 Monitoring	62
4.3.1	Pipeline Overview	62
4.3.2	Error-Adjusted Area Estimation and Uncertainty Estimation	63
4.3.3	Bootstrap-based Uncertainty Estimation and Propagation	66
4.4	Illustrative Applications	66
4.4.1	Application 1: Hypothetical Application Under Alternative Dependence Scenarios	67
4.4.1.1	Baseline Specification and Marginal Input Uncertainty	67
4.4.1.2	Alternative Dependence Structures	67

4.4.1.3	Joint Distribution Specification and Dependence Modeling	68
4.4.1.4	Monte Carlo Implementation	69
4.4.1.5	Results and Discussion	71
4.4.1.5.1	Distributional Behavior of LCR, PGR, and LCRPGR Across Dependence Scenarios	71
4.4.1.5.2	Summary Statistics and Deterministic Agreement	72
4.4.1.5.3	Probabilistic LUE Classification of City A	73
4.4.1.6	Application 1 Summary	75
4.4.2	Application 2: Uncertainty-Aware SDG Indicator 11.3.1 Monitoring from EO- Derived Built-Up Data	75
4.4.2.1	Study Area Description	76
4.4.2.2	Temporal Scope and Built-up and Population Datasets	76
4.4.2.3	Sampling Design and Sample Size Planning	78
4.4.2.4	Reference Data and Sample Labeling	79
4.4.2.5	Implementation of Error-Adjusted Area Estimation and Stratified Bootstrap Uncertainty Propagation	80
4.4.2.6	Results and Discussion	81
4.4.2.6.1	Conventional SDG Indicator 11.3.1 Results from Map-Based Pixel Counting (Fixed BUA)	81
4.4.2.6.2	Error-Adjusted SDG Indicator 11.3.1 Results from Post- stratified Estimation (Full Sample)	81
4.4.2.6.3	Uncertainty-Aware SDG Indicator 11.3.1 Results (Bootstrap- Based)	82
4.4.2.7	Application 2 Summary	84
4.5	Discussion	84
4.6	Limitations and Further Work	86
4.7	Chapter Conclusion	87
5	Data-Driven, Uncertainty-Aware Deep Learning Forecasting of Built-Up Area and Population for Forward-Looking Probabilistic Assessment of LUE	88
5.1	Chapter Overview	88
5.2	Related Work	89
5.3	Methods	92
5.3.1	Overview	92
5.3.2	Notation and Data Representation for Multiple Localities	92
5.3.3	Forecasting Problem Formulation	93
5.3.4	Global Joint-Target Sequence-to-Sequence Learning Setup	93
5.3.5	Probabilistic Forecasting via Quantiles	94
5.3.6	Deep Learning Architectures for Global Joint BUA–Pop Forecasting	94
5.3.6.1	1D CNN	95
5.3.6.2	LSTM	96
5.3.6.3	GRU	97
5.3.6.4	Hybrid Architectures	98

5.3.6.5	Transformer	98
5.3.7	Model Training Objective: Multi-Horizon Quantile Loss	99
5.3.8	Deep Ensemble Learning	99
5.3.9	Baseline Model: Exponential Smoothing	100
5.3.10	Model Performance Evaluation Strategy and Metrics	101
5.3.11	Model Selection for Downstream SDG Indicator 11.3.1 Propagation	103
5.3.12	Forecast Uncertainty Propagation to SDG Indicator 11.3.1	104
5.3.13	Dependence Modeling of Forecast Uncertainties	105
5.3.14	Probabilistic LUE Classification	107
5.4	Global Application to Urban Centers	107
5.4.1	Data Source	107
5.4.2	Data Partitioning and Normalization	108
5.4.3	DL Model Design and Training	110
5.4.3.1	1D CNN	111
5.4.3.2	LSTM	111
5.4.3.3	GRU	111
5.4.3.4	1D CNN+LSTM	112
5.4.3.5	1D CNN+GRU	112
5.4.3.6	Transformer	112
5.4.4	Deep Ensembling	113
5.4.5	Baseline Modeling Using ES	113
5.4.6	Model Evaluation and Selection	114
5.4.7	Forecast Uncertainty Propagation to SDG Indicator 11.3.1 and Probabilistic LUE Classification on the Test Set	115
5.4.8	Evaluation of Projected SDG Indicator 11.3.1 Estimates and Probabilistic LUE Classification	116
5.4.9	Forecast-Based Projection of SDG Indicator 11.3.1 for 2025–2030	116
5.5	Results	117
5.5.1	Forecasting Performance on the 2015–2020 Evaluation Horizon	117
5.5.1.1	BUA Point Forecast Accuracy and Prediction-Interval Performance	117
5.5.1.2	Pop Point Forecast Accuracy and Prediction-Interval Performance	118
5.5.1.3	Short- and Medium-Term Forecast Performance	119
5.5.2	Illustrative UC-Level Forecast Examples	120
5.5.3	Statistical Comparison and Final Model Selection	123
5.5.4	Deep Ensemble Behavior and Stability	124
5.5.5	Forecast Uncertainty Dependence Modeling	126
5.5.6	Effects of Forecast Uncertainty on SDG Indicator 11.3.1 Estimates	126
5.5.7	Predictive Agreement of Projected LCR, PGR, and LCRPGR with Observed Values	129
5.5.8	Characteristics and Accuracy of Projected Probabilistic LUE Classification	130
5.5.9	SDG Indicator 11.3.1 Projections for 2025–2030	131
5.5.10	Probabilistic LUE Classification and Confidence Patterns	134

5.6	Discussion	138
5.6.1	Implications of Formulating Joint BUA–Pop Forecasting as a Global Pooled Learning Task	138
5.6.2	Comparative Behavior of the Candidate DL Architectures	140
5.6.3	Role of Deep Ensembling in Stabilizing Predictive Uncertainty	140
5.6.4	Added Value of the Global DL Workflow Relative to the ES Baseline	141
5.6.5	Implications for Forecast–Uncertainty Propagation and Probabilistic LUE Assessment	142
5.6.6	Interpretation of the 2020–2025 and 2025–2030 Projections	143
5.6.7	Synthesis in Relation to RQ4	144
5.7	Limitations and Further Work	145
5.8	Chapter Conclusion	147
6	Enhancing LUE Assessment Through Built-Up Area–Built-Up Volume Trajectories	148
6.1	Chapter Overview	148
6.2	Related Work	148
6.3	The BUA–BUV Trajectory Framework	150
6.4	Operationalizing the BUA–BUV Framework for Typology Classification	153
6.5	Comparison with Other Trajectory Typology Classifications	154
6.6	Interpretation in the Context of SDG Indicator 11.3.1	155
6.7	Application to Global Urban Centers for Enhanced LUE Assessment	156
6.7.1	Dataset Used, Preparation, and Temporal Scope of Analysis	156
6.7.2	Conventional LUE Assessment	157
6.7.3	Trajectory Typology Assignment and Efficiency Evaluation	158
6.7.4	Statistical Assessment of Distributional Differences	158
6.7.5	Sensitivity Analysis	158
6.8	Results	159
6.8.1	Temporal Trends in LCR, PGR, and LCRPGR	159
6.8.2	Spatiotemporal Patterns of LUE	160
6.8.3	BUA–BUV Trajectories of UCs	160
6.8.4	Prevailing Built Forms of UCs	160
6.8.5	Trajectory Typologies of UCs	164
6.8.6	Linking Urban Growth Typologies to Efficiency Outcomes	164
6.8.7	Sensitivity of the BUA–BUV Trajectory Framework	167
6.8.7.1	Effect of Thresholds on Prevailing Built Form	167
6.8.7.2	Effect of Slope Parameters on Trajectory Typology	169
6.9	Discussion	170
6.9.1	Enhanced Interpretation of Efficiency Outcomes Through BUV Integration	170
6.9.2	Global Patterns of Built Form and Growth Modality	171
6.9.3	Intersection of Structural Dynamics and Efficiency Metrics	171
6.9.4	Structural Sustainability and Policy Implications	172
6.10	Limitations and Further Work	173

6.11 Chapter Conclusion	174
7 Synthesis, Summary of Implications, and Future Research Directions	176
7.1 Synthesis of Key Findings Across Research Questions	176
7.2 Summary of Implications for SDG Indicator 11.3.1 Monitoring and Broader Urban Sustainability Assessment	177
7.3 Future Research Directions	180
Bibliography	181
Appendix	205
Curriculum Vitae	209

Abbreviations

1D CNN	One-Dimensional Convolutional Neural Network
AIC	Akaike Information Criterion
AICc	Corrected Akaike Information Criterion
B3DI	Building 3D Density Index
BUA	Built-up area
BUA-BUV	Built-up area-Built-up volume
BU _p C	Built-up area per capita
BUV	Built-up volume
CA	Cellular Automata
CDF	Cumulative Distribution Function
Conv1D	One-Dimensional Convolution
CRS	Coordinate Reference System
DEGURBA	Degree of Urbanisation
DL	Deep Learning
dof	Degrees of Freedom
EO	Earth Observation
ES	Exponential Smoothing
ETS	Error-Trend-Seasonality
GAUD	Global Annual Urban Dynamics
GHS-BUILT-S	Global Human Settlement-Built-up Surface
GHSL	Global Human Settlement Layer
GHS-POP	Global Human Settlement-Population
GHS-UCDB	Global Human Settlement-Urban Centre Database
GIS	Geographic Information System
GISA	Global Impervious Surface Area
GISD30	Global 30 m Impervious-Surface Dynamic dataset
GLAD	Global Land Analysis and Discovery
GLCLUC	Global Land Cover and Land Use Change
GRU	Gated Recurrent Unit
IQR	Interquartile Range
LCR	Land Consumption Rate
LCRPGR	Ratio of Land Consumption Rate to Population Growth Rate
LGU/LGUs	Local Government Unit/Units
LSTM	Long Short-Term Memory
LUE	Land-Use Efficiency

MAD	Median Absolute Deviation
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MOE	Margin of Error
OFAT	One-Factor-At-A-Time
PGR	Population Growth Rate
PIAW	Prediction Interval Average Width
PICP	Prediction Interval Coverage Probability
PIW	Prediction Interval Width
Pop	Population
PSA	Philippine Statistics Authority
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
RQ	Research Question
SDG/SDGs	Sustainable Development Goal/Goals
SSP/SSPs	Shared Socioeconomic Pathway/Pathways
TCBUA	Total Change in Built-Up Area
UC/UCs	Urban Center/Centers
UN	United Nations
VHG	Vertical-to-Horizontal Growth
WB	World Bank
WS	Winkler Score
WSF-Evolution	World Settlement Footprint Evolution
WV02	WorldView-2

1 Introduction

1.1 Background and Motivations

Urbanization is among the most consequential processes shaping contemporary human–environment systems (Bai et al., 2017). Cities concentrate population, economic activity, and infrastructure, while simultaneously transforming land systems, altering ecosystems, and intensifying resource demands. As urban areas expand and intensify, a central challenge for sustainable development is not merely the pace of urban growth, but how efficiently land resources are consumed relative to population change. This challenge lies at the intersection of spatial planning, environmental sustainability, and social equity, making the efficient use of land a critical concern for both scientific analysis and policy action (Melchiorri et al., 2018; UN-Habitat, 2018b; Khan et al., 2025).

Within this context, Sustainable Development Goal (SDG) Indicator 11.3.1 was introduced by the United Nations (UN) to provide a globally comparable measure of land-use efficiency (LUE) (Figure 1.1). Defined as the ratio of the land consumption rate (LCR) to the population growth rate (PGR), commonly denoted as LCRPGR, the indicator operationalizes LUE by relating horizontal changes in built-up area to changes in population over time (UN Statistics Division, 2025). Its conceptual appeal lies in its simplicity: efficiency is achieved when population growth is accommodated without disproportionate expansion of built-up area (Estoque et al., 2021; UN-Habitat, 2018b). This formulation has enabled widespread adoption across national reporting systems, international urban assessments, and academic research, positioning SDG Indicator 11.3.1 as a central reference point for evaluating progress toward sustainable urbanization (UN-Habitat, 2018b; UN Statistics Division, 2025). This widespread adoption has been facilitated by the increasing availability of Earth observation (EO)–derived time series that enable consistent, spatially explicit measurement of built-up area and population change across large spatial extents and long temporal horizons.

1.1.1 The Uncertainty of Indicator-Based Assessment

The increasing use of SDG Indicator 11.3.1 for policy-relevant evaluation has raised fundamental questions about how such indicators represent urban processes and support decision-making (Estoque et al., 2021; Schiavina et al., 2019). Indicator-based assessments necessarily simplify complex and dynamic systems, enabling comparisons and benchmarking across cities and regions. Urban systems, however, are inherently uncertain, dynamic, and multidimensional (Yang & Zhang, 2025). In practice, these assessments are commonly implemented and interpreted deterministically, with built-up area, population, and derived indicators treated as exact values (UN Statistics Division, 2025).

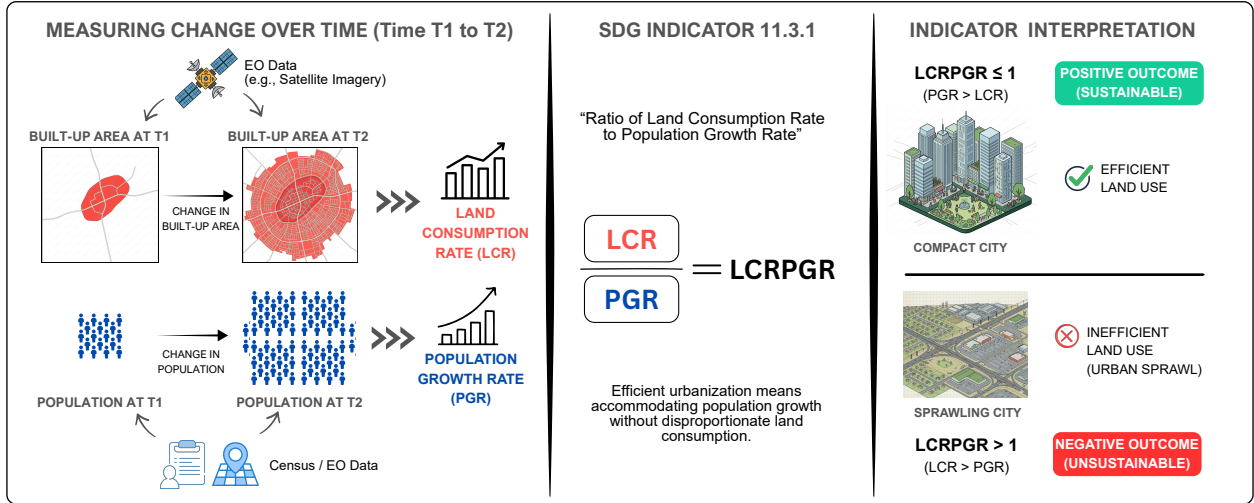


Figure 1.1: Conceptual overview of SDG Indicator 11.3.1 for monitoring land-use efficiency (LUE). Built-up area and population changes between two time points are used to calculate LCR and PGR, whose ratio defines LCRPGR. The interpretation panel assumes that both LCR and PGR are positive; under this condition, $LCRPGR \leq 1$ indicates relatively efficient land use, while $LCRPGR > 1$ indicates relatively inefficient land use.

This deterministic framing contrasts with the nature of the quantities underlying SDG Indicator 11.3.1. The built-up area and population data used to calculate LCR and PGR are modeled representations derived from heterogeneous data sources, including EO-derived built-up products, census or administrative population counts, gridded population models, and ancillary spatial covariates. As such, both inputs embed assumptions regarding definition, measurement, spatial allocation, temporal harmonization, and aggregation (Chakraborty et al., 2024; Leyk et al., 2019). While these input data characteristics are widely recognized, conventional applications typically rely on point estimates and do not explicitly account for uncertainty in these inputs.

The implications of this simplification are especially pronounced for ratio-based indicators such as SDG Indicator 11.3.1 because uncertainty in either input component can be amplified nonlinearly in the resulting value. Small variations in built-up area or population may therefore lead to disproportionately large changes in the indicator, particularly when population growth is low and the denominator becomes small, resulting in shifts in the interpreted LUE (Calka et al., 2022; Zhong et al., 2025b). When such values are treated as precise measures of LUE performance, there is a risk that apparent differences across cities or periods reflect data uncertainty rather than substantive differences in urban development dynamics.

As SDG Indicator 11.3.1 increasingly informs assessments of urban development progress and LUE, and as it is being extended to support more equitable and policy-relevant comparisons across cities and regions (Cai et al., 2020; Mudau et al., 2020; Schiavina et al., 2022b), this implicit assumption of precision becomes a critical limitation. Ensuring that indicator-based LUE assessments are robust to uncertainty is therefore a central motivation for re-examining how SDG Indicator 11.3.1 is calculated, interpreted, and applied in policy contexts.

1.1.2 The Need for Forward-Looking Assessment

A further challenge arises from the predominantly retrospective orientation of existing indicator-based assessments. While SDG Indicator 11.3.1 is intended to support long-term sustainability objectives, its application has largely focused on summarizing past patterns of urban expansion and population change rather than using historical evidence to assess plausible future land-use pressures and urban development trajectories (Schiavina et al., 2022b). As a result, indicator values are commonly interpreted as assessments of historical performance, offering limited insight into how current trajectories may evolve under continued urbanization or under alternative future conditions.

This retrospective focus constrains the relevance of SDG Indicator 11.3.1 for planning and decision-making in rapidly changing urban contexts. Urban development decisions that encompass infrastructure investment, land-use regulation, and climate adaptation are inherently forward-looking. Any forward-looking assessment must necessarily build on existing, and therefore historical, data; however, its value lies not only in describing past change but in using those data to evaluate plausible future pathways. These decisions must be made under uncertainty regarding future population growth, land availability, and development intensity. Indicators that rely exclusively on historical observations may provide little guidance on the range of plausible future efficiency outcomes, the uncertainties surrounding them, or the risks associated with alternative development paths and scenarios.

As cities face increasingly complex challenges related to infrastructure provision, climate adaptation, and land scarcity (UN-Habitat, 2020), decision-makers require analytical tools that move beyond descriptive assessment of past performance. Forward-looking, risk-aware evaluation of LUE is needed to support proactive planning and to anticipate potential inefficiencies before they become established. This necessity motivates extending SDG Indicator 11.3.1 beyond a purely retrospective monitoring instrument. Grounded in historical data, such an extension can support prospective, uncertainty-aware assessment of urban development trajectories through forecasting, scenario-based analysis, and the explicit treatment of uncertainty. This extension is important both for informing the remaining SDG implementation period up to 2030 and for sustaining evidence-based monitoring and planning beyond it.

1.1.3 The Dimensional Constraints of Planar Metrics

The conceptual scope of LUE extends beyond two-dimensional land consumption. While SDG Indicator 11.3.1 evaluates efficiency through changes in planar built-up area relative to population growth, this representation captures only one aspect of urban development (Estoque et al., 2021). In many contemporary urban contexts, population growth is increasingly accommodated through vertical intensification, such as higher building densities, increased floor space, and taller structures, rather than through outward horizontal expansion (Liu et al., 2024; Froking et al., 2024). Therefore, indicators that focus exclusively on planar changes in built-up area risk obscuring critical aspects of how cities grow and accommodate population. Urban areas with similar LCRPGR values may exhibit fundamentally different development patterns, with some relying on low-rise

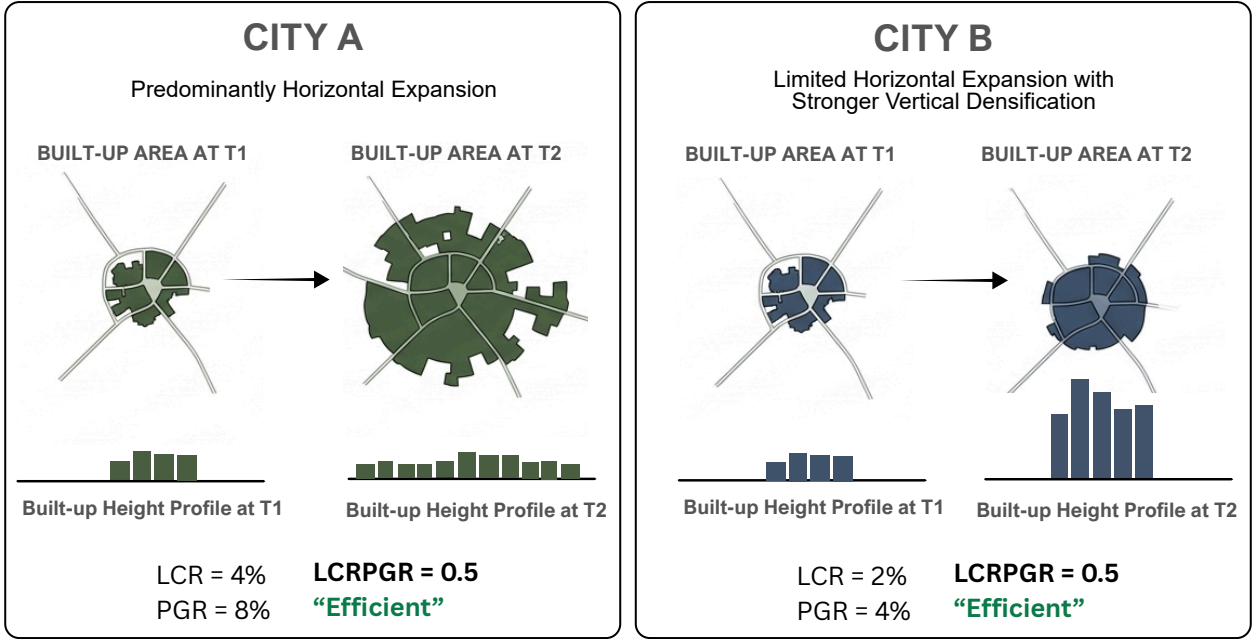


Figure 1.2: Conceptual illustration of two cities with identical LCRPGR values but contrasting modes of urban growth. City A accommodates population increase mainly through horizontal expansion, whereas City B accommodates growth through more limited planar expansion combined with stronger vertical densification. Despite these different development patterns, both cities yield the same LCRPGR value and LUE interpretation, highlighting how area-based indicators alone can obscure important structural differences in urban trajectories.

horizontal expansion and others on vertical densification (Figure 1.2). When these distinctions are not accounted for, efficiency outcomes derived from area-based metrics alone may be difficult to interpret, as they conflate structurally distinct urban trajectories into a single numerical value.

This dimensional limitation constrains the explanatory power of SDG Indicator 11.3.1, particularly in contexts where vertical development plays a central role in managing land scarcity, infrastructure demand, and environmental impacts (Grace Wong, 2004; Munn & Dragičević, 2021). Without complementary representations that capture both horizontal and vertical dimensions of growth (Koomen et al., 2009; Bruyns et al., 2021), indicator-based assessments remain incomplete, limiting their ability to explain how LUE is achieved and to support differentiated interpretation of urban development pathways.

Overall, the uncertainty of indicator inputs, the predominantly retrospective orientation of SDG Indicator 11.3.1 applications, and the planar treatment of urban growth reveal a deeper structural issue in current SDG Indicator 11.3.1-based assessments. The indicator is commonly interpreted as a precise, retrospective, and planar measure of LUE, despite being derived from uncertain inputs, applied to inherently dynamic urban systems, and blind to vertical modes of urban development. As a result, SDG Indicator 11.3.1 is often interpreted and applied in ways that go beyond what its prevailing analytical framing can reliably support, particularly in comparative, evaluative, and policy-oriented contexts. Addressing these limitations does not require abandoning the indicator; rather, it requires rethinking how it is calculated, interpreted, and extended. Figure 1.3 summarizes this conceptual motivation and frames the dissertation’s analytical direction.

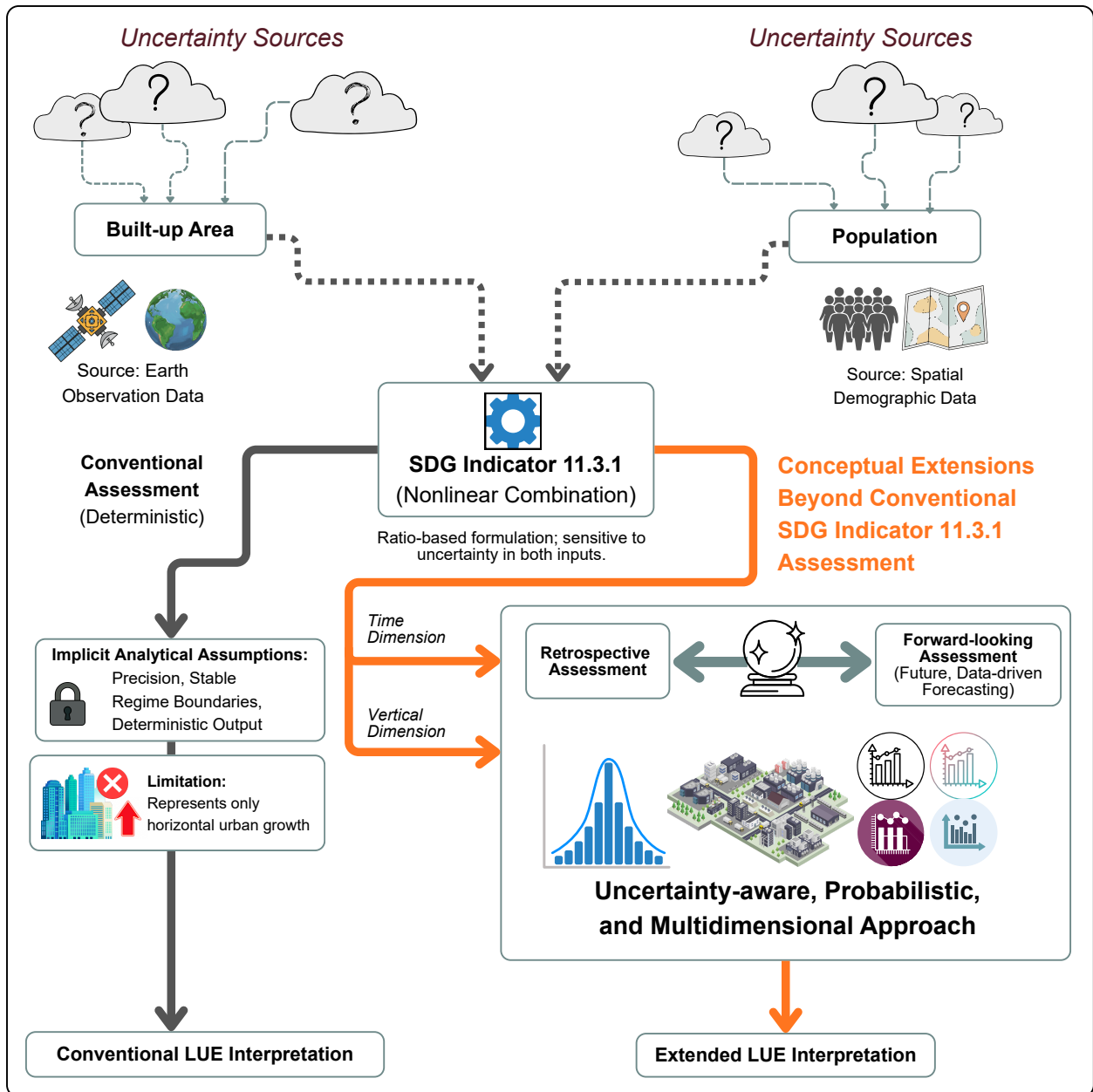


Figure 1.3: Conceptual framing of the dissertation, showing the transition from conventional deterministic SDG Indicator 11.3.1 assessment toward an uncertainty-aware, probabilistic, forward-looking, and multidimensional framework for LUE assessment.

1.2 Central Thesis Statement

This dissertation advances the thesis that prevailing uses of SDG Indicator 11.3.1 remain insufficient for a reliable, policy-relevant assessment of urban LUE because they treat the indicator as deterministic, retrospective, and two-dimensional. Built-up area and population are uncertain quantities whose errors propagate nonlinearly through LCR, PGR, and LCRPGR, affecting both indicator magnitudes and LUE classification. At the same time, contemporary urban growth is increasingly shaped by vertical intensification, which area-only indicators cannot capture. These limitations motivate the development of uncertainty-aware, forward-looking, and multidimensional analytical frameworks to improve robustness in LUE assessment.

1.3 Objective and Research Questions

The general objective of this dissertation is to develop, implement, and empirically evaluate analytical frameworks for assessing urban LUE in the context of SDG Indicator 11.3.1. These frameworks are designed to be uncertainty-aware, forward-looking, and multidimensional.

To achieve this objective, the dissertation addresses the following research questions (RQs):

- RQ1. How sensitive are SDG Indicator 11.3.1 values and LUE interpretations to uncertainty in built-up area and population arising from measurement, definition, and data-source selection under retrospective conditions?
- RQ2. How does explicit propagation of uncertainty in built-up area and population affect the magnitude, stability, and LUE classification outcomes of SDG Indicator 11.3.1 compared to conventional deterministic approaches?
- RQ3. Given uncertain SDG Indicator 11.3.1 components and outputs, how does probabilistic LUE classification improve the robustness, comparability, and interpretability of LUE assessments relative to deterministic classification?
- RQ4. To what extent can data-driven, uncertainty-aware forecasting models accurately and reliably predict built-up area and population time series across diverse localities, while producing predictive uncertainty that enables subsequent SDG Indicator 11.3.1-based LUE assessments to be well-calibrated and decision-relevant?
- RQ5. To what extent does integrating built-up area (BUA) and built-up volume (BUV) within a trajectory-based framework enhance the interpretability and robustness of SDG Indicator 11.3.1-based LUE assessment?

1.4 Contributions

The contributions of this dissertation are conceptual, methodological, and analytical:

- Conceptually, the dissertation reframes SDG Indicator 11.3.1 as an uncertainty-sensitive, distributional measure rather than a deterministic scalar. It shows that uncertainty in built-up area and population propagates nonlinearly through LCR, PGR, and especially LCRPGR, affecting indicator magnitude, stability, and LUE regime¹ interpretation, thereby challenging the implicit assumptions of precision common in deterministic SDG Indicator 11.3.1 applications.

¹Here, and throughout this dissertation, “regime” is used strictly in an analytical sense. In the context of LUE classification, it denotes a distinct interpretive state of urban LUE defined by a specific combination of LCR, PGR, and LCRPGR conditions. Accordingly, regime membership refers to the assignment, or probability of assignment, of a spatial entity (e.g., a city, urban center, or country) to one of these defined LUE states.

- Methodologically, the dissertation operationalizes probabilistic workflows for SDG Indicator 11.3.1 assessment that integrate (i) uncertainty estimation for built-up area from design-based reference sampling and error-adjusted area estimation, (ii) uncertainty propagation through the indicator transformations, (iii) uncertainty-aware forecasting of built-up area and population from empirical urban time series, and (iv) probabilistic LUE classification. A key contribution is the end-to-end integration of dependence-aware uncertainty propagation—treating built-up area and population as a joint uncertain system rather than independent inputs—with LUE regime membership probability outputs, applied consistently across both retrospective monitoring and forecast-based assessment.
- Analytically, the dissertation develops a trajectory-based framework that integrates BUA and BUUV, enabling horizontal expansion and vertical intensification to be represented in a unified analytical space. This framework reveals structural growth patterns that are not captured by area-based indicators alone and complements SDG Indicator 11.3.1-based assessment with a multidimensional context.

These contributions establish an integrated foundation for an uncertainty-aware, forward-looking, and multidimensional assessment of urban LUE, thereby extending both the scientific interpretation and the policy relevance of SDG Indicator 11.3.1.

2 Conceptual Foundations and Related Work

2.1 Conceptual Foundations

2.1.1 LUE and SDG Indicator 11.3.1

In its broadest sense, land-use efficiency (LUE) refers to the capacity of land systems to coordinate competing demands for food production, socio-economic development, and ecological maintenance under increasing population growth, economic expansion, and urbanization pressures, to mitigate land-use conflicts and promote sustainable development (Liu et al., 2020a). Within this framing, LUE is understood as a dynamic concept aligned with the sustainable development paradigm, shaped by the combined influences of natural conditions, economic activities, and social processes rather than by static land-use outcomes alone (Liu et al., 2020a; Zhu et al., 2019).

Within the UN SDGs framework, this broad, multidimensional understanding of LUE is translated into an operational concept suitable for global comparison and monitoring. In this framework, LUE is recognized as a cross-cutting concept that influences multiple aspects of sustainable development, including food security, economic productivity, and environmental protection (Khan et al., 2025). To enable systematic monitoring, LUE is formalized through a nested indicator hierarchy that links global development objectives to operational monitoring metrics.

At the highest level, the 17 SDGs define a comprehensive agenda for sustainable development. Within this framework, SDG 11 focuses on making cities and human settlements inclusive, safe, resilient, and sustainable. Under SDG 11, Target 11.3 specifically aims to enhance inclusive and sustainable urbanization and capacity for participatory, integrated, and sustainable human settlement planning (UN Department of Economic and Social Affairs, 2024). Indicator 11.3.1 operationalizes this target into a metric called “LCRPGR,” which quantifies LUE through the relationship between built-up expansion (LCR) and population change (PGR) (UN Statistics Division, 2025). Under the UN SDG monitoring framework, this indicator is classified as Tier II, indicating that while its methodological standards are well-established, global data coverage remains inconsistent (CEOS, 2023).

A key motivation for SDG Indicator 11.3.1 is the need to identify and monitor urban sprawl. Urban sprawl is a widely recognized manifestation of inefficient land use in which built-up land expands faster than population growth. This pattern often results in low-density development, higher infrastructure and service provision costs, and increased environmental pressure (UN-Habitat, 2018b; Zhou et al., 2021). By monitoring SDG Indicator 11.3.1, policymakers and stakeholders can receive timely, relevant information. This information can support informed decision-making to promote more inclusive and sustainable urbanization. At a minimum, such monitoring can help

slow the spread of urban sprawl. Where possible, it can also support efforts to maintain or increase urban compactness over time (UN-Habitat, 2018b).

Formally, SDG Indicator 11.3.1 and its components are calculated as follows (UN Statistics Division, 2025):

$$\text{LCR} = \frac{\text{BUA}_{t_2} - \text{BUA}_{t_1}}{\text{BUA}_{t_1}} \cdot \frac{1}{\Delta t} \quad (2.1)$$

$$\text{PGR} = \frac{\ln\left(\frac{\text{Pop}_{t_2}}{\text{Pop}_{t_1}}\right)}{\Delta t} \quad (2.2)$$

$$\text{LCRPGR} = \frac{\text{LCR}}{\text{PGR}} \quad (2.3)$$

Here, BUA denotes the total built-up area and Pop the total population at two epochs (t_1 and t_2), with Δt as the time interval. In the context of SDG Indicator 11.3.1, built-up areas are defined as “all areas occupied by buildings” (UN Statistics Division, 2025). The dimensionless indicator, LCRPGR, is based on rates rather than absolute quantities. As such, it facilitates comparison across spatial scales and time periods. At the same time, its ratio-based structure implies that variation in either built-up area or population can influence the resulting indicator value.

2.1.2 LUE Interpretation in SDG Indicator 11.3.1 Context

An LCRPGR value close to 1 indicates balanced urban expansion, where land consumption increases at a rate similar to population growth. Values larger than one indicate that land consumption grows faster than population and are commonly interpreted as inefficient land use and excessive spatial expansion. Values below one indicate that population growth is faster than land consumption and are usually associated with densification or more compact urban development (UN-Habitat, 2018b).

This simplified interpretation, however, is not generally sufficient and applies only when LCR and PGR are both positive. The official metadata explicitly requires that LCR and PGR be examined separately before interpreting their ratio, as similar LCRPGR values may arise from fundamentally different land–population dynamics (UN Statistics Division, 2025). For example, identical ratio values may reflect simultaneous growth in both components, concurrent decline, or asymmetric change in which one component increases while the other decreases. As illustrated by Figure 2.1 and described further in Table 2.1, meaningful interpretation therefore depends on both the sign and magnitude of each component rate. Based on these illustrations, efficient LUE refers to urban growth or contraction conditions in which land consumption is proportional to, lower than, or more strongly reduced than the observed population change, so that no excess land is taken relative to population dynamics.

From this perspective, SDG Indicator 11.3.1 functions not merely as a summary ratio but as a structured diagnostic indicator. Joint examination of LCR and PGR enables efficiency outcomes

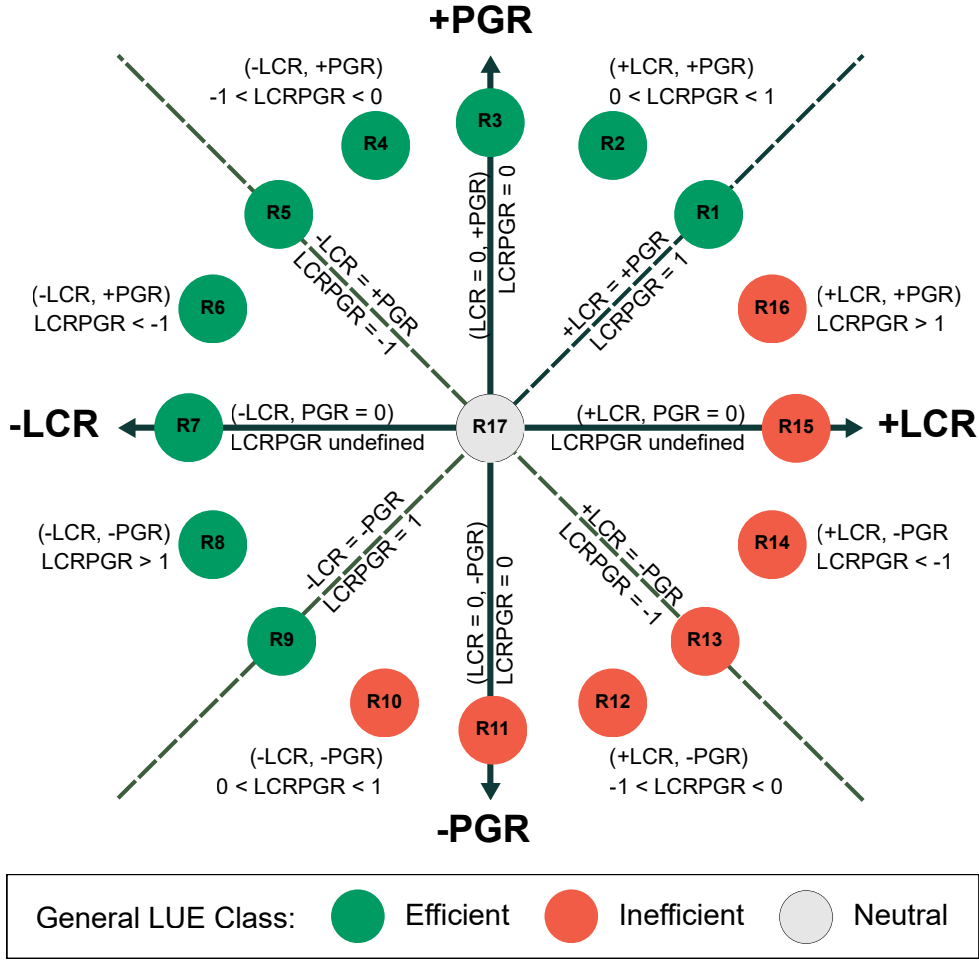


Figure 2.1: Conceptual 17-regime LUE classification space in the LCR–PGR plane. The figure partitions all possible sign and relative-magnitude combinations of LCR, PGR, and LCRPGR into 17 regimes. R1–R9 denote efficient LUE outcomes, where land is not consumed unnecessarily relative to population change. R10–R16 denote inefficient LUE outcomes, where land consumption exceeds or is insufficiently reduced relative to population change. R17 denotes the neutral case, where both LCR and PGR are zero. The corresponding regime definitions, interval rules, and interpretations are summarized in Table 2.1. Figure by the author based on UN Statistics Division (2025), UN-Habitat (2018b), Calka et al. (2022), and Cimini et al. (2023).

to be linked directly to either horizontal expansion/contraction, demographic growth/population decline, or combinations thereof. For this reason, component-wise interpretation is an integral element of the official monitoring methodology.

Conceptually, this interpretation framework assumes that built-up area and population inputs are known with sufficient precision to allow unambiguous classification into growth regimes. Recognizing this deterministic assumption is essential for understanding both the strengths and limitations of current SDG Indicator 11.3.1 monitoring practice. This recognition also provides the methodological motivation for later chapters, which examine how uncertainty in built-up area and population inputs propagates through the indicator and affects the robustness of LUE interpretation.

Table 2.1: Land-use efficient (LUE) regime classification based on the signs and relative magnitudes of LCR, PGR, and LCRPGR. *Continued from previous page.*

Code	LCR	PGR	LCRPGR	Interpretation	General LUE Class	LUE Regime
R12	+	−	$(-1, 0)$	Built-up area expands while population declines, although population decline exceeds built-up expansion in absolute terms. This still reflects inefficient land use, as built-up areas continue to increase despite declining population demand.	Inefficient	Inefficient Built-up Expansion with Population-Decline Dominance
R13	+	−	$= -1$	Built-up expansion and population decline occur at the same rate in opposite directions. Although balanced in magnitude, this remains inefficient because land consumption continues to increase despite a declining population.	Inefficient	Inefficient Proportional Built-up Expansion and Population Decline
R14	+	−	< -1	Built-up area expands faster than population declines. This represents the strongest inefficient expansion under population decline, with clear excess built-up growth relative to demographic need.	Inefficient	Inefficient Built-up Expansion under Population Decline (Expansion-Dominant)
R15	+	$= 0$	Undefined	Built-up area expands while population remains stable. This indicates demographic-detached expansion and unnecessary built-up growth without a corresponding population increase.	Inefficient	Inefficient Built-up Expansion under Stable Population
R16	+	+	> 1	Both built-up area and population grow, but built-up expansion exceeds population growth. This indicates inefficient outward growth or sprawl, where land is consumed faster than needed relative to population increase.	Inefficient	Inefficient Built-up Expansion under Slower Population Growth
R17	$= 0$	$= 0$	Undefined	No detectable change in land consumption and population over the interval.	Neither efficient nor inefficient	Neutral

2.2 State of the Art in SDG Indicator 11.3.1 Monitoring

Since the adoption of the UN SDGs in 2015, SDG Indicator 11.3.1 has been widely implemented as the primary operational tool for empirical assessment of LUE across global, national, and local contexts. A substantial body of literature has employed the indicator to examine LUE within the context of urban land–population dynamics (e.g., Melchiorri et al., 2019; Estoque et al., 2021; Schiavina et al., 2022b; Wang et al., 2020; Faye et al., 2022; Angel et al., 2024; Santillan & Heipke, 2024a). These studies have enabled comparative analyses across cities, regions, and countries and supported large-scale monitoring efforts. This body of work indicates that SDG Indicator 11.3.1 has reached a high level of operational maturity, characterized by standardized calculation procedures, broad data availability, and applicability across diverse urban settings, while maintaining a relatively low implementation barrier.

2.2.1 Earth Observation (EO)-based Data Sources and Implementation Practices

Applications of SDG Indicator 11.3.1 rely on EO data to measure changes in built-up area and population across space and time. In accordance with a series of methodological guidelines issued by UN-Habitat and the UN Statistics Division (UN-Habitat, 2018a, 2018b; UN Statistics Division, 2021, 2025), operational workflows typically emphasize procedural transparency and simplicity. At a minimum, calculating LCRPGR requires delineating a spatial unit of analysis (e.g., an urban center, metropolitan area, city, or country) and corresponding built-up area and population values for at least two reference years. This simplicity has facilitated the broad adoption of the indicator across diverse geographical and institutional contexts.

Although SDG Indicator 11.3.1 is formally standardized, differences persist in how the LCR component is implemented across studies. Earlier applications, particularly those conducted before the 2021 revision of the official metadata (UN Statistics Division, 2021, 2025), commonly used a logarithmic form of the LCR as described in the 2018 guidelines (UN-Habitat, 2018a, 2018b). While the updated metadata replaced this with a linear rate-based formulation, several studies published after 2021 continue to apply the earlier logarithmic version (e.g., Gao et al., 2022; Bhandari et al., 2023; Yiannakou & Zografos, 2025; Liu et al., 2025; Zhong et al., 2025a), often citing older documentation or previous literature. As a result, LCR and LCRPGR values reported in the literature are not always directly comparable, even when similar EO data sources are used. At the same time, some studies explicitly justify their continued use of the logarithmic formulation. For example, Angel et al. (2024) adopt a logarithmic-based LCR to maintain consistency with the logarithmic formulation of PGR, thereby facilitating component-wise comparability between land and population growth rates.

Despite differences in formula choice, spatial scale, and study objectives, existing SDG Indicator 11.3.1 applications show strong similarity in the EO-based data sources used to represent built-up area. Among global products, the Global Human Settlement Layer (GHSL) is the most widely applied dataset for SDG Indicator 11.3.1 monitoring at global and regional scales (e.g., Estoque et al., 2021; Laituri et al., 2021; Schiavina et al., 2022b; Liu et al., 2025). GHSL is particularly valuable

in data-scarce contexts where reliable local statistics are unavailable (e.g., Santillan & Heipke, 2023). Its extensive temporal coverage (1975–2020), combined with harmonized layers of built-up area (GHS-BUILT-S), population (GHS-POP), and urban–rural classification (GHS-SMOD), enables the consistent derivation of land consumption and population metrics within aligned spatial units. These characteristics substantially lower implementation barriers and support multi-scale analyses under an open-access policy. As a result, GHSL is widely regarded as the state of the art for SDG Indicator 11.3.1 monitoring and has been identified as a key dataset capable of supporting the transition of the indicator from Tier II to Tier I by providing a global baseline for the required variables (Melchiorri et al., 2019; Estoque et al., 2021).

In addition to GHSL, other global impervious surface and land-cover datasets, such as GlobeLand30, the Global Artificial Impervious Area (GAIA), the Global Land Cover with Fine Classification System (GLC_FCS), and the Global Impervious Surface Area (GISA), have been used to estimate built-up area and calculate LCR values (e.g., Cai et al., 2020; Bhandari et al., 2023; Bounoua et al., 2023; Huang et al., 2024; Zhong et al., 2025b). At regional and national scales, existing land-use or land-cover (LULC) maps produced by governmental agencies are also utilized for LCR calculations (e.g., Nicolau et al., 2019; Ling et al., 2023). In parallel, many studies generate built-up area maps directly from satellite imagery using supervised classification and related digital image analysis techniques, often following the spatial processing steps recommended in the SDG Indicator 11.3.1 metadata and methodological guidelines. These approaches range from traditional machine-learning methods, such as Support Vector Machines and Random Forests, to more recent deep learning-based workflows for impervious surface mapping (e.g., Ghazaryan et al., 2021; Han et al., 2022; Jalilov et al., 2021; Li et al., 2020). Some studies further combine multiple EO datasets and ancillary information within more integrated processing frameworks to improve the accuracy and precision of built-up data (e.g., Jiang et al., 2021; Zhou et al., 2021; Liu et al., 2023; Cardenas-Ritzert et al., 2024a; Xu et al., 2024).

For population inputs, SDG Indicator 11.3.1 applications commonly rely on census-derived population counts when the unit of analysis aligns with administrative or census boundaries (e.g., Aquilino et al., 2021; Cai et al., 2020; Jalilov et al., 2021; Koroso et al., 2021; Bhandari et al., 2023). Population estimates from international sources, such as the World Population Prospects, are also used, particularly in national- and global-scale studies (e.g., Estoque et al., 2021; Jiang et al., 2021; Li et al., 2021b; Faye et al., 2022). In contrast, studies employing non-administrative spatial units (e.g., urban areas, functional urban areas, metropolitan regions, or urban agglomerations) often lack directly available census aggregates. In such cases, population inputs are derived from gridded population datasets, including GHS-POP and WorldPop, or through dasymetric mapping and spatial disaggregation techniques (e.g., Zhou et al., 2021; Schiavina et al., 2022b; Holobacă et al., 2022; Wang et al., 2022a; Cardenas-Ritzert et al., 2024a). Even in analyses conducted at administrative levels, gridded population datasets are sometimes preferred to ensure spatial consistency with EO-derived built-up area data or to compensate for incomplete, outdated, or inconsistent census information (e.g., Calka et al., 2022; Santillan & Heipke, 2024a).

2.2.2 Result Interpretation and Reporting Practices

SDG Indicator 11.3.1 monitoring results are commonly reported at the level of a defined spatial unit, such as an administrative unit (e.g., city, municipality, province, or country), an urban center, a metropolitan region, or other forms of urban agglomeration. For official SDG reporting, UN-Habitat distributes a standardized reporting template to countries on an annual basis for the submission of the most recent SDG Indicator 11.3.1 data at the city and national levels (UN-Habitat, 2024). The reported information typically includes built-up area, population, LCR, PGR, and LCRPGR for each reporting unit and analysis period.

In the research literature, a common reporting format is a tabular presentation listing LCR, PGR, and LCRPGR for each unit and time interval (e.g., Schiavina et al., 2019, 2022b; Zhou et al., 2021; Cimini et al., 2023). In some cases, these tables are further organized according to broader analytical groupings, such as administrative level, development context (e.g., World Bank income groups or SDG regions), settlement characteristics, and city size (e.g., Schiavina et al., 2019; Wang et al., 2020; Estoque et al., 2021; Huang et al., 2024).

Two main approaches are used to summarize indicator values within such groupings. The first relies on aggregation, where built-up area and population are first summed within each group and then used to calculate group-level LCR, PGR, and LCRPGR values. This approach is commonly applied in regional, national, or global comparisons and supports high-level synthesis of built-up–population dynamics across large and heterogeneous sets of urban areas (e.g., Schiavina et al., 2019; Estoque et al., 2021; Schiavina et al., 2022b). The second approach focuses on the distribution of unit-level indicator values within each group. In this case, summaries are usually based on measures of central tendency, most often the mean or median, sometimes accompanied by standard deviation to describe variability (e.g., Nicolau et al., 2019; Li et al., 2021a; Santillan & Heipke, 2024a). Some studies further present histograms of LCRPGR values to examine distributions before interpretation (e.g., Calka et al., 2022).

Interpretation of SDG Indicator 11.3.1 monitoring results is typically conveyed through explanatory text or tables that clarify how the ratio values should be interpreted. In most applied studies, LCRPGR values larger than 1 are interpreted as indicating that land consumption is growing faster than population, while values below one are associated with relatively more compact or population-efficient development (e.g., Schiavina et al., 2019; Estoque et al., 2021; Schiavina et al., 2022b). Several studies go beyond this basic interpretation by adopting regime-based classifications that explicitly consider the signs and magnitudes of both LCR and PGR. These classifications distinguish, for example, efficiency driven by population densification, inefficiency driven by outward land expansion, and cases where built-up area continues to increase despite population decline (e.g., Jiang et al., 2021; Calka et al., 2022; Santillan & Heipke, 2024a). By considering the component rates alongside the ratio, such approaches support a more detailed interpretation of urban development trajectories. In addition, some studies examine the temporal evolution of LCRPGR values to assess progress toward SDG Indicator 11.3.1, focusing on whether spatial units show improvement or deterioration in LUE over time in relation to observed urban growth patterns (e.g., Schiavina et al., 2022b; Liu et al., 2025).

Cartographic and statistical visualizations typically support reporting. Dot maps and choropleth maps of LCR, PGR, and LCRPGR are widely used to display spatial patterns and regional contrasts (e.g., Melchiorri et al., 2019; Wang et al., 2020; Schiavina et al., 2022b; Calka et al., 2022; Huang et al., 2024; Santillan & Heipke, 2024a; Liu et al., 2025). In many cases, LCRPGR maps classify spatial units into efficiency-related categories or value ranges, distinguishing “efficient” spatial units from inefficient ones. In contrast, others apply multi-level classifications ranging from very low to high efficiency (e.g., Faye et al., 2022). Additional summaries include pie charts or bar charts showing the proportion of spatial units within specific LCRPGR classes for a given grouping (e.g., Estoque et al., 2021; Schiavina et al., 2022), scatterplots of LCR versus PGR to explore their relationship and implications for LUE assessment (e.g., Wang et al., 2020; Estoque et al., 2021; Ghazaryan et al., 2021), and time-series plots illustrating changes in LCR, PGR, and LCRPGR across successive intervals (e.g., Jiang et al., 2021; Ling et al., 2023; Zhou et al., 2021). Web-based mapping platforms are also increasingly used to visualize the status and progress of SDG Indicator 11.3.1, particularly to support interactive exploration and communication of results (e.g., Santillan & Heipke, 2024b).

2.2.3 Methodological Characteristics and Known Challenges

Existing studies on SDG Indicator 11.3.1 document a set of methodological characteristics that influence how the indicator behaves across different spatial, temporal, and data contexts. These characteristics arise from both the mathematical structure of the indicator and from common choices in EO-based data preparation and implementation workflows, and they have been examined across a growing body of empirical applications.

2.2.3.1 Mathematical Behavior and Interpretation of LCRPGR under Population Change

A recurrent topic in the literature on SDG Indicator 11.3.1 concerns the mathematical behavior and interpretability of LCRPGR across different population change regimes. Because LCRPGR is formulated as a ratio between land consumption and population growth rates, its numerical properties are directly influenced by both the magnitude and sign of population change.

One commonly documented feature is the occurrence of negative LCRPGR values, which arise when the LCR and PGR have opposite signs. In empirical applications, this most frequently occurs when built-up area continues to expand while population declines, a pattern observed across diverse geographical contexts and population datasets (e.g., Nicolau et al., 2019; Cai et al., 2020; Li et al., 2021a; Calka et al., 2022; Angel et al., 2024; Ma et al., 2024). Negative LCR values, reflecting net reductions in built-up area, are theoretically possible but relatively uncommon in city-scale applications and therefore contribute less frequently to negative LCRPGR outcomes in practice. In some studies, negative LCR values are attributed to inconsistencies or errors in built-up area data sources rather than to genuine land contraction processes (e.g., Aquilino et al., 2021; Calka et al., 2022). While negative LCRPGR values clearly indicate a decoupling between land consumption and population change, their substantive interpretation is not straightforward. As highlighted by Angel et al. (2024), it is difficult to assess whether a larger negative value corresponds to a more

favorable LUE outcome than a smaller one, limiting the usefulness of negative LCRPGR magnitudes for comparative evaluation.

Beyond sign-related ambiguity, the literature has also emphasized that LCRPGR is unbounded by construction. Angel et al. (2024) formalize this property by demonstrating that no theoretical upper or lower limits constrain the indicator’s magnitude. As population growth approaches zero, LCRPGR values can increase or decrease without bound, regardless of the absolute scale of land consumption. Empirical studies show that LCRPGR values can vary widely across cities and metropolitan areas, reaching very large positive or negative magnitudes that complicate interpretation and limit meaningful comparison across spatial contexts (e.g., Li et al., 2021a; Calka et al., 2022; Ma et al., 2024). This unboundedness is intrinsic to the ratio-based formulation of LCRPGR and holds independently of data source, spatial resolution, or implementation workflow. Additionally, Zhong et al. (2025b) demonstrated that a minimal demographic change—that is, when population growth is close to zero—can trigger sharp transitions between positive and negative efficiency regimes, reflecting nonlinear amplification effects inherent in the indicator’s mathematical structure.

To support interpretation considering these properties, the literature reports several established practices. A widely adopted approach is to report LCR and PGR alongside LCRPGR and to interpret the ratio in conjunction with the magnitude and sign of its components (e.g., Estoque et al., 2021; Calka et al., 2022; Cimini et al., 2023). This recommendation is also emphasized in the official SDG Indicator 11.3.1 methodology (UN Statistics Division, 2025). Other studies employ sign-based interpretation tables that explicitly consider whether LCR, PGR, and LCRPGR are positive, zero, or negative, enabling differentiation between population-driven densification, land-intensive expansion, and expansion under population decline (e.g., Calka et al., 2022; Cimini et al., 2023; Jiang et al., 2021; Zhou et al., 2021). Histogram-based analyses are further used to examine the empirical distribution of LCRPGR values, identify extreme or atypical values, and characterize the observed value ranges before aggregation or classification (e.g., Calka et al., 2022). Additional handling strategies include filtering or bounding extreme values for statistical summaries and aggregating results to higher spatial levels to reduce the influence of extreme local outcomes (e.g., Li et al., 2021a; Santillan & Heipke, 2024a).

These studies show that negative values, unbounded magnitudes, and oscillatory behavior are not isolated anomalies but systematic consequences of the ratio-based structure of LCRPGR under certain population change conditions. The interpretation practices reviewed above can improve clarity when LCR and PGR are treated as fixed quantities. However, their usefulness becomes more limited when built-up area and population are themselves uncertain estimates. Under such conditions, even modest uncertainty in the component rates may alter their sign or bring PGR close to zero, thereby amplifying instability in LCRPGR and producing extreme or ambiguous values. These points suggest that deterministic component-wise or rule-based interpretation, while still informative, may not be sufficient on its own when uncertainty in the indicator inputs is explicitly acknowledged.

2.2.3.2 Operational Definitions of Built-up Area in SDG Indicator 11.3.1 Monitoring

Across EO-based applications of SDG Indicator 11.3.1, built-up area is operationalized using a range of data products and thematic definitions that reflect differences in land-cover classification systems, data availability, and analytical objectives. These operational choices directly shape how land consumption is represented within LCR calculations and are documented across a wide body of empirical studies. This variability in built-up area definitions within SDG Indicator 11.3.1 monitoring was explicitly articulated by Calka et al. (2022), who framed such variation as a form of semantic uncertainty arising from differences in how built-up area is conceptually and thematically represented across datasets and implementations.

The official SDG Indicator 11.3.1 metadata defines built-up area as “all areas occupied by buildings” (UN Statistics Division, 2025). However, this building-centric conceptual definition is not always directly aligned with the semantic structure of the built-up or land-cover datasets commonly used in practice, giving rise to what Calka et al. (2022) describe as semantic uncertainty in SDG Indicator 11.3.1 input data. As a result, LCR is frequently calculated using proxy representations that depart from the SDG definition. Examples include artificial areas (Nicolau et al., 2019; Calka et al., 2022), artificial surfaces, including those that also encompass industrial and mining lands and transportation infrastructure (Cai et al., 2020), impervious surfaces (Jiang et al., 2021; Bhandari et al., 2023; Ling et al., 2023), construction land (Wang et al., 2020), and built-up area defined as an aggregation of residential and bare land classes (Jalilov et al., 2021). In some cases, LCR is derived from the fusion of multiple land-cover representations, such as built-up area (SDG definition), impervious surface, and artificial surface classes, drawn from different data products (Zhou et al., 2021).

In contrast to this diversity of operational definitions, studies that rely on the GHSL most closely adhere to the intended definition of built-up area in SDG Indicator 11.3.1. The GHSL conceptualization aligns with the Infrastructure for Spatial Information in Europe (INSPIRE) framework, in which built-up area refers explicitly to roofed constructions intended for human shelter, economic production, or service delivery (Pesaresi et al., 2013; European Commission, 2023; Sikder et al., 2025). This building-centric definition aligns with the SDG metadata formulation and distinguishes GHSL-based implementations from approaches that equate built-up area with broader categories of artificial or impervious surfaces.

Despite these substantive definitional differences, the choice of built-up area representation is rarely articulated as an explicit methodological decision in applications of SDG Indicator 11.3.1. Moreover, indicator results are often discussed without reference to the specific built-up area definition employed. One notable exception is Calka et al. (2022), who explicitly justify their use of “artificial areas” by arguing that the land type addressed by the indicator corresponds to urban land comprising both built-up areas and associated open urban spaces; on this basis, they adopt artificial areas from the Coordination of Information on the Environment (CORINE) land cover dataset. In most studies, however, an implicit assumption of equivalence across alternative definitions of built-up areas persists. This assumption complicates cross-study and cross-context comparisons, as observed differences in LCR and LCRPGR values may reflect variation in the semantic representation of built-up areas rather than substantive differences in LUE outcomes.

2.2.3.3 Treatment of Accuracy and Uncertainty in Input Data

Most implementations of SDG Indicator 11.3.1 incorporate some form of accuracy assessment or validation of built-up area and population inputs during the data preparation stage. However, the manner in which accuracy and uncertainty information is subsequently treated within indicator calculation and interpretation varies across studies and is typically discussed through a combination of methodological descriptions and qualitative interpretive caveats.

Studies that generate built-up area maps through image classification, multi-data integration, or data fusion approaches report accuracy metrics such as overall accuracy, user's and producer's accuracies, or related commonly used measures (e.g., Mudau et al., 2020; Cai et al., 2020; Jalilov et al., 2021; Koroso et al., 2021; Zhou et al., 2021; Han et al., 2022; Karthikeya et al., 2024). In most cases, however, these accuracy assessment results are not propagated into subsequent calculations of LCR, PGR, or LCRPGR, and indicator values are calculated using mapped or classified inputs as fixed quantities. Because classification error is unavoidable in thematic mapping, direct area estimates obtained by counting mapped pixels can be biased (Stehman, 2013). Well-established design-based methods address this problem by using probability samples and an error matrix to derive error-adjusted class-area estimates and associated uncertainties, supported by widely used good-practice guidance (e.g., Stehman, 2009, 2013; Olofsson et al., 2013, 2014; Olofsson, 2021; FAO, 2016). However, these approaches are rarely employed in SDG Indicator 11.3.1 workflows to quantify built-up data uncertainty and propagate it into the indicator components.

A similar pattern is observed in studies that rely on readily available land-use, land-cover, built-up, or population data products, including global EO-derived datasets. In such applications, validation results reported in prior assessments or external documentation are typically cited to justify dataset suitability (e.g., Estoque et al., 2021; Li et al., 2021a; Bhandari et al., 2023; Bounoua et al., 2023; Ling et al., 2023; Santillan & Heipke, 2024a). In some cases, dataset selection is explicitly justified by reference to meeting commonly cited minimum accuracy thresholds (e.g., around 85%), as in the analysis conducted by Yiannakou & Zografos (2025). While a limited number of studies explicitly acknowledge data limitations and sources of uncertainty in their input datasets, typically through qualitative caveats or recommendations for cautious interpretation (e.g., Estoque et al., 2021; Santillan & Heipke, 2024a), such information is not explicitly incorporated into LUE calculations.

Calka et al. (2022) provide an explicit discussion of input data uncertainty, emphasizing that uncertainty is an inherent attribute of the spatial datasets used in SDG Indicator 11.3.1 monitoring. Their analysis identifies multiple sources of uncertainty, including semantic ambiguity in land-cover definitions, limitations in thematic accuracy, commission and omission errors, spatial generalization effects, and positional inaccuracies in land-cover datasets. In their LUE assessment, they report that LCRPGR outliers are mainly due to uncertainty in the source data rather than substantive urban processes. At the same time, they observe that such extreme values constitute a small fraction of all observations and are spatially dispersed, and, therefore, do not undermine the overall validity of aggregated indicator results. While these uncertainty sources and their effects are discussed in detail, they are not formally propagated into the calculation of LCR, PGR, or LCRPGR, and the indicator is ultimately evaluated within a deterministic framework.

As reflected in the literature, SDG Indicator 11.3.1 workflows predominantly rely on fixed representations of built-up area and population, with uncertainty addressed primarily through validation reporting and qualitative caveats rather than formal integration into indicator calculation. This approach enables broad applicability but limits insight into how data uncertainty, combined with semantic ambiguity, interacts with the ratio-based structure of LCRPGR, particularly when indicator values are derived from heterogeneous input sources.

2.2.3.4 Emerging Sensitivity and Robustness Analyses

Building on established EO-based implementation practices, a recent strand of literature has begun to explicitly examine the sensitivity of SDG Indicator 11.3.1 outcomes to controlled variations in input data characteristics, spatial resolution, and population redistribution methods. Rather than treating anomalous or extreme values as isolated artifacts, these studies adopt experimental and comparative designs that systematically perturb inputs or workflows to assess the robustness and stability of indicator behavior under alternative data and methodological configurations. Against a backdrop of largely standardized EO-based workflows, these analyses directly challenge the implicit assumption that built-up area and population datasets are interchangeable within SDG Indicator 11.3.1 implementations. Cardenas-Ritzert et al. (2024b), for example, isolate spatial resolution effects by varying the granularity of land-use inputs while holding other components constant. Their results show that coarser spatial representations tend to omit small or fragmented urban features, alter the detection of urban boundaries, and bias estimated land consumption rates, thereby affecting downstream LCR and LCRPGR values. Related work has sought to reduce such sensitivities by improving the spatial precision of built-up area inputs. Wang et al. (2025), for instance, demonstrate that high-resolution, high-accuracy urban extraction workflows can attenuate extreme LCRPGR values and yield smoother spatial and temporal patterns, suggesting that part of the observed instability arises from the quality of input delineation.

Some complementary sensitivity analyses focus on population inputs rather than land consumption. For instance, Zhong et al. (2025a) examine the robustness of LCRPGR to alternative global gridded population datasets while holding built-up area inputs constant, showing that differences in population redistribution alone can produce substantial divergence in LCRPGR estimates. Their findings indicate that uncertainty introduced during population gridding is not merely additive but can be strongly amplified through the logarithmic calculation of PGR and the ratio-based formulation of LCRPGR. Extending this line of inquiry, Zhong et al. (2025b) demonstrate through controlled perturbation experiments that negligible population change and even minor errors in population data can trigger extreme indicator values, frequent sign reversals, and sharp inter-interval oscillations, resulting in unstable temporal trajectories despite relatively smooth land consumption trends.

Collectively, these studies demonstrate that the limitations of SDG Indicator 11.3.1 arise not only from data quality, resolution, or semantic choices but also from the nonlinear interaction between uncertain inputs and the indicator’s ratio-based mathematical structure. Although these recent sensitivity and robustness analyses provide important diagnostic insight, they remain grounded in

deterministic workflows that address LCRPGR instability primarily through data selection, resolution adjustment, or accuracy improvement. Residual errors, uncertainty ranges, and their nonlinear propagation through LCR and PGR are not explicitly quantified, leaving a gap in the development of SDG Indicator 11.3.1–aligned frameworks that can characterize indicator behavior probabilistically and support uncertainty-aware interpretation.

2.2.3.5 Dimensional and Spatial Scope of Land Consumption Representation

Beyond data quality and implementation choices, the literature identifies structural limitations in SDG Indicator 11.3.1 stemming from its reliance on two-dimensional representations of land consumption. By construction, the indicator captures horizontal expansion of built-up area but does not explicitly represent vertical development processes, thereby constraining its ability to reflect the full dimensionality of contemporary urban growth (Estoque et al., 2021; Huang et al., 2024).

Empirical studies show that cities following markedly different development pathways, such as compact vertical intensification versus extensive horizontal expansion, can exhibit similar LCRPGR values (Estoque et al., 2021). This limitation has been reported across diverse geographical contexts, demonstrating that even when LCR and PGR are interpreted jointly, the indicator remains insufficient to distinguish among structurally distinct urban growth regimes (Schiavina et al., 2019; Estoque et al., 2021). Evidence from global analyses further suggests that weak or unstable relationships between land consumption and population growth in many cities are increasingly driven by vertical construction processes that are not captured by area-based metrics alone (Estoque et al., 2021).

Closely related to this dimensional constraint is the lack of spatial explicitness inherent in the LCRPGR formulation. As a dimensionless ratio aggregated at the urban unit level, the indicator provides no information on the spatial configuration of land consumption or population change. Consequently, it cannot differentiate between key spatial urbanization processes such as infill development, edge expansion, leapfrogging, or settlement dispersion (Melchiorri et al., 2019; Han et al., 2022; Cimini et al., 2023). This aggregation further contributes to interpretive ambiguity, as identical LCRPGR values may arise from fundamentally different spatial patterns and land-use trajectories. Moreover, path dependency effects imply that similar efficiency values can result from different historical sequences of land take and population change, even when the underlying urban development trajectories are substantially different (Schiavina et al., 2019).

As a result, the literature emphasizes that LCRPGR captures only a limited subset of urban development dynamics. By focusing exclusively on horizontal land consumption and population change, the indicator does not account for vertical growth, intra-urban heterogeneity, or the broader economic, institutional, and environmental contexts that shape LUE. Consequently, efficiency classifications derived from SDG Indicator 11.3.1 may obscure important structural and functional differences among urban areas, limiting their explanatory power and policy relevance when applied in isolation.

2.2.4 Conceptual and Methodological Extensions of LCRPGR

In response to the documented methodological characteristics and limitations of SDG Indicator 11.3.1, the literature has proposed a range of conceptual extensions and complementary analytical approaches to enhance the interpretation of LUE outcomes. These efforts seek to mitigate known interpretive ambiguities, sensitivity to input data, and structural blind spots of LCRPGR, while largely retaining its role within the SDG monitoring framework.

A common strategy involves contextualizing LCRPGR by using supplementary indicators that more explicitly characterize specific dimensions of urban expansion. Metrics such as Built-up Area per Capita (BUpC) and Total Change in Built-up Area are formally included in the official SDG Indicator 11.3.1 methodology as secondary indicators intended to support interpretation of LCRPGR values and clarify the underlying nature of urban growth within each spatial unit (UN Statistics Division, 2025). Additional metrics, including Abstract Achieved Population Density in Expansion Areas (AAPDEA) and Marginal Land Consumption per New Inhabitant (MLCNI), have been proposed to provide further insight into the intensity, efficiency, and spatial manifestation of land consumption (Schiavina et al., 2019, 2022b). These indicators are typically reported alongside LCRPGR to aid interpretation, particularly in distinguishing land-intensive expansion from population-driven densification in newly urbanizing areas. Other studies extend interpretation by linking LCRPGR outcomes to broader socioeconomic or institutional dimensions. For example, LUE has been examined in relation to economic productivity through indicators such as the ratio of economic growth to land consumption (EGRLCR) (Jiang et al., 2021) or contextualized using governance-related variables such as the World Governance Indicators (Jaber, 2025). Similarly, Huang et al. (2024) combine urban expansion intensity (UEI), compactness index (CI), and BUpC with LCRPGR value ranges to construct a composite LUE evaluation system that captures internal urban structural change. Collectively, these approaches broaden the analytical framing of SDG Indicator 11.3.1; however, they remain auxiliary in nature and do not alter the indicator's underlying mathematical structure or its reliance on built-up area and population change rates.

Building on earlier evidence that LCRPGR is highly sensitive to input data uncertainty and exhibits nonlinear error amplification (Section 2.2.3), several authors have questioned whether the indicator should continue to serve as the primary summary metric for LUE assessment. Rather than explicitly reformulating the indicator or modeling uncertainty, these studies propose alternative strategies to improve interpretability and stability within a deterministic framework. Zhong et al. (2025b), for example, respond to LCRPGR instability by advocating mitigation through indicator substitution and spatial aggregation. They emphasize increased reliance on BUpC and its associated derivatives (e.g., absolute values, growth rates, and increments), demonstrating that BUpC-based measures are generally less sensitive to incidental data errors and yield smoother, more stable temporal patterns than LCRPGR in comparative analyses. This approach improves interpretability by emphasizing per-capita representations of land consumption. However, it does not alter the underlying land-population relationship and addresses uncertainty only indirectly through indicator choice and aggregation.

A more fundamental response is proposed by Angel et al. (2024), who argue that the ratio-based formulation of SDG Indicator 11.3.1 should be replaced altogether. Motivated by the unbounded,

sign-ambiguous, and difficult-to-interpret behavior of LCRPGR, they propose two alternative rate-based indicators: the average annual rate of change in land consumption per person and the average annual rate of change in population density. These indicators are derived from the same built-up area and population components as LCRPGR. They are expressed as differences rather than ratios, yielding bounded, sign-consistent values that are easier to interpret and compare across cities and regions. However, despite improving mathematical transparency, this reformulation likewise remains deterministic, treating built-up area and population inputs as fixed quantities and leaving their uncertainties unquantified and unpropagated.

Overall, these conceptual and methodological extensions reflect growing recognition that LCRPGR, despite its operational simplicity and widespread adoption, provides an incomplete and, in some contexts, fragile representation of urban LUE. Existing efforts have largely focused on supplementing or substituting the indicator through alternative metrics or contextual variables, rather than directly addressing deeper issues related to uncertainty propagation, nonlinear error amplification, and the incomplete representation of multidimensional urban growth processes inherent in the original formulation. Consequently, while these approaches enhance interpretability and analytical breadth, important methodological challenges remain unresolved within the current SDG Indicator 11.3.1 monitoring framework.

2.2.5 Extensions of SDG Indicator 11.3.1 to LUE Forecasting

While SDG Indicator 11.3.1 is conceptually aligned with long-term sustainability objectives, the literature reviewed in the previous sections primarily focuses on retrospective assessments of LUE using historical built-up area and population data. These studies have provided valuable insights into long-term urbanization trends and spatial heterogeneity in LUE but offer limited guidance for anticipatory planning and policy formulation. Such analyses deepen understanding of how urbanization has evolved and affected efficiency; however, they provide little insight into how these dynamics may unfold in the future.

A small but growing body of literature has begun to extend SDG Indicator 11.3.1 into a forward-looking context by projecting future built-up area and population and computing prospective LCR, PGR, and LCRPGR values. These efforts generally adopt scenario-based or model-driven projections. Most rely on predefined socioeconomic or policy scenarios, most notably the Shared Socioeconomic Pathways (SSPs), a set of globally consistent narratives and quantitative projections describing alternative future trajectories of population, economic development, urbanization, and technological change used to support long-term sustainability and climate-related assessments (O'Neill et al., 2017).

One prominent line of work employs SSP-consistent statistical downscaling frameworks to generate future built-up area and population inputs for SDG Indicator 11.3.1 assessment. Jiang et al. (2022) project city-level urban land and population under multiple SSPs by first estimating provincial urban land demand using regression models that link per-capita built-up area to cumulative Gross Domestic Product (GDP), per-capita GDP, and urbanization rate. These province-level projections are then downscaled to cities, assuming that city-level growth follows provincial urban

population growth trajectories. Future LCRPGR values are subsequently calculated from these projected inputs to evaluate urban LUE across SSPs. In this framework, forecasting is explicitly tied to socioeconomic scenario narratives, with the primary purpose of assessing the sustainability implications of alternative development pathways rather than predicting city-specific trajectories from historical EO observations.

A related but conceptually distinct approach is adopted by Wang et al. (2024), who rely directly on existing gridded SSP projection datasets for population, GDP, and land-use change to calculate future values of SDG Indicator 11.3.1 over multi-decadal horizons. Rather than explicitly modeling urban expansion processes or demographic change, this approach treats SSP-consistent projections as exogenous inputs and focuses on evaluating the temporal evolution and spatial differentiation of LCRPGR across alternative development scenarios. In this case, SDG Indicator 11.3.1 is used as an evaluative metric within a broader scenario analysis framework, including coupling analyses with other efficiency concepts, rather than as an object of forecasting.

In contrast to SSP-driven statistical projections, several studies adopt process-based land-use simulation models to generate future built-up area and then combine these outputs with projected or externally sourced population data. Lu et al. (2022), for example, apply a cellular automata (CA)–Markov modeling framework to simulate future land-use patterns under policy and environmental constraints, using historical land-cover transitions to estimate future built-up expansion. Population inputs are obtained from gridded population projection datasets, and SDG Indicator 11.3.1 values are calculated for future time horizons by combining simulated land expansion with projected population change. This approach emphasizes spatial explicitness and local land-use dynamics. However, it implicitly assumes that past transition rules and suitability relationships remain valid in the future and that population projections are sufficiently reliable to support indicator calculation.

An even more planning-oriented extension is presented by Zhang et al. (2026), who integrate multi-objective land-use optimization, genetic algorithms, and spatial simulation models to generate alternative future land-use scenarios under explicit policy and sustainability constraints. Built-up area trajectories are produced under multiple development scenarios, while population is projected using logistic growth models with scenario-specific carrying capacities. SDG Indicator 11.3.1 and related indicators are then used to evaluate whether planned development pathways align with sustainability objectives. In their framework, forecasting serves primarily as a decision-support and planning evaluation tool, with explicit assumptions embedded in scenario design, growth limits, and policy targets.

Collectively, these studies demonstrate that extending SDG Indicator 11.3.1 to future-oriented analysis is technically feasible and increasingly applied. However, they also reveal several common characteristics and limitations. First, most forecasting efforts rely on scenario-based projections or process-driven simulations. Second, projected built-up area and population inputs are typically treated as deterministic realizations of scenarios or models, with uncertainty either implicitly embedded in scenario narratives or addressed qualitatively. Third, uncertainty associated with projected inputs is not formally propagated through the calculation of LCR, PGR, or LCRPGR, despite the indicator’s known sensitivity to small variations in its components. As a result, existing forecasting

extensions provide valuable exploratory and planning-oriented insights but do not yet offer a probabilistic, uncertainty-aware framework for forward-looking SDG Indicator 11.3.1 monitoring. This gap motivates the development of forecasting approaches that explicitly model uncertainty in both built-up area and population data and examine how this uncertainty propagates probabilistically through LUE indicators.

2.3 Synthesis and Summary of Research Gaps

The literature reviewed in the previous sections demonstrates that SDG Indicator 11.3.1 is operationally mature and widely applied across global, national, and local contexts. Supplementary indicators, component-wise interpretation, and standardized workflows have improved its usability and interpretability, while emerging forecasting efforts have begun to extend the indicator toward prospective analysis. Nevertheless, several fundamental research gaps remain unresolved:

- **Continued reliance on a deterministic treatment of inherently uncertain inputs.**

Built-up area and population data are typically treated as precise and semantically consistent quantities, despite well-documented differences in classification accuracy, thematic definition, and data quality across EO products and population datasets. Although these uncertainties are frequently acknowledged and, in some cases, assessed or mitigated, they are rarely propagated through the calculation of LCR, PGR, or LCRPGR. Consequently, it remains unclear how differences in built-up area definitions and uncertainties in built-up area and population estimates influence indicator values, their stability, and their interpretation. Systematic analyses quantifying the sensitivity of SDG Indicator 11.3.1 to input uncertainty and semantic variability remain limited.

At the same time, standard remote sensing methods for class area estimation and uncertainty quantification from land cover maps are well established. However, they have not been consistently translated into SDG Indicator 11.3.1 monitoring practice as an integrated, end-to-end pipeline. In other words, the methodological building blocks are in place. Yet, procedures that systematically connect EO-derived built-up inputs to uncertainty-aware BUA estimates and then propagate these uncertainties through to LCR, PGR, and LCRPGR are still rarely implemented in routine reporting. This limitation constrains the robustness, interpretability, and comparability of reported LUE outcomes across data sources and study contexts.

- **Limited development of uncertainty-aware LUE forecasting.** Existing forward-looking studies predominantly rely on scenario-based assumptions or deterministic projections of built-up area and population, from which future LCR, PGR, and LCRPGR are subsequently calculated. In most cases, however, forecast uncertainty and dependence structures are not explicitly modeled. These approaches do not fully exploit the growing availability of long-term EO-derived urban time series, nor do they provide probabilistic projections of LCR, PGR, or LCRPGR. As a result, current forecasting efforts are unable to support uncertainty-aware interpretation, risk-sensitive planning, or probabilistic policy assessment, leaving a gap between retrospective monitoring and forward-looking decision support.

- **Omission of vertical urban growth.** Existing implementations and extensions of SDG Indicator 11.3.1 are grounded in a two-dimensional representation of land consumption that implicitly equates urban growth with horizontal expansion. Vertical development processes, such as densification through increased building height or volume, are therefore invisible to the indicator, despite their growing importance in many high-density and land-constrained urban contexts. Because land consumption is measured only in two dimensions, LCRPGR collapses diverse urban development pathways into a single scalar value, thereby obscuring vertical growth dynamics and limiting the indicator's capacity to explain how cities accommodate population change. As a result, LCRPGR alone cannot fully distinguish whether apparent compactness arises from horizontal containment, vertical intensification, or a combination of both. Consequently, the indicator offers an incomplete representation of the multidimensional nature of contemporary urban development, providing only limited insight into the processes through which LUE is achieved and the structural sustainability of observed urban development trajectories.

Addressing these gaps requires a set of analytical approaches that extend SDG Indicator 11.3.1 beyond deterministic input treatment and purely planar interpretation. Such approaches must explicitly consider the sensitivity of the indicator to variations in built-up area definitions, data-source selection, and uncertainty in built-up area and population estimates, while also incorporating probabilistic forecasting and complementary representations of vertical urban growth. The following chapters develop and evaluate these extensions in a structured manner within the SDG Indicator 11.3.1 framework. As an immediate next step, Chapter 3 conducts a sensitivity analysis to examine how alternative built-up area and population inputs affect LCR, PGR, LCRPGR, and the resulting LUE interpretation.

3 Sensitivity of SDG Indicator 11.3.1 to Alternative Representations of Built-up Area and Population²

3.1 Chapter Overview

This chapter examines the sensitivity of SDG Indicator 11.3.1 to alternative representations of built-up area (BUA) and population (Pop) within a strictly controlled, deterministic analytical setting. The objective is to assess how plausible and commonly used data choices influence calculated values of LCR, PGR, and LCRPGR when all other elements of the analysis are held constant.

The chapter is specifically directed at answering the first research question (RQ1): *How sensitive are SDG Indicator 11.3.1 values and LUE interpretations to uncertainty in built-up area and population arising from measurement, definition, and data-source selection under retrospective conditions?*

The analysis is intentionally diagnostic and is not meant to explain urban growth processes, identify causal relationships, or assess long-term trends. Its purpose is to establish a consistent empirical baseline that enables the isolation and interpretation of the effects of input data selection. The EO-derived BUA and Pop datasets are retained at their native spatial resolutions rather than resampled to a common grid. In this way, differences in spatial resolution are treated as part of the real-world variability associated with dataset choice.

This chapter serves as a foundation for the uncertainty-aware analyses developed in the subsequent chapters. By characterizing the magnitude, structure, and direction of data-driven variability in SDG Indicator 11.3.1 outcomes under retrospective conditions, it provides the empirical basis for the uncertainty propagation, probabilistic LUE classification, and forecast-based assessments that follow.

²This chapter builds upon the author's published work: Santillan et al. (2025a). *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. <https://doi.org/10.5194/isprs-annals-X-G-2025-753-2025>. The deterministic input data sensitivity analysis presented here has been substantially extended and revised. Compared to the published article, this chapter incorporates additional built-up area and population datasets, revised data preparation procedures, and an explicitly controlled one-factor-at-a-time sensitivity design. The scope of the analysis has been expanded to provide a systematic, nationwide assessment at the city and municipality level, and the interpretation of sensitivity results has been refined to serve as an empirical foundation for the uncertainty-aware and probabilistic analyses developed in subsequent chapters.

3.2 Sensitivity Analysis Methodology

The sensitivity analysis is implemented by repeatedly calculating LCR, PGR, and LCRPGR using alternative representations of BUA and Pop. For each calculation, the same spatial reporting units, temporal interval, and indicator formulations are applied to ensure direct comparability across results. The indicator values are calculated following the standard SDG Indicator 11.3.1 formulations described in Section 2.1.1, using a fixed temporal duration Δt . All inputs are treated as deterministic quantities, consistent with prevailing SDG Indicator 11.3.1 monitoring practice.

To ensure unambiguous attribution of observed differences to a single source of variation, sensitivity is examined using a one-factor-at-a-time (OFAT) design. In the built-up sensitivity experiments, alternative built-up datasets are introduced while population inputs are held fixed to a single baseline. In the population sensitivity experiments, alternative population datasets are introduced while the built-up representation is held constant. This structure is particularly appropriate for SDG Indicator 11.3.1 because LCR and PGR depend on different denominators, BUA_{t_1} and Pop_{t_1} , respectively. Additionally, simultaneous substitutions can obscure whether changes in LCRPGR originate from built-up change, population change, or their interaction. Under OFAT, changes in indicator values can be interpreted directly as deterministic sensitivity responses to the varying input component.

Indicator values are calculated independently for each spatial unit and dataset configuration. For each configuration, LUE class labels are also assigned using the classification scheme defined in Table 2.1. The resulting distributions of LCR, PGR, LCRPGR, and LUE classes are then compared across configurations to evaluate the magnitude, direction, and spatial variability of dataset-driven sensitivity.

3.3 Experimental Setup and Implementation Details

3.3.1 Study Area

The study area was the Philippines (Figure 3.1a), an archipelagic country in Southeast Asia with a land area of about 300,000 km², including inland water bodies (Fabian Jr., 1991). It comprises 7,641 islands grouped into Luzon, Visayas, and Mindanao (Licuanan et al., 2019). This fragmented geography contributes to substantial heterogeneity in settlement patterns, population distribution, and urban development intensity across the country (Berse, 2024; Santillan & Heipke, 2024a).

Administratively, the Philippines has a four-tier structure of regions, provinces, cities, municipalities, and barangays. Regions mainly serve planning and coordination functions, while the lower three levels are generally referred to as local government units. Among these, cities and municipalities are the most relevant units for analyzing SDG Indicator 11.3.1 because they are the primary level for socio-economic planning, land-use regulation, and urban development. As of 2020, the country had 17 regions, 82 provinces, 149 cities, and 1,485 municipalities. Of the cities, 38 were independent, and 33 were highly urbanized cities, 16 of which were in the National Capital Region (Metro Manila) (PSA, 2021). Census data show sustained population growth in the Philippines

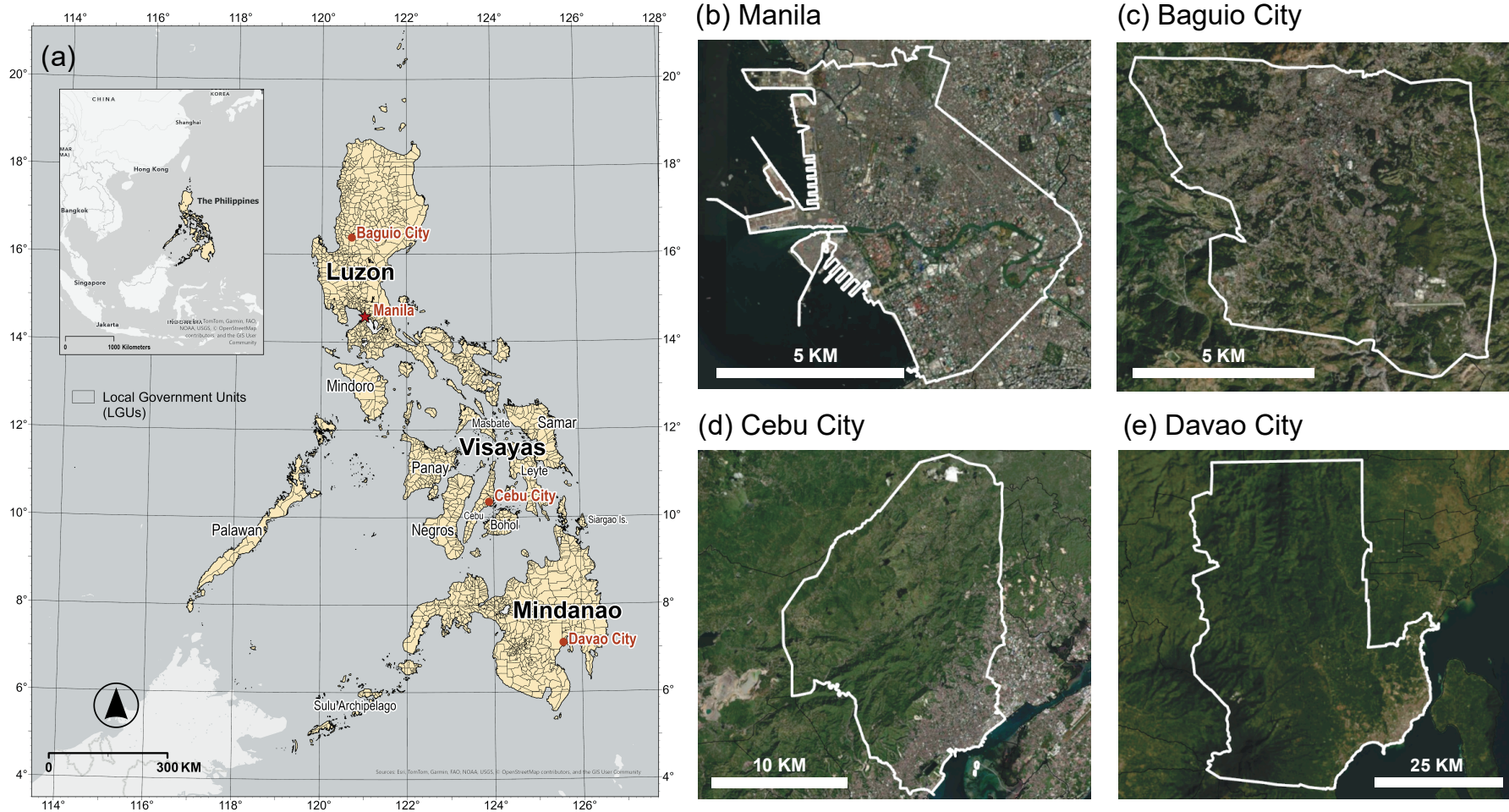


Figure 3.1: Study area and reference cities used to aid the interpretation of the nationwide sensitivity results. (a) Geographic context of the Philippines, showing the three major island groups (Luzon, Visayas, and Mindanao) and the locations of the four selected urban centers; the inset situates the Philippines within East and Southeast Asia. (b)–(e) Satellite image views of the selected cities, with administrative boundaries shown as white outlines, included to provide visual context on differences in urban form and surrounding landscape (Imagery Source: Esri, Vantor, Earthstar Geographics, and the GIS User Community).

up to 2020, with the total population reaching about 109 million. At the same time, the annual population growth rate gradually declined, from 1.90% in 2000–2010 to 1.72% in 2010–2015 and 1.63% in 2015–2020 (PSA, 2017, 2021).

The country’s fragmented geography, uneven settlement structure, and continuing demographic change make the Philippines a suitable setting for examining the sensitivity of LUE metrics to alternative representations of BUA and Pop. Earlier national-scale LUE assessments in the country have mainly relied on GHSL as the source of BUA and Pop data (Santillan & Heipke, 2023, 2024a, 2024b). Although these studies provide useful baseline evidence on built-up land consumption and population change, they offer limited insight into how alternative EO-derived datasets influence SDG Indicator 11.3.1 results under otherwise controlled conditions. Building on this earlier work, the present study uses the Philippines as a testbed to systematically assess dataset-driven sensitivity in LUE outcomes rather than to re-estimate national urbanization trends.

3.3.2 Spatial Units of Analysis, Temporal Scope, and Analytical Interval

The sensitivity analysis was conducted nationwide, using all administratively defined Philippine cities and municipalities (1,634 in total) as the spatial units of analysis. Hereafter, these city and municipality units are collectively referred to as local government units (LGUs). To aid interpretation of nationwide results, selected cities, namely Manila (the nation’s capital), Baguio City, Cebu City, and Davao City, are referenced as illustrative examples (Figure 3.1b–e).

The sensitivity analysis used the 2000–2015 interval ($\Delta t=15$ years) as a single cumulative period to emphasize differences arising from dataset choice. A longer interval was considered preferable in this context because it captures cumulative built-up change and reduces the influence of short-term fluctuations and small absolute changes in BUA or Pop that can obscure dataset-driven effects. The selection of 2000–2015 was also guided by data availability and consistency. Both endpoints have census-derived Pop data, whereas more recent years, such as 2020, lack temporally consistent and harmonized coverage across the several BUA and Pop datasets used in this chapter.

3.3.3 Datasets Used

BUA was represented using six global datasets: GHS-BUILT-S, GISD30, GISA 3.0, GAUD, WSF-Evolution, and the Global Land Analysis and Discovery (GLAD) Global Land Cover and Land Use Change (GLCLUC) dataset (Figure 3.2). Their main characteristics and BUA definitions are summarized in Table 3.1. Except for GHS-BUILT-S, which combines multitemporal Landsat and Sentinel-2 imagery, all datasets were derived primarily from multitemporal Landsat observations. Reported validation summaries indicate generally high but non-uniform accuracies across these BUA datasets (Table 3.2).

The BUA datasets were selected because they are widely used in urban and SDG Indicator 11.3.1 studies but differ substantially in their spatial resolution, temporal structure, classification logic, and underlying interpretation of built-up land, spanning both conservative and more inclusive representations. Most were designed specifically for mapping built-up or settlement areas. In

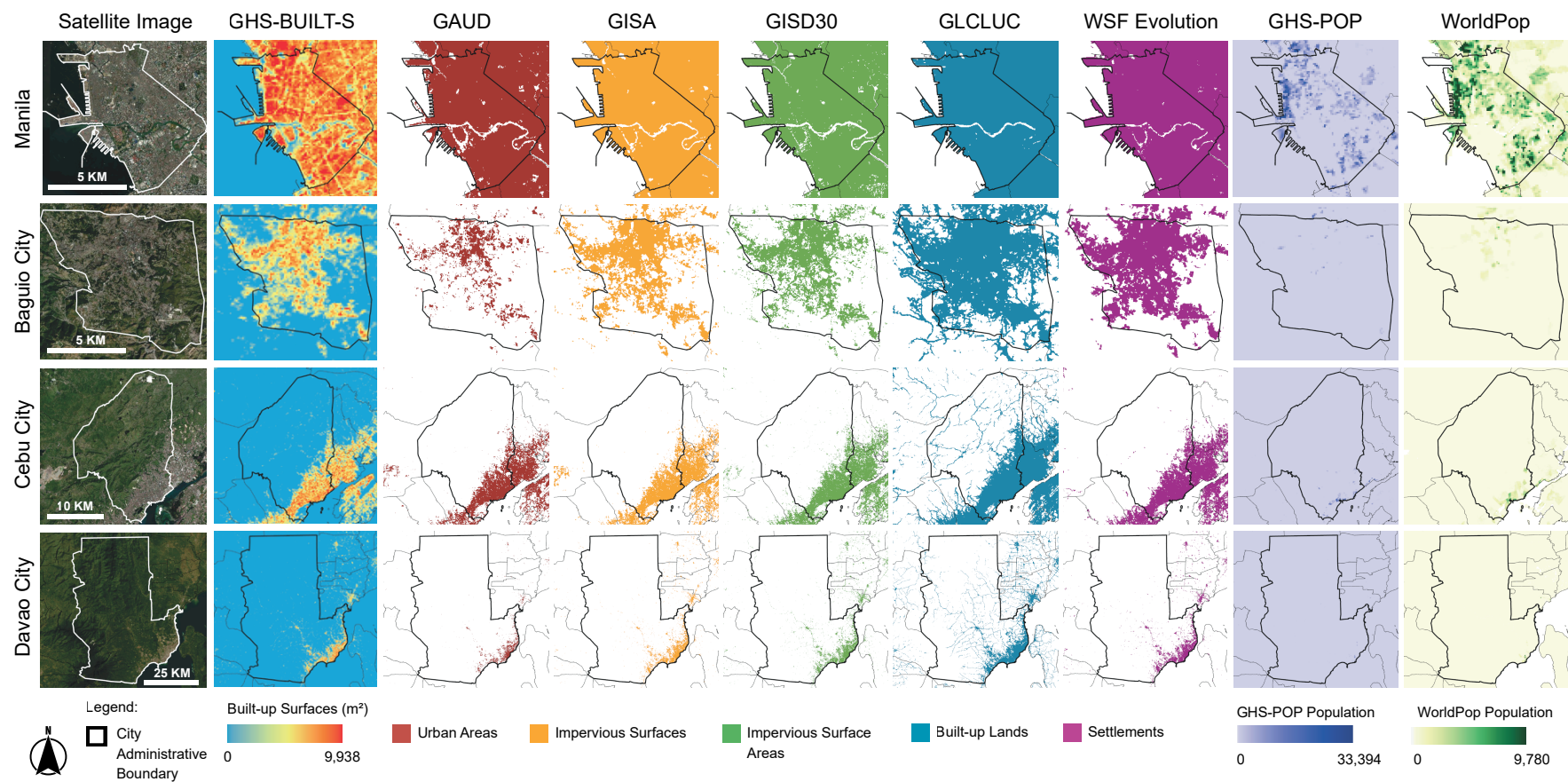


Figure 3.2: Built-up area and population representations for four representative Philippine cities in 2015, highlighting dataset-driven differences in mapped extent and spatial pattern. Basemap imagery in the first column is shown for visual reference only (Imagery Source: Esri, Vantor, Earthstar Geographics, and the GIS User Community).

Table 3.1: Major characteristics of the global built-up area datasets included in the sensitivity analysis. GHS-BUILT-S is a continuous, area-based built-up dataset at 100 m spatial resolution, in which each grid cell stores the estimated built-up surface area. The other datasets are discrete 1 arc-second products, approximately 30 m at the equator, where each pixel is assigned a single category. These comprise year-of-detection datasets (GISD30, GAUD, and WSF-Evolution) and one year-specific, class-based land-cover dataset (GLCLUC).

Dataset	Temporal Coverage	Minimum Temporal Resolution (years)	‘Built-up’ Name	‘Built-up’ Definition
Global Human Settlement – BUILT-S (GHS-BUILT-S)	1975–2020	5	Built-up surface	“The built-up surface is the gross surface excluding the thickness of the walls beyond the building wall perimeter with a spatial generalization matching the 1 ha topographic map specifications, that is also informally called ‘building footprint’. Building is ‘any roofed structure erected above ground for any use’” (European Commission, 2023).
Global Annual Urban Dynamics (GAUD)	1985–2015	1	Urban areas	Urban areas are “pixels that are dominated by built elements (for example, buildings, roads, runways and so on)” (Liu et al., 2020b).
Global 30 m Impervious Surface Dynamic Dataset (GISD30)	1985–2020	5	Impervious surfaces	Impervious surfaces are “surfaces preventing the surface water from penetrating into the ground” and are composed of anthropogenic materials, such as cement, asphalt, brick, and stone” (Zhang et al., 2022).
Global Impervious Surface Area (GISA)	1972–2019	1	Impervious surface area	“Impervious surface area (ISA) generally refers to artificial structures such as parking lots, roads, roofs, open spaces, etc.” (Huang et al., 2022).
Global Land Analysis and Discovery (GLAD) Global Land Cover and Land Use Change (GLCLUC)	2000–2020	5	Built-up lands	“Built-up land consists of man-made land surfaces associated with infrastructure, commercial and residential land uses”. Built-up land pixels “include man-made surfaces, even if such surfaces do not dominate within the pixel” (Potapov et al., 2022; Hansen et al., 2022).
World Settlement Footprint Evolution (WSF Evolution)	1985–2015	5	Settlements	Extent human settlements; Settlements are “all areas covered by buildings, building lots or roads/paved-surfaces” (Marconcini et al., 2020, 2021).

contrast, GLCLUC is a land-cover and land-use classification product; it was included to reflect the common practice of deriving BUA indirectly from land-cover or land-use maps. Although GLCLUC is distributed as a single product, its thematic classes were produced using separately calibrated machine learning models, which were subsequently merged into a unified land-cover/land-use dataset. In particular, the built-up class (“built-up lands”) was mapped using a deep learning convolutional neural network trained with OpenStreetMap data (Potapov et al., 2022). The datasets also differ in output type. GHS-BUILT-S is a continuous area-based layer reporting the estimated BUA per grid cell. GISD30, GISA, GAUD, and WSF-Evolution are discrete-event-based products that record the first year of built-up detection, and GLCLUC is a categorical land-cover and land-use dataset from which the built-up class is extracted. In terms of spatial properties, all datasets except GHS-BUILT-S are provided at approximately 1 arc-second resolution in a geographic World Geodetic System 1984 (WGS84) coordinate reference system (CRS), whereas GHS-BUILT-S is available at multiple resolutions and projections; this study used the native 100 m Mollweide version because the 3 arc-second WGS84 product is a reprojected form of that layer, and using the native equal-area product avoids precision loss and grid-warping effects (European Commission, 2023).

Table 3.2: Reported accuracies of the BUA and Pop datasets used in the sensitivity analysis. MAE = mean absolute error; UA = user’s accuracy; PA = producer’s accuracy. Unless otherwise stated, UA and PA refer to the built-up class represented by the dataset. TAA = total allocation accuracy, a summary metric expressing the agreement between the estimated and reference population across the spatial units used in the validation. The values reported here are intended only as summary indicators of dataset performance and are reproduced from the corresponding dataset documentation or validation studies as presented in the original sources. They do not necessarily correspond to the study area, the validation design, or the specific years used in this analysis. The cited references should therefore be consulted for the full validation context, exact metric definitions, and proper interpretation of the reported values.

Dataset	Reported Accuracy
Global Human Settlement – BUILT-S (GHS-BUILT-S)	MAE = 6% of grid cell area (100 × 100 m) (Pesaresi et al., 2024)
Global Annual Urban Dynamics (GAUD)	UA = 86.9%, PA = 88.1% (Ren et al., 2025)
Global 30 m Impervious Surface Dynamic Dataset (GISD30)	UA = 77.9–93.5%, PA = 67.1–92.3% (Zhang et al., 2022)
Global Impervious Surface Area (GISA)	UA = 86.1%, PA = 92.5% (Ren et al., 2025)
Global Land Analysis and Discovery (GLAD) Global Land Cover and Land Use Change (GLCLUC)	Stable built-up lands (2000–2020): PA = 39.1%, UA = 63.7%. Built-up land gains (2000–2020): PA = 59.6%, UA = 74.1% (Potapov et al., 2022)
World Settlement Footprint Evolution (WSF Evolution)	UA = 84.5%, PA = 92.8% (Ren et al., 2025)
GHS-POP	TAA: 83% (Pesaresi et al., 2024)
WorldPop	TAA: 72% (Pesaresi et al., 2024)

Pop inputs comprised official census-based counts, GHS-POP, and WorldPop; the reported accuracies of the gridded population datasets are summarized in Table 3.2. Census data from the

Philippine Statistics Authority (PSA) provided administratively reported population totals for May 1, 2000, and August 1, 2015, and served as the baseline population data in this study, in line with common practice in SDG Indicator 11.3.1 monitoring.

GHS-POP and WorldPop are EO-derived gridded population products that redistribute population counts using different ancillary data and modeling approaches. GHS-POP shares the same spatial resolution and coordinate reference system options as GHS-BUILT-S, and in this study, the 100 m Mollweide version was used for consistency with the built-up area input. Its estimates are derived from the Center for International Earth Science Information Network Gridded Population of the World, version 4.11. They are disaggregated to grid cells using the spatial distribution, density, and classification of built-up areas from GHS-BUILT-S (European Commission, 2023).

WorldPop similarly provides gridded population estimates as the number of people per pixel at 3 arc-second resolution in a geographic WGS84 CRS. The dataset is generated using a Random Forest-based dasymetric redistribution approach that reallocates census population counts to grid cells using a suite of ancillary predictors (Stevens et al., 2015; WorldPop and CIESIN, 2018). For this study, the WorldPop products for 2000 and 2015 correspond to those generated using the top-down unconstrained methodology, in which population is redistributed across all grid cells without applying a settlement or built-up mask (WorldPop, 2026). As a result, population may be allocated to both built-up and non-built-up areas, allowing representation of dispersed rural populations but potentially introducing population presence in areas unlikely to be inhabited. Despite this limitation, the unconstrained WorldPop products provide a useful contrast to settlement-constrained population representations, such as GHS-POP, and enable assessment of the sensitivity of LUE metrics to alternative population modeling assumptions.

All BUA GeoTIFF tiles covering the Philippines for 2000 and 2015 were downloaded from their respective sources³ and mosaicked. City- and municipality-level census data were obtained from the PSA website⁴, while GHS-POP and WorldPop data⁵ were obtained from their respective data repositories. The administrative boundary shapefile was obtained from the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) through the Humanitarian Data Exchange⁶ and was used as the common boundary layer in all analyses.

³BUA datasets were downloaded from the following links: GHS-BUILT-S, <https://human-settlement.emergency.copernicus.eu/download.php> (accessed February 14, 2025); GAUD, https://figshare.com/articles/dataset/A_global_dataset_of_annual_urban_extents_1992-2020_from_harmonized_nighttime_lights/16602224/1 (accessed September 4, 2024); GISD30, <https://zenodo.org/records/5220816> (accessed September 3, 2024); GISA, <https://zenodo.org/records/14848113> (accessed January 26, 2024); GLCLUC, <https://storage.googleapis.com/earthenginepartners-hansen/GLCLU2000-2020/v2/download.html> (accessed January 26, 2026); WSF Evolution, https://download.geoservice.dlr.de/WSF_EVO/ (accessed June 3, 2024).

⁴<https://psa.gov.ph/> (accessed April 19, 2023).

⁵GHS-POP: <https://human-settlement.emergency.copernicus.eu/download.php> (accessed February 14, 2025); WorldPop: <https://hub.worldpop.org/geodata/listing?id=29> (accessed January 26, 2026).

⁶<https://data.humdata.org/dataset/cod-ab-phl> (accessed April 8, 2024).

3.3.4 Data Preparation

To ensure consistent area-based calculations, all spatial datasets were reprojected to a common equal-area CRS before aggregation. A Philippines-specific Albers Equal Area Conic projection was used (standard parallels 7°N and 18°N; central meridian 121.5°E; latitude of origin 0°; false easting and northing 0; datum WGS84). This equal-area CRS preserves areal comparability across the country’s archipelagic extent and reduces latitudinal distortion, so that differences in derived BUA, gridded Pop totals, and indicator values primarily reflect dataset choice rather than projection effects.

The GHS-BUILT-S, GHS-POP, and WorldPop datasets were reprojected to the common Albers equal-area CRS using a 100 m output grid, consistent with their reported nominal spatial resolution. All remaining built-up datasets were reprojected to the same Albers CRS and output on a uniform 30 m grid, reflecting their native spatial resolution.

BUA and Pop totals per LGU were then derived using zonal statistics on the projected layers. For continuous area-based datasets (GHS-BUILT-S) and continuous population counts (GHS-POP, WorldPop), LGU totals were calculated by summing pixel values within each LGU boundary. For discrete event-based year-of-detection built-up datasets (GISD30, GISA, GAUD, and WSF-Evolution), the year-of-detection layer was first thresholded to generate binary “built-up by year” masks for the endpoint years (2000 and 2015). The built-up area per LGU was then calculated by counting built-up pixels within each LGU and converting the counts to area (m²) using the corresponding pixel area. For the discrete class-based GLCLUC dataset, each year-specific land-cover/land-use map was first reclassified into built-up and non-built-up classes, after which built-up area was calculated via pixel counting and converted to area at the LGU level. For population, census totals were available directly at the LGU level and were joined to the LGU boundary layer using a common LGU identifier.

3.3.5 Indicator Calculation and Sensitivity Analysis

Indicator calculation was performed at two spatial levels: the national level, where BUA and Pop were aggregated across all LGUs before indicator calculation, and the LGU level, where the indicators were calculated independently for each city and municipality. Eight deterministic BUA–Pop dataset configurations were implemented for the OFAT sensitivity analysis (Table 3.3). The baseline configuration (BUA0) used GHS-BUILT-S for built-up area and census population totals. BUA sensitivity runs (BUA1–BUA5) substituted alternative built-up datasets while holding Pop fixed to census totals. Pop sensitivity runs (Pop1–Pop2) substituted gridded Pop datasets while holding BUA fixed to GHS-BUILT-S.

For each configuration and spatial level, LCR, PGR, and LCRPGR were calculated for the 2000–2015 interval. In addition, each LGU and national result was assigned a general LUE class (efficient/inefficient) based on the general classification scheme defined in Table 2.1. Sensitivity was summarized as differences in LCR, PGR, LCRPGR, and corresponding LUE class assignments relative to the baseline configuration.

Table 3.3: Deterministic built-up area–population dataset configurations used for one-factor-at-a-time (OFAT) sensitivity analysis (2000–2015). Each configuration calculates LCR, PGR, and LCRPGR for all LGUs and for the national aggregate using identical formulae, spatial units, and the same 2000–2015 interval. The baseline configuration (BUA0) serves as the reference for all difference calculations. Arrows in the “Metrics that vary” column indicate propagation of differences relative to the baseline, for example, built-up substitutions change LCR, which then propagates to LCRPGR; population substitutions change PGR, which then propagates to LCRPGR.

Configuration	Sensitivity block	Built-up area dataset	Population dataset	What varies vs baseline	Metrics that vary
BUA0	Baseline	GHS-BUILT-S	Census	—	—
BUA1	BUA sensitivity	GAUD	Census	BUA only	LCR → LCRPGR
BUA2	BUA sensitivity	GISA	Census	BUA only	LCR → LCRPGR
BUA3	BUA sensitivity	GISD30	Census	BUA only	LCR → LCRPGR
BUA4	BUA sensitivity	GLCLUC	Census	BUA only	LCR → LCRPGR
BUA5	BUA sensitivity	WSF-Evo	Census	BUA only	LCR → LCRPGR
Pop1	Pop sensitivity	GHS-BUILT-S	GHS-POP	Pop only	PGR → LCRPGR
Pop2	Pop sensitivity	GHS-BUILT-S	WorldPop	Pop only	PGR → LCRPGR

3.3.6 Data Completeness and Indicator Calculability Across Configurations

Data completeness screening indicated that indicator calculability is not uniform across dataset configurations because some BUA and Pop products reported zero values for a subset of LGUs, most commonly at the baseline year (2000). Among the built-up alternatives, exclusions were concentrated in GAUD (57 LGUs with zero built-up at t_1 , and 16 at t_2), whereas the other built-up products yielded only small numbers of non-calculable LGUs (2–10 LGUs with zero built-up at t_1 , and 1–9 at t_2 , depending on the product). Pop-related exclusions were rare across both gridded products (1 LGU under GHS-POP and 2 LGUs under WorldPop, with zero population at one or both endpoints). In total, 65 LGUs were non-calculable in at least one configuration due to zero BUA or zero Pop at one or both endpoints. To ensure like-for-like comparisons, all nationwide sensitivity results were therefore calculated on a harmonized intersection sample of $N = 1,569$ LGUs, obtained by excluding any LGU that reported a zero built-up area or zero population in any dataset at either endpoint.

3.4 Results

3.4.1 Dataset Differences in BUA and Pop

Using the harmonized intersection set described in Section 3.3.6, national totals showed substantial cross-dataset differences relative to the baseline configuration (Figure 3.3). In 2000, the baseline dataset, GHS-BUILT-S, reported 1,762 km² of BUA, whereas estimates from the alternative datasets ranged from 1,628 km² for GAUD, corresponding to -7.6% relative to the baseline, to 11,337 km² for GLCLUC, corresponding to $+543.3\%$. GLCLUC therefore reported by far the largest built-up extent among all datasets. The remaining products also exceeded the baseline by wide margins: GISD30 reported 3,549 km² ($+101.41\%$), GISA reported 3,985 km² ($+126.14\%$), and WSF Evolution reported 5,346 km² ($+203.35\%$). By 2015, the divergence had become even larger, with estimates spanning 2,190 km² for GAUD, only -0.69% relative to the baseline value of 2,205 km², to 17,024 km² for GLCLUC ($+671.98\%$). The intermediate 2015 totals also remained far above the baseline, including 4,997 km² for GISD30 ($+126.61\%$), 5,050 km² for GISA ($+128.99\%$), and 6,863 km² for WSF Evolution ($+211.20\%$).

Figure 3.3a further shows that these national-scale differences were also expressed at the LGU level, although the magnitude and structure of divergence varied across localities. In Manila, the baseline BUA total changed only slightly between 2000 and 2015. Most alternative datasets already reported BUA values much higher in 2000. By 2015, the alternative datasets yielded nearly identical BUA estimates. This pattern suggests that the spread in Manila was mainly driven by differences in built-up representation rather than by a large increase in mapped built-up extent from 2000 to 2015. In Baguio City, the cross-dataset spread was large in both years, with GAUD remaining comparatively low while GLCLUC and WSF-Evolution yielded much larger totals, and the mid-range products (GISA and GISD30) fell between these extremes. Cebu City showed a similar ordering, with GLCLUC remaining the largest in both years. Davao City exhibited the strongest divergence among the four cities. All built-up alternatives exceeded the baseline by wide margins. GLCLUC was especially high in both 2000 and 2015.

In contrast, there was less variation in national Pop totals across sources (Figure 3.3). Relative to census totals of 75.736 million in 2000 and 100.067 million in 2015, GHS-POP was nearly aligned, reporting 75.418 million (-0.42%) in 2000 and 100.229 million ($+0.16\%$) in 2015. WorldPop, however, yielded consistently lower totals: 64.399 million in 2000 (-14.97%) and 79.996 million in 2015 (-20.06%). The same general pattern is evident in the four cities. WorldPop was lower than the census in both years for Manila, Baguio City, Cebu City, and Davao City. GHS-POP remained close to the census in most cases, with small city-dependent differences. In Cebu City and Davao City, GHS-POP closely tracked the census in both years. In Baguio City, GHS-POP was lower than the census in both years. Manila showed the most noticeable deviation, where GHS-POP exceeded the census total in 2000 but was slightly lower than the census in 2015.

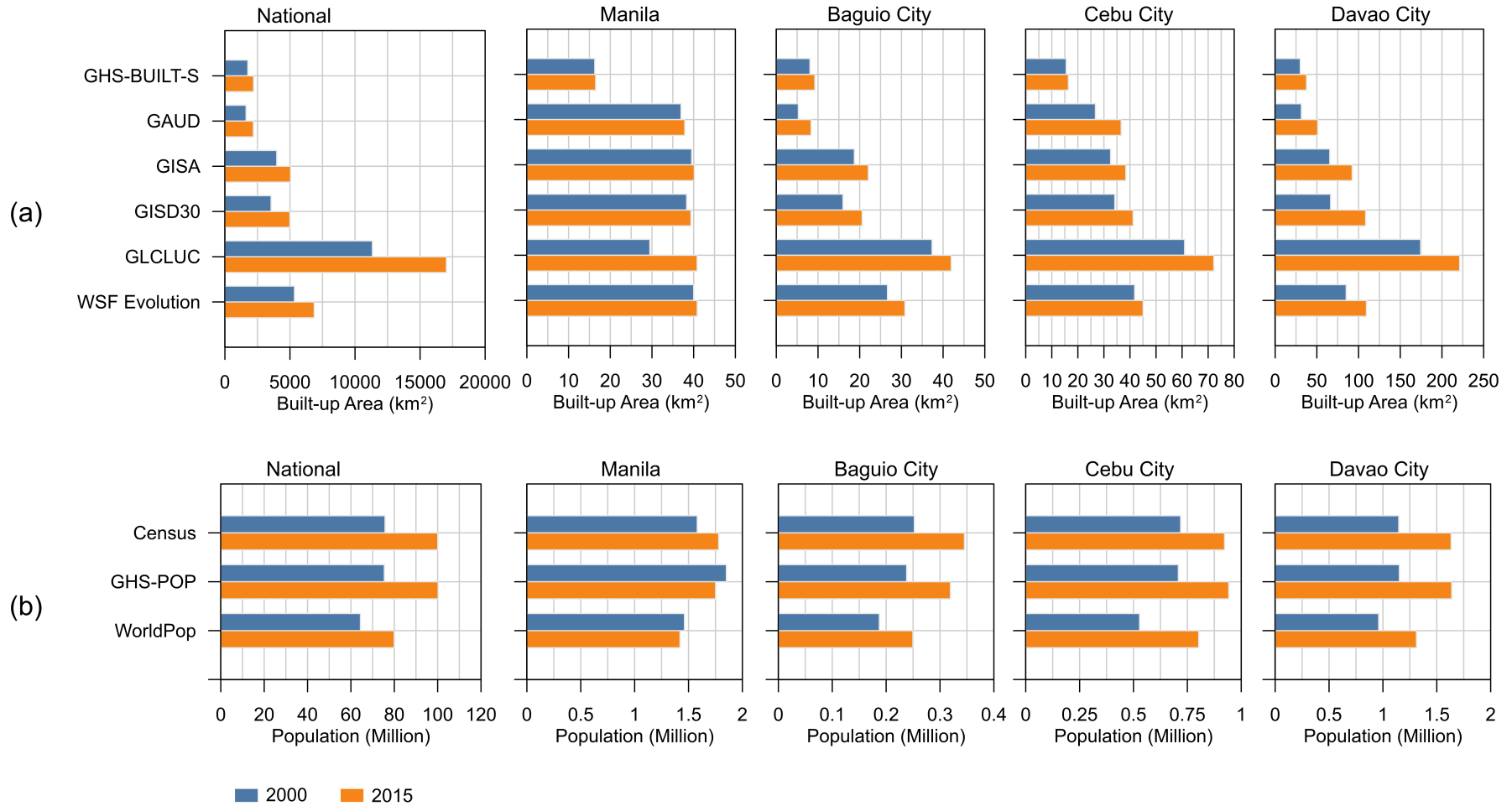


Figure 3.3: Cross-dataset totals of built-up area and population for the Philippines, summarized for the national aggregate and four example cities at the endpoint years 2000 and 2015. Panel (a) shows built-up area totals derived from the baseline (GHS-BUILT-S) and alternative built-up datasets, and panel (b) shows population totals derived from census counts and gridded population products. Note: “National” totals are aggregated over the harmonized intersection set of $N = 1,569$ local government units (LGUs, composed of cities and municipalities) retained for all configurations after excluding any LGU with zero built-up area or zero population in any dataset at either endpoint.

3.4.2 Baseline SDG Indicator 11.3.1 Metrics (GHS-BUILT-S + Census)

Under the baseline configuration (BUA0), national-level SDG Indicator 11.3.1 metrics for the period 2000–2015 yielded LCR and PGR values of 1.68% and 1.86% per year, respectively, resulting in an LCRPGR of 0.90 (Table 3.4). These values placed the Philippines in an efficient LUE category at the national scale, indicating that built-up land expanded more slowly than population over the study period.

At the LGU level, baseline patterns for LCR, PGR, and LCRPGR were spatially heterogeneous (Figure 3.4a–c) and exhibited substantial distributional spread (Figure 3.5). PGR was relatively more concentrated, whereas LCR showed broader dispersion. Figure 3.5 further shows that while the central distributions of LCR and PGR were comparatively compact, LCRPGR was more skewed and included extreme values, particularly in LGUs where PGR was close to zero. The national LCRPGR value (0.90), calculated from totals aggregated across LGUs, differed from the central tendency of the LGU-level distribution (1.26); consequently, the national baseline indicated efficiency even though many LGUs exhibited LCRPGR values above unity (Table 3.4; Figure 3.5).

The corresponding general LUE classification (Figure 3.4d) was therefore non-uniform, with 542 LGUs (35%) classified as efficient and 1,027 LGUs (65%) classified as inefficient.

Table 3.4: Nationwide sensitivity of SDG Indicator 11.3.1 metrics (LCR, PGR, and LCRPGR) to alternative built-up area and population inputs for the Philippines (2000–2015). Values are national aggregates calculated from LGU-level totals ($N = 1,569$ LGUs after filtering). LGUs reporting zero built-up area or zero population in any candidate dataset at either time point were excluded to ensure a consistent set of reporting units across configurations. LCR and PGR are annual rates. National LUE classes are reported as E (Efficient) and I (Inefficient).

Configu- ration	Dataset Combination	LCR (%)	PGR (%)	LCRPGR	Δ LCRPGR	LUE Class
BUA0	GHS-BUILT-S + Census	1.68	1.86	0.90	—	E
BUA1	GAUD + Census	2.30	1.86	1.24	+0.34	I
BUA2	GISA + Census	1.78	1.86	0.96	+0.06	E
BUA3	GISD30 + Census	2.72	1.86	1.46	+0.56	I
BUA4	GLCLUC + Census	3.34	1.86	1.80	+0.90	I
BUA5	WSF Evolution + Census	1.89	1.86	1.02	+0.12	I
Pop1	GHS-BUILT-S + GHS-POP	1.68	1.90	0.88	−0.02	E
Pop2	GHS-BUILT-S + WorldPop	1.68	1.45	1.16	+0.26	I

3.4.3 Sensitivity to BUA Representation

Replacing the baseline built-up dataset with alternative EO-derived representations (BUA1–BUA5), while holding population fixed to the census, produced pronounced changes in LCR and, by propagation, LCRPGR at the national level (Table 3.4). Relative to the baseline, national LCR increased under all built-up substitutions. The largest increase occurred under GLCLUC (3.34% versus 1.68% in the baseline), raising national LCRPGR from 0.90 to as high as 1.80 and reversing the national

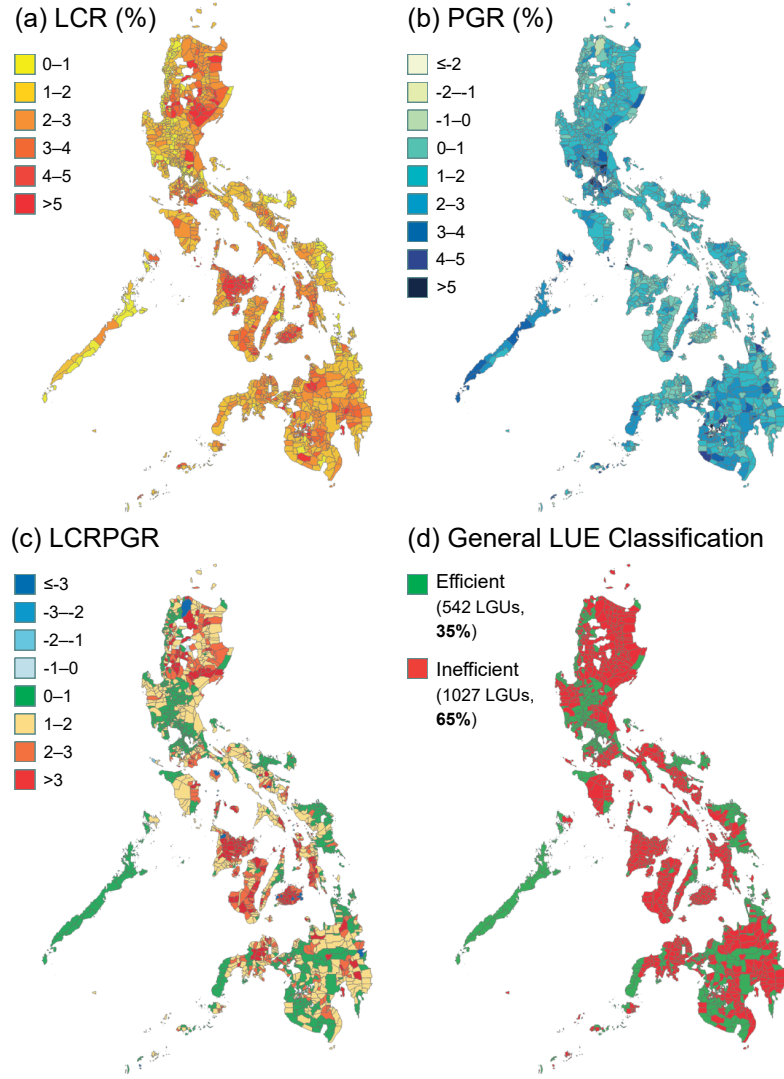


Figure 3.4: Baseline LGU-level maps of LCR, PGR, and LCRPGR for the Philippines (2000–2015), calculated using GHS-BUILT-S built-up area and census population inputs. Each value range includes its upper limit. Results are shown for the harmonized intersection set of $N = 1,569$ LGUs (out of 1,634); excluded LGUs are shown in white.

general LUE classification to inefficient for most configurations (Table 3.4). GISA was the main exception at the national scale, where LCRPGR remained close to unity and the national class remained efficient.

At the LGU level, built-up substitutions shifted the distributions of both ΔLCR and ΔLCRPGR predominantly upward, with configuration-dependent magnitude and spread (Table 3.5; Figure 3.6a–b). Across the harmonized LGU set, LCR was positive for every LGU under both the baseline and all alternative built-up datasets. Accordingly, ΔLCR and ΔLCRPGR reflected differences in the magnitude of positive built-up growth rates across datasets rather than sign reversals. GAUD (BUA1) produced the largest typical departures from baseline (median $\Delta\text{LCR} = 2.70\%$; median $\Delta\text{LCRPGR} = 1.89$) and the widest interquartile ranges (Table 3.5). GLCLUC (BUA4) also showed a clear positive central shift (median $\Delta\text{LCRPGR} = 1.02$). GISA and GISD30 yielded smaller central

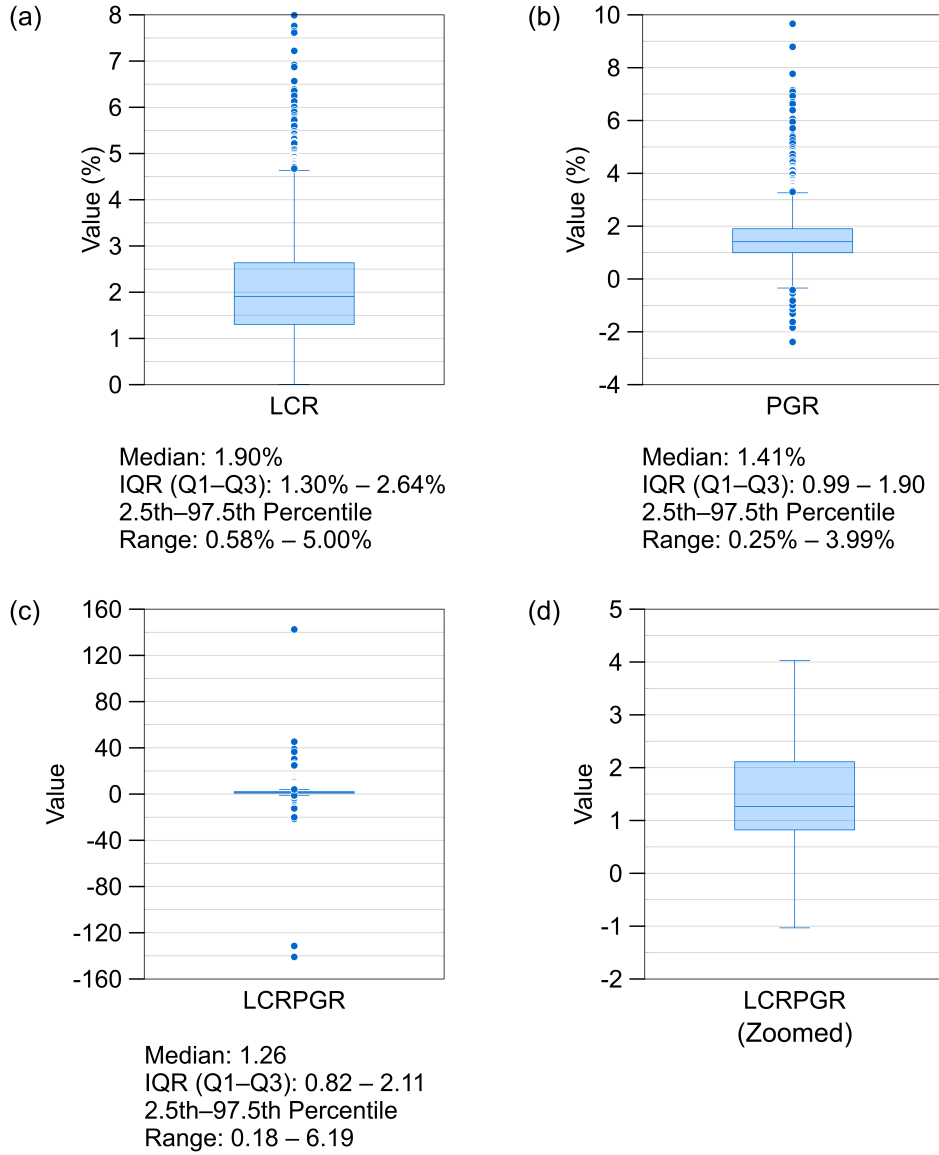


Figure 3.5: Baseline LGU-level distributions of LCR, PGR, and LCRPGR for the Philippines (2000–2015) under the baseline configuration (GHS-BUILT-S + Census), summarized across the harmonized intersection set of $N = 1,569$ LGUs. Boxplots show the median and interquartile range (IQR), with whiskers following the $1.5 \times \text{IQR}$ convention; points beyond the whiskers are retained and shown. Summary statistics reported below each panel (median, IQR, and 2.5th–97.5th percentiles) describe the distribution across LGUs. In the zoomed LCRPGR boxplot (d), outlier points (values beyond the $1.5 \times \text{IQR}$ whisker limits) are not displayed for visual clarity.

departures (median $\Delta \text{LCRPGR} \leq 0.49$) with tighter spreads. WSF-Evolution (BUA5) was closest to the baseline, with the smallest median shifts ($\Delta \text{LCR} = 0.25\%$; $\Delta \text{LCRPGR} = 0.16$) and the narrowest interquartile ranges (Table 3.5; Figure 3.6a–b). The full-range panels show that a small number of LGUs exhibited extreme ΔLCR and ΔLCRPGR values; the zoomed panels highlight differences in the central distributions that characterized the majority of LGUs (Figure 3.6a–b).

These distributional shifts corresponded to changes in LUE classification (Table 3.5; Figure 3.6c). Across configurations, the majority of LGUs remained stable ($\approx 70\text{--}84\%$). However, class transitions occurred in both directions. GLCLUC yielded the largest share of *Efficient-to-Inefficient* transitions

Table 3.5: LGU-level descriptive summary of built-up area sensitivity relative to the baseline configuration (BUA0: GHS-BUILT-S + Census) for 2000–2015 ($N = 1,569$ LGUs; harmonized intersection). Census population inputs are held constant across configurations; only the built-up dataset varies (BUA1–BUA5). Reported values summarize the distributions of ΔLCR and ΔLCRPGR across LGUs using the median and interquartile range (Q1–Q3), together with the percentage of LGUs with stable LUE class and class transitions (*Efficient-to-Inefficient*, *Inefficient-to-Efficient*) relative to the baseline.

Configuration	Built-up Dataset	ΔLCR (%) Median (Q1–Q3)	ΔLCRPGR Median (Q1–Q3)	Stable (%)	Efficient- to- Inefficient (%)	Inefficient- to- Efficient (%)
BUA1	GAUD	2.70 (0.57–8.44)	1.89 (0.31–6.56)	72.47	22.75	4.78
BUA2	GISA	0.31 (−0.62–2.12)	0.19 (−0.43–1.60)	70.87	14.21	14.91
BUA3	GISD30	0.65 (−0.21–1.88)	0.46 (−0.16–1.14)	70.24	20.84	8.92
BUA4	GLCLUC	1.56 (0.46–3.13)	1.02 (0.24–2.53)	72.02	25.62	2.36
BUA5	WSF Evolution	0.25 (−0.13–0.55)	0.16 (−0.09–0.40)	83.94	11.98	4.08

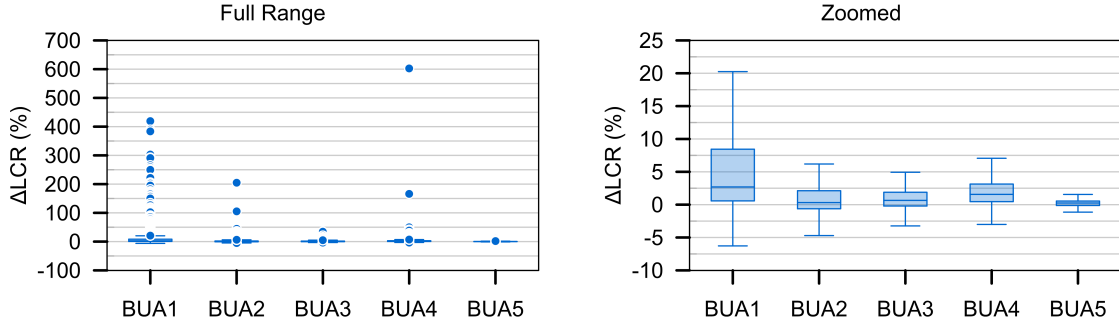
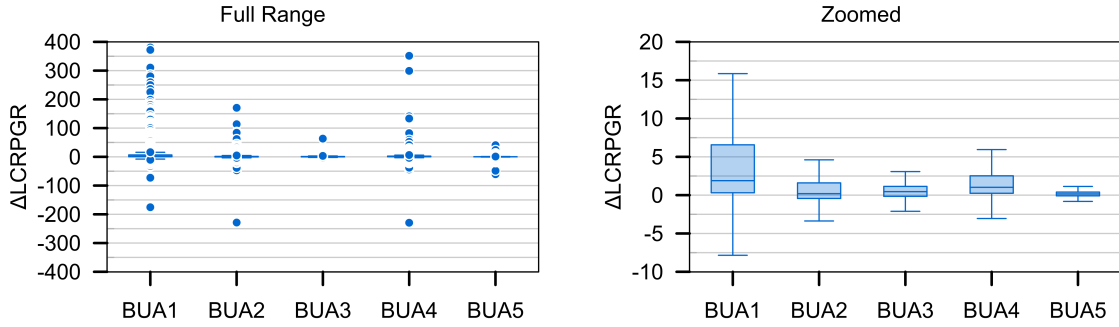
($\approx 25.6\%$), whereas WSF-Evolution showed the highest stability ($\approx 83.9\%$) and the smallest shares of transitions overall (Table 3.5).

Spatially, differences relative to the baseline were geographically structured. Clusters of LGUs with similar ΔLCR and ΔLCRPGR values appeared across multiple built-up substitutions (Figure 3.7a–b), and class transition maps show localized areas where LGUs were repeatedly reclassified under particular built-up datasets (Figure 3.7c).

3.4.4 Sensitivity to Pop Representation

When Pop inputs were varied while holding BUA constant (Pop1–Pop2), national-level effects were more moderate but dataset-dependent (Table 3.4). Substitution with GHS-POP resulted in a marginal increase in national PGR from 1.86% to 1.90%, yielding an LCRPGR of 0.88 and preserving the efficient national classification. In contrast, the use of WorldPop lowered the national PGR to 1.45%, increased the LCRPGR to 1.16, and reversed the national LUE classification to inefficient.

At the LGU level, both population substitutions produced positive shifts in ΔPGR and negative shifts in ΔLCRPGR relative to the baseline (Table 3.6; Figure 3.8a–b). Median ΔPGR was 0.86% for GHS-POP (Pop1) and 0.20% for WorldPop (Pop2), while median ΔLCRPGR was -0.53 for Pop1 and -0.30 for Pop2 (Table 3.6). The interquartile ranges indicate that Pop1 yields a stronger and more consistently negative ΔLCRPGR distribution than Pop2 (Table 3.6). A small number of LGUs exhibited extreme values in the full-range panels, while the zoomed panels emphasize the central differences between Pop1 and Pop2 (Figure 3.8a–b).

(a) Δ LCR Distribution(b) Δ LCRPGR Distribution

(c) LUE Class Transition

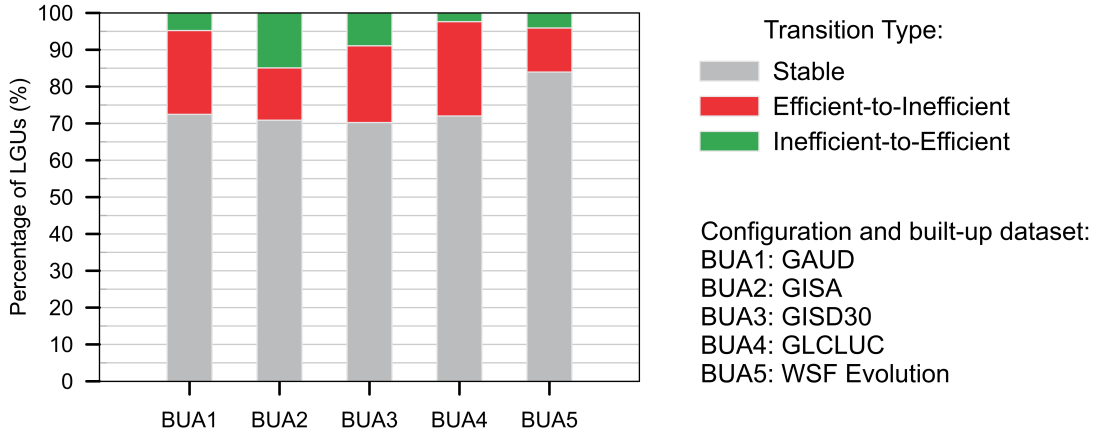


Figure 3.6: LGU-level built-up area sensitivity relative to the baseline configuration (BUA0: GHS-BUILT-S + Census) for built-up area substitutions (BUA1–BUA5), 2000–2015 ($N = 1,569$ LGUs; harmonized intersection). Census population is held constant; only the built-up area dataset is varied. Panels show (a) distributions of Δ LCR (%), (b) distributions of Δ LCRPGR with full-range (left) and zoomed (right) views, and (c) LUE class transitions relative to the baseline, expressed as the percentage of LGUs remaining stable or changing class. Boxplots show the median, interquartile range, and $1.5 \times \text{IQR}$ whiskers. In the zoomed panels, outliers are omitted for clarity but retained in the calculations.

These shifts translated into substantial LUE class transitions (Table 3.6; Figure 3.8c). Under Pop1 (GHS-POP), stability decreased to 65.71%, with 28.36% of LGUs transitioning from *Inefficient-to-Efficient* and 5.93% transitioning from *Efficient-to-Inefficient*. Under Pop2 (WorldPop), stability was higher (72.15%), with 17.34% *Inefficient-to-Efficient* and 10.52% *Efficient-to-Inefficient* transi-

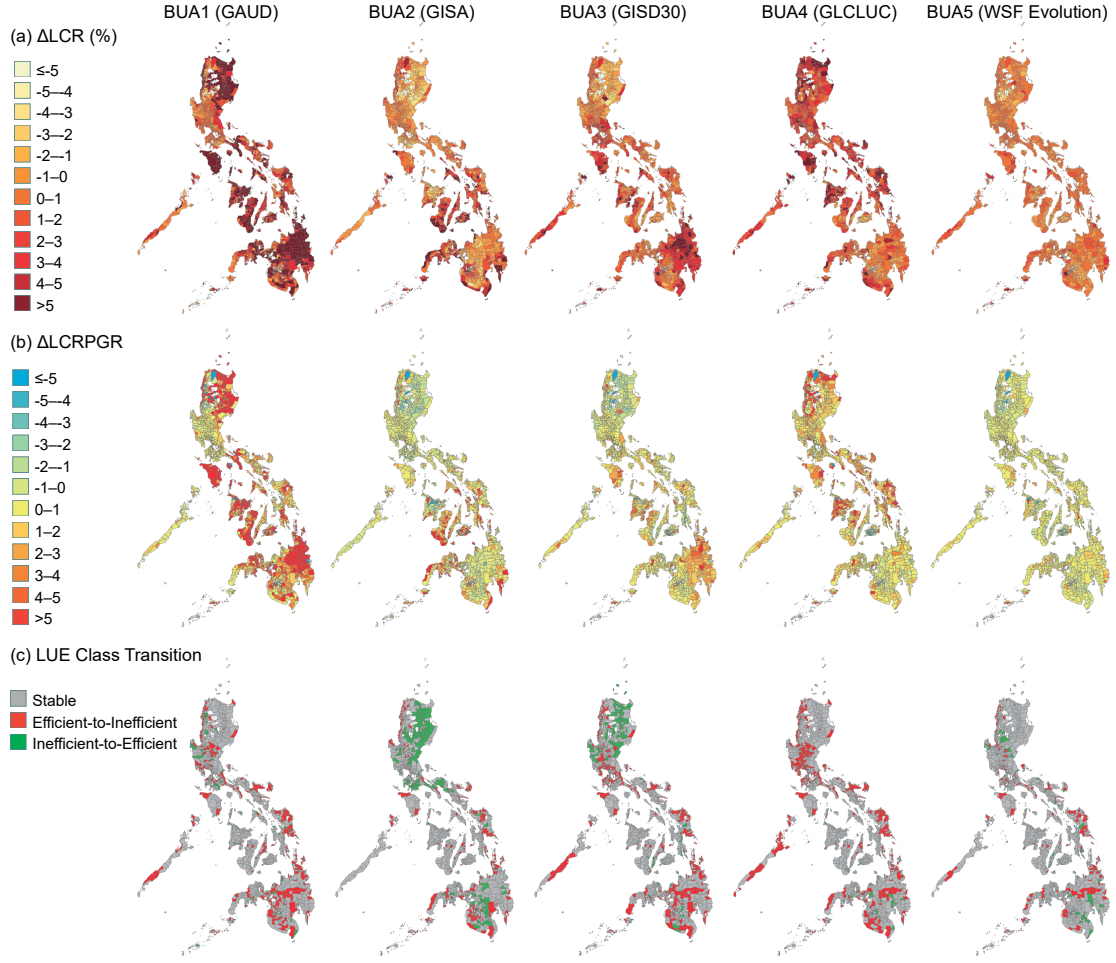


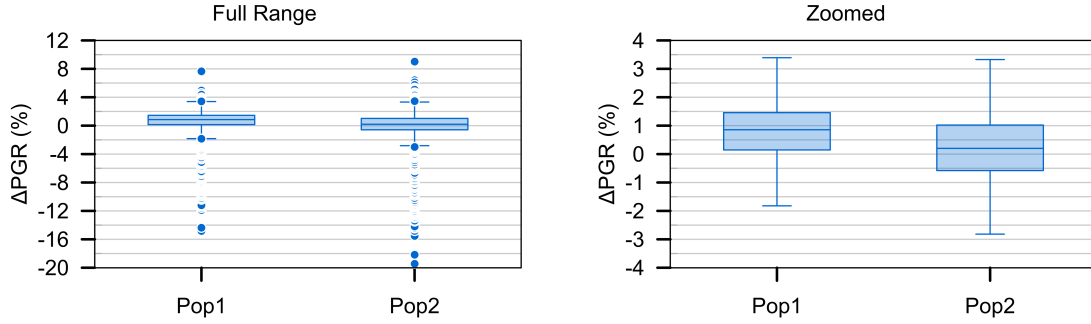
Figure 3.7: Philippines LGU-level sensitivity maps for 2000–2015 showing changes in (a) LCR, (b) LCRPGR, and (c) LUE class transition relative to the baseline when substituting the baseline built-up input with alternative datasets while holding population constant. Each value range includes its upper limit. Results are shown for the harmonized intersection set of $N = 1,569$ LGUs (out of 1,634); excluded LGUs are shown in white.

Table 3.6: LGU-level descriptive summary of Pop sensitivity relative to the baseline configuration (BUA0: GHS-BUILT-S + Census) for 2000–2015 ($N = 1,569$ LGUs; harmonized intersection). Built-up area inputs are held constant across configurations; only the population dataset varies (Pop1–Pop2). Reported values summarize the distributions of ΔPGR and $\Delta LCRPGR$ across LGUs using the median and interquartile range (Q1–Q3), along with the percentage of LGUs with stable LUE class and class transitions (*Efficient-to-Inefficient*, *Inefficient-to-Efficient*) relative to the baseline.

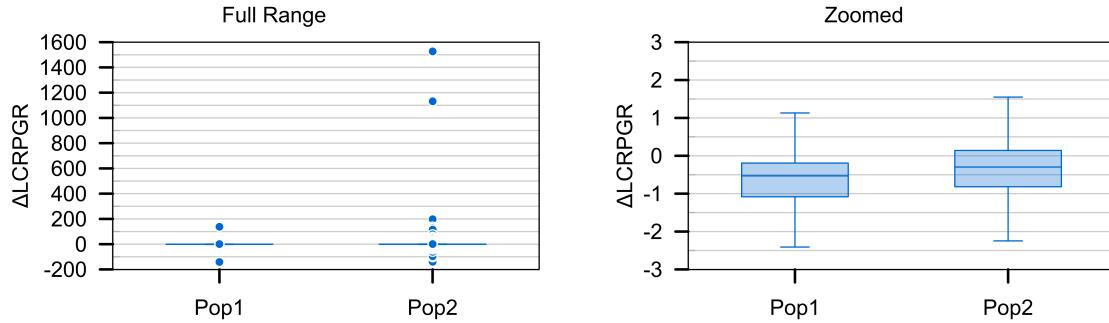
Configuration	Population Dataset	ΔPGR (%) Median (Q1–Q3)	$\Delta LCRPGR$ Median (Q1–Q3)	Stable (%)	Efficient-to-Inefficient (%)	Inefficient-to-Efficient (%)
Pop1	GHS-POP	0.86 (0.15–1.46)	−0.53 (−1.08–0.19)	65.71	5.93	28.36
Pop2	WorldPop	0.20 (−0.58–1.02)	−0.30 (−0.82–0.14)	72.15	10.52	17.34

tions (Table 3.6). Spatially, ΔPGR , ΔLCRPGR , and LUE transition patterns were geographically clustered rather than uniformly distributed (Figure 3.9a–c).

(a) ΔPGR Distribution



(b) ΔLCRPGR Distribution



(c) LUE Class Transition

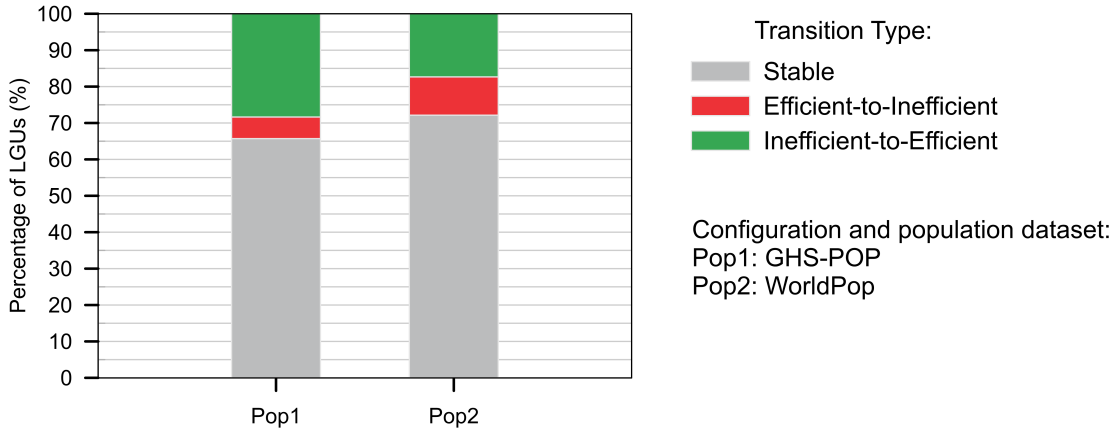


Figure 3.8: LGU-level population sensitivity relative to the baseline configuration (BUA0: GHS-BUILT-S + Census) for population substitutions (Pop1–Pop2), 2000–2015 ($N = 1,569$ LGUs; harmonized intersection). GHS-BUILT-S is held constant; only the population dataset varies. Panels show (a) distributions of ΔPGR (%), (b) distributions of ΔLCRPGR with full-range (left) and zoomed (right) views, and (c) LUE class transitions relative to the baseline, expressed as the percentage of LGUs remaining stable or changing class. Boxplots show the median, interquartile range, and $1.5 \times \text{IQR}$ whiskers. In the zoomed panels, outliers are omitted for clarity but retained in the calculations.

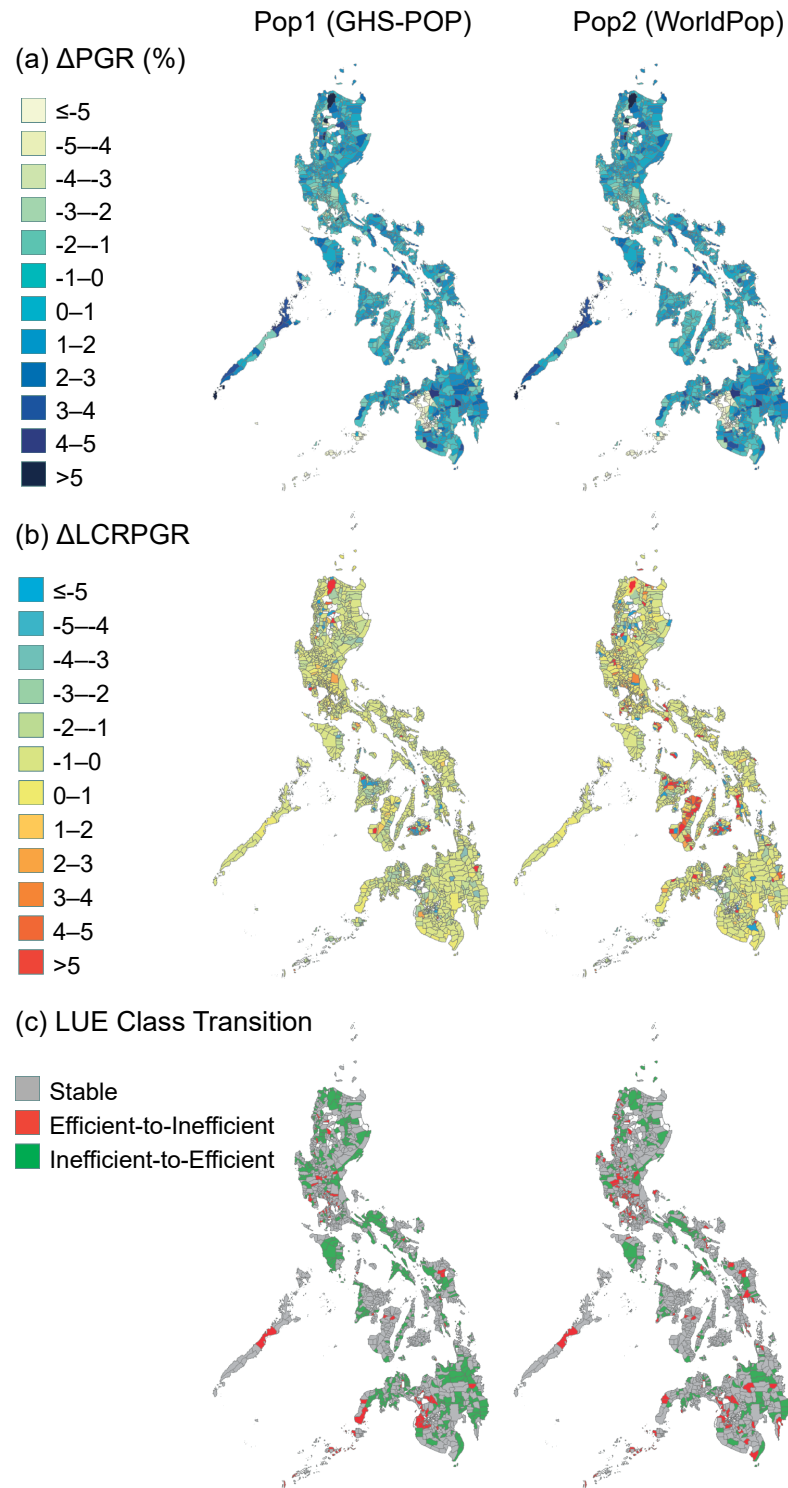


Figure 3.9: Philippines LGU-level sensitivity maps for 2000–2015 showing changes in (a) PGR, (b) LCRPGR, and (c) LUE class transition relative to the baseline when substituting the baseline population input with alternative datasets while holding built-up area constant. Each value range includes its upper limit. Excluded LGUs are depicted in white.

3.4.5 City Examples for Interpreting Deterministic Sensitivity

To complement the nationwide patterns, Table 3.7 presents baseline and alternative-configuration LCRPGR values and corresponding LUE classes for four representative cities: Manila, Baguio City, Cebu City, and Davao City.

Under the baseline configuration (BUA0), all four cities were classified as efficient, with LCRPGR values well below unity. Substitution of built-up datasets produced markedly different LUE classifications across cities. In Manila, most built-up alternatives maintained an efficient classification, with only GLCLUC (BUA4) shifting LCRPGR sharply upward to 3.24, resulting in an inefficient classification. These results mirrored the strong positive Δ LCRPGR observed for GLCLUC at the LGU level (Section 3.4.3). In Baguio City and Cebu City, GAUD (BUA1) induced the largest upward shifts, pushing both cities into the inefficient category, while other built-up alternatives generally preserved efficiency. In Davao City, sensitivity was more pronounced: GAUD, GISA, and GISD30 all exceeded the threshold, resulting in inefficient classifications, whereas GLCLUC and WSF-Evolution remained efficient. These patterns are consistent with earlier results showing that spatially extensive or heterogeneous contexts exhibited stronger deterministic sensitivity to built-up representation.

Population substitutions (Pop1–Pop2) generated responses in opposite directions. For Manila, both GHS-POP and WorldPop yielded negative PGR values, denoted by asterisks in Table 3.7, producing negative LCRPGR values and shifting the city from efficient to inefficient. This represented a case of built-up expansion despite population decline. In contrast, Baguio City, Cebu City, and Davao City remained efficient under both population substitutions, with only moderate changes in LCRPGR. These results align with Section 3.4.4, where population substitutions generally reduced LCRPGR, expressed as negative Δ LCRPGR, but did not uniformly trigger class reversals across cities.

Table 3.7: Baseline (BUA0) and alternative-configuration LCRPGR values for Manila, Baguio City, Cebu City, and Davao City over 2000–2015, and corresponding general LUE class. E = Efficient and I = Inefficient. An asterisk (*) denotes cities where negative PGR was obtained under the indicated population substitution.

Configuration	Dataset Combination	Manila		Baguio City		Cebu City		Davao City	
		LCR PGR	LUE Class	LCR PGR	LUE Class	LCR PGR	LUE Class	LCR PGR	LUE Class
BUA0	GHS-BUILT-S + Census	0.11	E	0.46	E	0.21	E	0.71	E
BUA1	GAUD + Census	0.21	E	1.84	I	1.48	I	1.78	I
BUA2	GISA + Census	0.13	E	0.57	E	0.72	E	1.16	I
BUA3	GISD30 + Census	0.23	E	0.92	E	0.82	E	1.79	I
BUA4	GLCLUC + Census	3.24	I	0.39	E	0.74	E	0.76	E
BUA5	WSF Evolution + Census	0.18	E	0.50	E	0.31	E	0.81	E
Pop1	GHS-BUILT-S + GHS-POP	−0.24*	I	0.49	E	0.19	E	0.71	E
Pop2	GHS-BUILT-S + WorldPop	−0.46*	I	0.51	E	0.13	E	0.81	E

3.5 Discussion

3.5.1 Directional Effects of BUA and Pop Substitution

The OFAT experiments show that deterministic variability in SDG Indicator 11.3.1 outcomes is strongly shaped by which input component is substituted, with BUA and Pop changes producing distinct and largely opposite responses in ΔLCRPGR and LUE transitions. When BUA was varied while Pop was held fixed (BUA1–BUA5), ΔLCRPGR shifts predominantly upward relative to the baseline, and reclassification was dominated by *Efficient-to-Inefficient* transitions. GAUD (BUA1) produced the strongest and most variable departures from the baseline, with a high median ΔLCRPGR and a wide interquartile range, indicating that the sensitivity signal was widespread across LGUs rather than driven by a few outliers. GLCLUC (BUA4) also induced a clear positive shift in ΔLCRPGR and yielded the largest share of *Efficient-to-Inefficient* transitions. By contrast, WSF-Evolution (BUA5) remained closest to the baseline, with ΔLCRPGR values concentrated near zero and the highest proportion of stable classifications, implying that its built-up representation produced the most baseline-consistent LUE outcomes in this deterministic setting.

A different pattern emerged when Pop is varied while BUA was held fixed (Pop1–Pop2). In both Pop substitutions, ΔLCRPGR shifted downward, as indicated by negative medians, and the dominant transition direction is *Inefficient-to-Efficient*. This effect was strongest under GHS-POP (Pop1), which produced the highest share of *Inefficient-to-Efficient* reclassifications, whereas World-Pop (Pop2) showed a similar but more moderate response. While Pop substitution can reclassify a substantial fraction of LGUs, the direction of change is consistently toward lower LCRPGR and more efficient LUE status relative to the baseline.

These results indicate that BUA substitutions tend to push LUE outcomes toward inefficiency. In contrast, Pop substitutions tend to pull them toward efficiency under the same spatial units, time interval, and indicator formulation. This contrast is reinforced by the city examples. In several cities, BUA substitutions, particularly GLCLUC, produced large upward shifts in LCRPGR and, in some cases, reversed the LUE class relative to the baseline. In contrast, Pop substitutions generally reduced LCRPGR, and where reclassification occurred, it tended to move cities toward more efficient status. These examples demonstrate that the directional behavior observed nationally is not an artifact of aggregation but is also evident in major urban centers.

3.5.2 Why BUA Substitutions Exert the Stronger Effect

Among the evaluated dataset combinations, the BUA component exerts the strongest influence on deterministic variability in SDG Indicator 11.3.1. This finding is evident in two ways. First, BUA substitutions generated the largest shifts in ΔLCRPGR and the widest spreads, including substantial interquartile ranges that reflected differences across many LGUs. Second, BUA substitutions accounted for the largest share of *Efficient-to-Inefficient* transitions, reaching roughly one quarter of LGUs in some datasets, with direct consequences for how local LUE performance is interpreted. Pop substitutions, although capable of producing many transitions, tended to move LGUs in the opposite direction and do so with more constrained ΔLCRPGR shifts.

A likely reason for this stronger BUA effect is that the built-up products do not consistently measure the same spatial phenomenon. The baseline dataset, GHS-BUILT-S, represents built-up surfaces as a continuous area-based measure, more closely tied to roofed structures or building footprints. By contrast, the alternative products correspond to broader but related concepts, including urban area, impervious surface, settlement extent, or land-cover classes, which allow the inclusion of developed features even when built-up is not the dominant pixel component. These thematic differences are compounded by differences in spatial encoding. For several alternative products, year-specific BUA layers are derived by thresholding event-based or categorical maps to produce binary “built-up by year” masks, so that once a pixel was classified as built-up, its entire area contributed to the total, regardless of the within-pixel built fraction. Relative to a continuous built-up surface layer, this logic can systematically inflate mapped BUA totals and, in turn, increase LCR and LCRPGR.

This interpretation is consistent with the product-specific behavior observed in the results. GLCLUC, which adopts the broadest land-use-oriented built-up concept among the evaluated products, produces the largest upward departures from the baseline and the largest share of *Efficient-to-Inefficient* transitions. GAUD also produced strong departures, suggesting that its urban-area representation captured spatial elements that extend beyond the narrower built-up surface concept of the baseline. By contrast, WSF-Evolution remained closest to the baseline, indicating that its settlement representation was more aligned, in this study context, with the spatial footprint captured by GHS-BUILT-S. Thus, the observed sensitivity pattern is not solely a matter of numerical disagreement across datasets, but more fundamentally reflects semantic and representational non-equivalence in how BUA is defined and encoded.

The reported accuracy summaries in Table 3.2 provide additional context for this interpretation. Although those values are drawn from different studies and are therefore not strictly comparable, they suggest that part of the observed upward shift in LCRPGR may also be associated with differences in dataset performance, particularly commission tendencies. In this regard, GLCLUC, which shows relatively low user’s accuracy for stable built-up lands and built-up gains in the cited source, may capture non-built-up areas as built-up, thereby contributing to larger estimated BUA totals and stronger upward shifts in LCRPGR. More generally, products with lower user’s accuracy for the built-up class imply higher commission error and therefore a larger potential to overestimate built-up extent. This accuracy-related explanation, however, should be regarded as supportive rather than definitive, since the reported metrics were obtained under different validation designs, study areas, and class definitions.

This dominance of the BUA component is also visible in the reference-city summaries (Table 3.7). Across the four cities, substituting the BUA dataset was sufficient to change the LUE class in some cases, whereas Pop substitutions more often shifted the LCRPGR value without changing the class. These examples underscore that differences in the built-up concept and mapping logic across products can materially alter SDG Indicator 11.3.1 efficiency outcomes even for prominent, well-established urban centers used as reference cases in this chapter. In practice, this means that the choice of built-up dataset is the primary driver of whether an LGU is labeled as efficient or inefficient in deterministic SDG Indicator 11.3.1 reporting. In contrast, population choice acts as a secondary modifier that can soften or amplify the apparent efficiency status.

The city examples further suggest that the magnitude of this sensitivity depends not only on the dataset but also on local geographic context. In Baguio City, mixed land cover and fragmented urban development likely amplified differences between continuous built-up surface estimation and broader urban or impervious representations. In Davao City, the large administrative extent increased the influence of dispersed settlements, transport corridors, and other linear or low-density developed features that may be included differently across datasets. Accordingly, dataset effects are not spatially uniform: they interact with the morphology, fragmentation, and administrative extent of the reporting unit.

3.5.3 Relation to Previous Studies

These findings are consistent with prior work arguing that SDG Indicator 11.3.1 behavior is shaped not only by urban change processes but also by uncertainty and semantic ambiguity in the underlying spatial datasets. Calka et al. (2022), for example, highlighted that sources of uncertainty, such as land-cover definition ambiguity, thematic misclassification, and spatial generalization, can contribute to extreme or unstable LCRPGR values. Complementary sensitivity and robustness studies (e.g., Cardenas-Ritzert et al., 2024b; Zhong et al., 2025a, 2025b) further show that controlled variations in land inputs, including resolution and delineation quality, as well as population gridding assumptions, can yield substantial divergence in LCRPGR estimates because of the nonlinear structure of PGR and the ratio form of LCRPGR. The results of this chapter extend these insights by showing that routine dataset substitution alone, without perturbing resolutions or injecting synthetic error, can produce systematic shifts in Δ LCRPGR and non-trivial LUE reclassification across LGUs and reference cities, with built-up representation exerting the stronger effect in this study context.

These findings also provide important context for earlier national-scale LUE assessments based exclusively on GHSL inputs (e.g., Santillan & Heipke, 2023, 2024a). Those studies evaluated Philippine LUE using a single built-up and population representation, namely GHS-BUILT-S and GHS-POP. The present results show that alternative, yet commonly used, BUA and Pop datasets can materially alter LCRPGR values and induce non-trivial LUE reclassification at both national and LGU levels. These results do not invalidate the earlier GHSL-based findings; rather, they indicate that those findings should be understood as valid for the specific dataset combination used, but not as the only plausible representation of LUE outcomes. The magnitude and direction of Δ LCRPGR observed here therefore suggest that reported efficiency patterns in single-dataset implementations should be interpreted as conditional on the chosen data representation.

3.5.4 Implications for SDG Indicator 11.3.1 Monitoring and Policy Use

The results have immediate implications for monitoring SDG Indicator 11.3.1. Dataset-driven changes in either LCR or PGR can shift an LGU across the efficiency threshold even when underlying urban dynamics are unchanged. The results of this chapter show that such threshold crossings are not rare: depending on the dataset substitution, about 15–30% of LGUs can change class. Consequently, LUE classifications derived from a single built-up or population product should not be

interpreted as uniquely “correct” or directly comparable across studies unless the input datasets and their built-up and population concepts are explicitly aligned. For monitoring frameworks that aim to track progress over time or compare performance across jurisdictions, the sensitivity highlighted here implies that changes in reported efficiency can arise from routine data choices rather than from actual changes in LUE.

From a policy perspective, these findings matter because SDG Indicator 11.3.1 outputs can be used to support accountability narratives, prioritize interventions, and allocate technical assistance. If an LGU’s classification can switch from efficient to inefficient, or vice versa, simply because a different built-up or population product is adopted, then policy decisions based on a single deterministic classification risk being unstable. A practical implication is that agencies and practitioners should document the specific datasets used and the concepts they represent, avoid treating LUE classes as definitive without sensitivity checks, and interpret “efficient/inefficient” labels as conditional on data representation rather than as absolute statements of performance. More broadly, the results support the need for monitoring approaches that report uncertainty or sensitivity ranges alongside point estimates, and for guidance that clarifies how differences in built-up definitions, such as roofed structures versus impervious, settlement, urban, or broader developed land representations, may systematically shift national and local conclusions under SDG Indicator 11.3.1.

3.6 Limitations and Further Work

This chapter was intentionally designed as a deterministic sensitivity analysis to isolate how alternative representations of BUA and Pop affect SDG Indicator 11.3.1 outcomes under controlled conditions. Accordingly, several limitations should be acknowledged. First, once selected, each dataset was treated as fixed and exact, such that within-product uncertainty was not explicitly represented. This includes, for example, built-up misclassification, spatial generalization effects, and uncertainty in gridded population estimates. As summarized in Table 3.2, the datasets considered in this chapter also differ in their reported accuracies, suggesting that part of the observed variation may reflect not only differences in thematic definition and spatial representation, but also differences in dataset performance. Second, the OFAT design varied BUA and Pop separately. While this was useful for isolating their individual effects, it did not capture their joint uncertainty or any dependence structure between them. Third, the larger number of BUA datasets relative to Pop datasets may have widened the observed range of BUA-related variation and therefore contributed to the stronger apparent influence of BUA substitution in this chapter.

These limitations also indicate the next step of the dissertation. Further work should move beyond deterministic dataset substitution toward explicit uncertainty representation and propagation, so that SDG Indicator 11.3.1 outputs can be interpreted not only as conditional on dataset choice, but also in terms of their uncertainty and degree of classification support.

3.7 Chapter Conclusion

This chapter examined the sensitivity of SDG Indicator 11.3.1 to alternative representations of BUA and Pop under a controlled deterministic design. By systematically substituting commonly used EO-derived BUA and gridded Pop datasets while holding spatial units, temporal interval, and indicator formulations constant, the analysis showed that routine input-data choices alone can materially alter LCR, PGR, LCRPGR, and resulting LUE classifications. The findings directly address RQ1 by demonstrating that SDG Indicator 11.3.1 is meaningfully sensitive to representation differences in retrospective LUE assessment, with BUA substitutions exerting the stronger influence in this study context. These results establish the need to move beyond single-value deterministic reporting toward approaches that explicitly account for uncertainty, which is the focus of Chapter 4.

4 Uncertainty Propagation and Probabilistic LUE Classification for SDG Indicator 11.3.1⁷

4.1 Chapter Overview

Building on the deterministic sensitivity results of Chapter 3, this chapter develops an uncertainty-aware extension of the SDG Indicator 11.3.1 framework in which BUA and Pop are treated as uncertain inputs. The objective is to examine how uncertainty in these inputs propagates through LCR, PGR, and LCRPGR, and how this propagation influences the stability and interpretation of LUE classification. Rather than relying solely on single-point estimates, the chapter presents indicator outputs as probability distributions that capture the effects of input uncertainty, nonlinear indicator transformations, and inter-variable dependence. It also extends this uncertainty-aware perspective to the classification stage through a probabilistic treatment of LUE regimes.

This chapter addresses two research questions:

- *RQ2. How does explicit propagation of uncertainty in built-up area and population affect the magnitude, stability, and LUE classification outcomes of SDG Indicator 11.3.1 compared to conventional deterministic approaches?*
- *RQ3. Given uncertain SDG Indicator 11.3.1 components and outputs, how does probabilistic LUE classification improve the robustness, comparability, and interpretability of LUE assessments relative to deterministic classification?*

To answer these questions, the chapter first introduces a conceptual and mathematical framework for uncertainty propagation that is general with respect to the source of uncertainty (Figure 4.1). In this framework, BUA and Pop are represented by probability distributions that may arise from sampling variability, classification error, measurement uncertainty, definitional ambiguity, or predictive modeling. These input distributions are propagated through the SDG Indicator 11.3.1 formulations using Monte Carlo simulation, yielding empirical distributions for LCR, PGR, and LCRPGR.

⁷This chapter builds upon the author's published work: Santillan et al. (2025a). *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. <https://doi.org/10.5194/isprs-annals-X-G-2025-753-2025>. The uncertainty propagation presented in the paper has been substantially extended and revised. Compared to the published article, this chapter formally introduces the general framework for uncertainty propagation and probabilistic LUE classification and makes use of hypothetical and real-world illustrations of the said framework utilizing new data and study area.

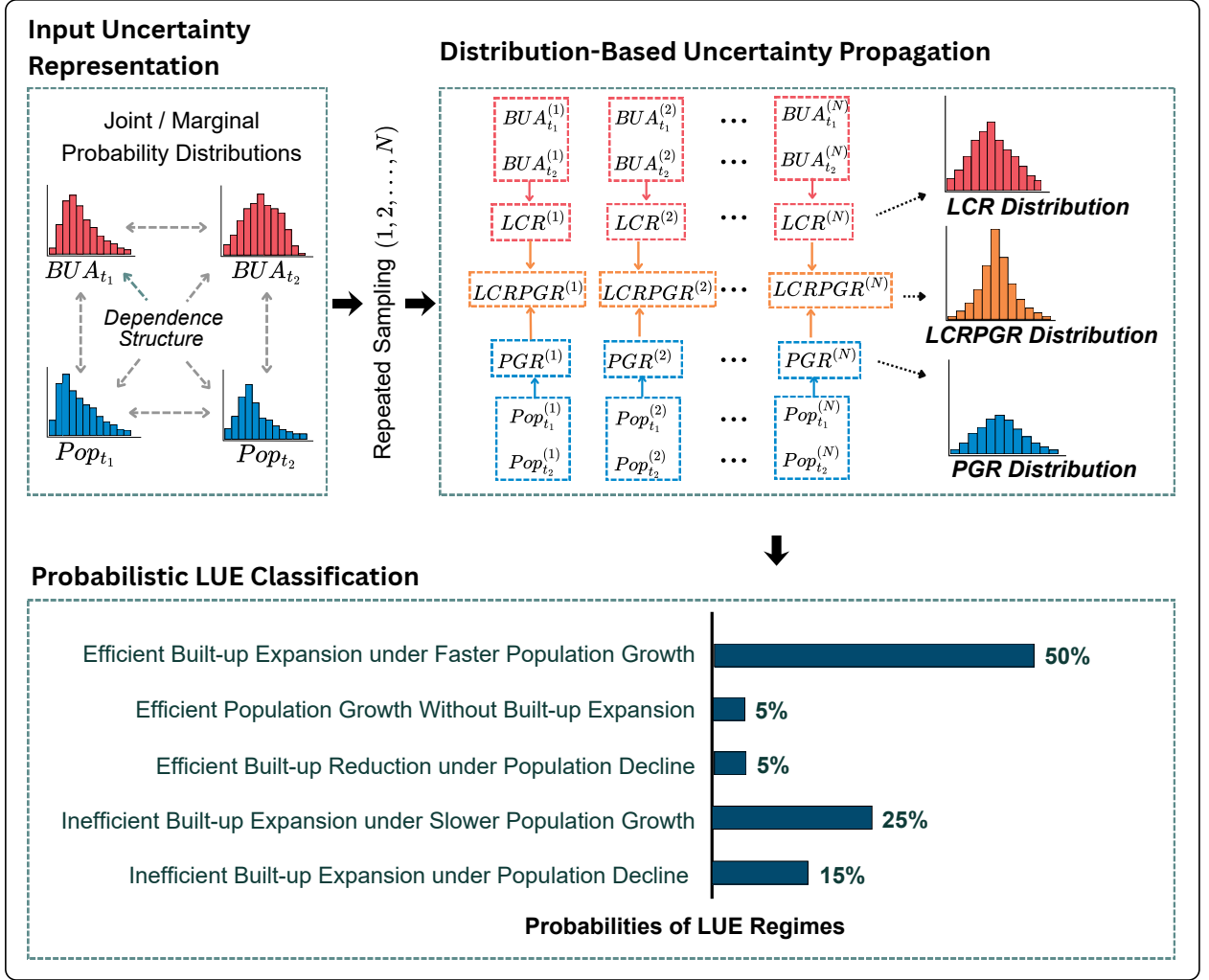


Figure 4.1: Conceptual framework for uncertainty propagation and probabilistic LUE classification for SDG Indicator 11.3.1. Input uncertainty in built-up area and population is represented through marginal or joint probability distributions, including their dependence structure. Repeated sampling from these distributions produces multiple realizations of built-up area, population, LCR, PGR, and LCRPGR. Each realization is classified into an LUE regime based on the corresponding LCR, PGR, and LCRPGR values. The percentages in the lower panel represent the proportion of realizations assigned to each regime and therefore indicate regime membership probabilities. The values shown are illustrative.

A key design principle is that the framework is not tied to a single uncertainty-generation method. The same propagation logic can be used whether uncertainty is obtained from design-based inference, model-based estimation, bootstrap procedures, or predictive distributions from forecasting models. This makes the framework suitable as a common basis for both retrospective monitoring and forward-looking assessment. In the present chapter, the framework is developed under retrospective conditions, with particular attention to uncertainty in EO-derived estimates of built-up area. Later, Chapter 5 extends this foundation by introducing probabilistic forecasting of BUA and Pop for the forward-looking assessment of SDG Indicator 11.3.1.

Beyond uncertainty propagation itself, the chapter introduces a probabilistic LUE classification approach. Instead of assigning each spatial unit to a single regime from point estimates alone,

the method classifies the ensemble of simulated indicator realizations, thereby producing estimated probabilities of regime membership. This provides a more informative basis for interpretation by showing not only the most likely regime, but also the degree of support for that regime relative to competing plausible outcomes.

The remainder of the chapter is structured as follows. Section 4.2 presents the general framework for uncertainty propagation and probabilistic LUE classification, including the conceptual representation of input uncertainty, indicator transformation, distribution-based propagation, extension to secondary SDG 11.3.1 indicators, and probabilistic regime assignment. Section 4.3 then focuses on uncertainty quantification of EO-derived BUA for SDG Indicator 11.3.1 monitoring. It outlines the pipeline adopted in this chapter, including error-adjusted area estimation and uncertainty estimation, followed by bootstrap-based uncertainty estimation and propagation. Section 4.4 contains two illustrative applications: a hypothetical case used to examine the influence of dependence assumptions, and a real-world case in Baguio City (Philippines) demonstrating uncertainty-aware SDG Indicator 11.3.1 monitoring using EO-derived built-up data. Section 4.5 discusses the main findings and methodological implications of the chapter. Section 4.6 outlines limitations and future work, and Section 4.7 concludes the chapter.

4.2 General Framework for Uncertainty Propagation and Probabilistic LUE Classification

4.2.1 Conceptual Representation of Input Uncertainty

In conventional implementations of SDG Indicator 11.3.1, BUA and Pop at each reference epoch are treated as fixed quantities. Under this deterministic interpretation, the indicator components (LCR, PGR, LCRPGR) are calculated as single-valued functions of observed inputs. However, as established in Chapter 3, these inputs depend on the measurement approach, semantic definition, and data source. More fundamentally, BUA and Pop are estimated quantities subject to sampling variability, classification error, spatial generalization, redistribution uncertainty, or predictive error (Olofsson et al., 2014; Leyk et al., 2019).

The uncertainty considered here is primarily epistemic, reflecting incomplete knowledge of the true values of BUA and Pop due to one or more of the factors stated above. Aleatoric variability in the underlying urban process is not explicitly modeled; instead, the framework propagates the uncertainty encoded in the input distributions, including cases where epistemic and aleatoric components are not separable.

To formalize this uncertainty, BUA and Pop are treated as random variables. For two reference epochs t_1 and t_2 of a spatial unit of analysis, define:

$$\text{BUA}_{t_1} \sim p_{\text{BUA}_{t_1}}, \quad \text{BUA}_{t_2} \sim p_{\text{BUA}_{t_2}}, \quad (4.1)$$

$$\text{Pop}_{t_1} \sim p_{\text{Pop}_{t_1}}, \quad \text{Pop}_{t_2} \sim p_{\text{Pop}_{t_2}}, \quad (4.2)$$

where $p_{(\bullet)}$ denotes the probability distribution characterizing uncertainty in each input quantity.

More generally, uncertainty may be represented through a joint distribution:

$$\mathbf{v} = (\text{BUA}_{t_1}, \text{BUA}_{t_2}, \text{Pop}_{t_1}, \text{Pop}_{t_2})^\top \sim p(\mathbf{v}). \quad (4.3)$$

where $\mathbf{v}$ denotes the joint random input vector and $p(\mathbf{v})$ its joint probability distribution.

This joint representation is useful because dependencies may exist both across epochs (e.g., correlated errors between t_1 and t_2 when inputs are generated using the same sensors, models, or reference data) and, in some workflows, across variables (e.g., shared covariates or coupled estimation procedures). In such a case, representing uncertainty only marginally (i.e., treating inputs as independent) can yield implausible combinations of input values that do not reflect their joint behavior (Jiang et al., 2017). Ignoring such dependence structures in this case may distort the indicator component distributions, particularly given the nonlinear transformation inherent in LCRPGR. However, dependence is not always present. In many SDG Indicator 11.3.1 monitoring applications, BUA and Pop are produced by distinct data pipelines (e.g., EO-derived built-up mapping on the one hand and census-based population estimation on the other), in which case cross-variable dependence may be negligible. The framework, therefore, treats the dependence structure as scenario-specific; it can be modeled when justified and omitted when independence is a defensible assumption.

Alternative factorizations of $p(\mathbf{v})$ can be used to encode different dependence assumptions. A common case is within-variable temporal dependence without cross-variable dependence, where:

$$p(\mathbf{v}) = p(\text{BUA}_{t_1}, \text{BUA}_{t_2}) p(\text{Pop}_{t_1}, \text{Pop}_{t_2}). \quad (4.4)$$

A more restrictive case is full independence, where:

$$p(\mathbf{v}) = p(\text{BUA}_{t_1}) p(\text{BUA}_{t_2}) p(\text{Pop}_{t_1}) p(\text{Pop}_{t_2}). \quad (4.5)$$

Conversely, when BUA and Pop share models, covariates, calibration data, or error sources, cross-variable dependence may be present and a full joint representation $p(\mathbf{v})$ is appropriate (Eq. 4.3).

The framework developed here does not assume a specific parametric form for $p(\mathbf{v})$ (or for its factorized components). Input uncertainty may be characterized via design-based inference (e.g., stratified sampling), model-based estimation, bootstrap resampling, ensemble prediction, or any other statistically coherent procedure (e.g., Stehman, 2000, 2009; Olofsson et al., 2013; Efron & Tibshirani, 1986; Gallaun et al., 2015; Lakshminarayanan et al., 2017). The only requirement is that uncertainty can be expressed in distributional form, either analytically or empirically.

4.2.2 Indicator Transformation

Given the (possibly factorized) joint input uncertainty representation specified in Section 4.2.1, the SDG Indicator 11.3.1 components are obtained through the deterministic transformations specified in Eqs. 2.1–2.3. To emphasize the transformation structure, denote the mapping from inputs to indicator components as

$$\mathbf{y} = (\text{LCR}, \text{PGR}, \text{LCRPGR})^\top = T(\mathbf{v}), \quad (4.6)$$

where $T(\bullet)$ represents the composite transformation implied by the SDG Indicator 11.3.1 formulations. Under deterministic inputs, $\mathbf{y}$ is a fixed vector. Under the probabilistic representation introduced in Section 4.2.1, however, $\mathbf{v}$ is random, and therefore $\mathbf{y}$ is also random. Essentially, the SDG Indicator 11.3.1 formulae transform input uncertainty into indicator uncertainty, and they do so in a nonlinear way.

Two structural features of this transformation are central for uncertainty propagation. First, the mapping from $\mathbf{v}$ to $\mathbf{y}$ is nonlinear. LCR depends on ratios of differences in BUA; PGR involves a logarithmic transformation of population change; and LCRPGR is itself a ratio of two derived quantities. Nonlinear transformations generally alter distributional shape, meaning that even symmetric or approximately Gaussian input distributions can produce skewed or heavy-tailed output distributions. This effect becomes more pronounced when denominators are small (e.g., low BUA_{t_1} in the LCR term, or near-zero population growth in the PGR term) and when uncertainties in numerators and denominators are correlated.

Second, the transformation couples uncertainties across epochs, and depending on the dependence assumptions adopted in Section 4.2.1, may also couple uncertainties across variables. Because LCR and PGR each depend on paired inter-epoch quantities, uncertainty in BUA_{t_1} influences LCR jointly with BUA_{t_2} , and similarly for Pop in PGR. Moreover, the ratio structure of LCRPGR implies that uncertainty in LCR and PGR interacts multiplicatively rather than additively. Consequently, propagated uncertainty in LCRPGR cannot, in general, be inferred from marginal uncertainty in any single input alone; it depends on the joint (or factorized) structure used to generate plausible realizations of $\mathbf{v}$, including any temporal dependence within BUA and/or Pop and any cross-variable dependence when such dependence is modeled.

These properties imply that simple linear error-propagation approximations (e.g., first-order Taylor expansion or delta-method approaches) may be unreliable for SDG Indicator 11.3.1, because they linearize T and can fail to capture skewness, heavy tails, and dependence-driven effects arising from ratios and logarithms, especially when PGR is small in magnitude. For this reason, and as described in the following section, a distribution-based propagation approach is adopted that approximates the induced distribution of $\mathbf{y} = T(\mathbf{v})$ directly while retaining the nonlinear transformation and whichever dependence structure (full, block, or independent) is specified for the inputs.

4.2.3 Distribution-Based Uncertainty Propagation

Uncertainty propagation is implemented by approximating the distribution of the SDG Indicator 11.3.1 components induced by the mapping $\mathbf{y} = T(\mathbf{v})$ through repeated evaluation of $T(\bullet)$ over plausible realizations of the input vector $\mathbf{v}$. The procedure adopted here is Monte Carlo-based uncertainty propagation, which constructs an empirical approximation of the output distribution through repeated sampling of plausible inputs under a specified uncertainty model and repeated evaluation of the transformation of interest (Metropolis & Ulam, 1949; Robert & Casella, 2004; JCGM, 2008). In this study, that transformation is given by the SDG Indicator 11.3.1 formulae.

Let N denote the number of simulation draws. For each draw $i = 1, \dots, N$, an input realization $\mathbf{v}^{(i)}$ is generated from the distributional specification defined in Section 4.2.1. Depending on

the dependence assumptions adopted, $\mathbf{v}^{(i)}$ may be generated (i) by sampling all four inputs independently from their marginal distributions, (ii) by sampling from two bivariate distributions ($\text{BUA}_{t_1}, \text{BUA}_{t_2}$) and ($\text{Pop}_{t_1}, \text{Pop}_{t_2}$) and concatenating them (temporal dependence within variables, no cross-variable dependence), or (iii) by sampling from a full four-dimensional joint model when cross-variable dependence is justified.

The joint distribution $p(\mathbf{v})$ may be specified using different modeling strategies depending on the nature of available uncertainty information. A common parametric approach is to assume a multivariate normal distribution defined by a mean vector, marginal standard deviations, and a correlation matrix (Robert & Casella, 2004; JCGM, 2008). When uncertainty arises from sampling-based accuracy assessment, model ensembles, or repeated estimation procedures, empirical or bootstrap-based joint distributions may instead be constructed directly from resampled realizations (e.g., Efron & Tibshirani, 1986; Gallaun et al., 2015). Alternatively, copula-based approaches (e.g., Nelsen, 2006; Nagler, 2025) can be employed to model dependence separately from marginal distributions, allowing flexible non-Gaussian marginals while preserving prescribed dependence properties. The uncertainty propagation framework developed here does not require a specific joint distributional form; it only requires that realizations of the joint input vector can be generated in a manner consistent with the assumed marginal uncertainties and dependence structure.

Once a realization $\mathbf{v}^{(i)}$ has been generated, the SDG Indicator 11.3.1 transformation is evaluated at that realization to obtain $\mathbf{y}^{(i)} = T(\mathbf{v}^{(i)})$. Repeating this procedure yields an empirical sample $\{\mathbf{y}^{(i)}\}_{i=1}^N$, which serves as a non-parametric approximation of the propagated distribution of LCR, PGR, and LCRPGR. Because the approximation is empirical, it does not require distributional assumptions about $\mathbf{y}$ and naturally retains skewness, heavy tails, or other non-Gaussian features generated by the nonlinear transformations.

The simulated sample provides a direct basis for summarizing indicator uncertainty. Point summaries, such as the mean or median, can be calculated for each component of $\mathbf{y}$, and uncertainty intervals can be reported using empirical quantiles, for example, by taking the 2.5th and 97.5th percentiles to form a central 95% interval. Because the resulting distributions may be asymmetric, quantile-based summaries are preferable to variance-only reporting. Beyond descriptive summaries, the propagated distribution enables probability statements that are directly aligned with interpretation needs, such as the probability that an indicator exceeds a threshold or falls within a particular efficiency regime. Such probability statements are especially relevant in the context of SDG Indicator 11.3.1, where LUE interpretation ultimately relies on threshold-defined regimes, so realizations near classification boundaries may be assigned to different regimes under relatively small changes in the uncertain BUA and Pop inputs.

As N increases, the empirical distribution converges to the true distribution induced by $p(\mathbf{v})$ under standard Monte Carlo convergence. In practice, N is selected to ensure stable estimation of the distributional features of interest, particularly tail quantiles and regime probabilities (JCGM, 2008; Janssen, 2013). The output of this propagation step (full distributions of LCR, PGR, and LCRPGR) provides the quantitative foundation for the probabilistic LUE classification framework introduced in Section 4.2.5.

4.2.4 Extension to SDG 11.3.1 Secondary Indicators

The uncertainty-propagation framework described above is not specific to LCR, PGR, or LCRPGR. More generally, once the uncertain inputs ($\text{BUA}_{t_1}, \text{BUA}_{t_2}, \text{Pop}_{t_1}, \text{Pop}_{t_2}$) are represented through a joint (or factorized) distribution and propagated via repeated draws, any SDG Indicator 11.3.1-related quantity that is a deterministic function of the same inputs can be assigned a propagated uncertainty distribution using the same Monte Carlo realizations. This extension is consistent with the SDG Indicator 11.3.1 methodology, which recommends reporting complementary indicators alongside the primary components to support interpretation.

In particular, two recommended secondary indicators, namely Built-up Area per Capita (BUpC) and Total Change in Built-up Area (%) (TCBUA) (UN Statistics Division, 2025) can be obtained directly from each simulated realization $\mathbf{v}^{(i)}$. BUpC at each epoch and TCBUA are calculated as

$$\text{BUpC}_t = \frac{\text{BUA}_t}{\text{Pop}_t}, \quad t \in \{t_1, t_2\} \quad (4.7)$$

$$\text{TCBUA} (\%) = 100 \times \frac{\text{BUA}_{t_2} - \text{BUA}_{t_1}}{\text{BUA}_{t_1}} \quad (4.8)$$

Evaluating these expressions for $i = 1, \dots, N$ yields empirical distributions of these secondary indicators, from which their medians, uncertainty intervals, and probability statements can be reported in the same manner as for the three primary indicator components.

4.2.5 Probabilistic LUE Classification

Once BUA and Pop are treated as uncertain quantities and propagated through the SDG Indicator 11.3.1 formulae, the resulting outputs (LCR, PGR, and LCRPGR) are no longer fixed scalars but random variables. Consequently, regime membership itself becomes uncertain. A spatial unit may plausibly belong to multiple LUE regimes depending on the realization of input uncertainty (Figure 2.1; Table 2.1). Deterministic assignment, therefore, conceals classification variability that may arise from uncertainty in BUA and Pop estimates.

To formalize probabilistic classification, let $\mathbf{y}^{(i)} = \left(\text{LCR}^{(i)}, \text{PGR}^{(i)}, \text{LCRPGR}^{(i)} \right)^\top$ denote the i -th simulated realization obtained from the uncertainty propagation procedure described in Section 4.2.3 under the specified input dependence scenario. For each realization, the corresponding LUE regime $R^{(i)}$ is determined by applying the deterministic classification rules defined in Table 2.1. That is, each simulated triplet $\mathbf{y}^{(i)}$ is mapped to one of the predefined efficiency regimes according to the sign and relative magnitude criteria for LCR, PGR, and LCRPGR.

Let K denote the total number of regimes defined in Table 2.1. The probability that a spatial unit belongs to regime k is then estimated as

$$\hat{P}(R = k) = \frac{1}{N} \sum_{i=1}^N \mathbb{I} \left(R^{(i)} = k \right), \quad (4.9)$$

where N is the number of simulation draws and $\mathbb{I}(\cdot)$ is the indicator function, which equals 1 if the condition is satisfied and 0 otherwise. The set of probabilities $\{\hat{P}(R=1), \hat{P}(R=2), \dots, \hat{P}(R=K)\}$ forms a discrete probability distribution over LUE regimes for each spatial unit.

This formulation yields several important interpretive advantages. First, it allows classification uncertainty to be quantified explicitly. A unit with $\hat{P}(R=1) \approx 1$ exhibits stable regime membership under plausible input variation, whereas a unit with probability mass distributed across multiple regimes is classification-unstable. Second, the approach supports risk-aware interpretation. For example, the probability that LCRPGR exceeds unity, or that LCR is positive while PGR is negative (corresponding to inefficient expansion under demographic decline in Table 2.1), can be reported directly. Such probability statements provide a clearer basis for decision-relevant interpretation than a single deterministic label.

Third, probabilistic classification enables the identification of threshold sensitivity. Because the regime definitions in Table 2.1 rely on sign changes and relative magnitudes of LCR and PGR, small perturbations near zero-growth boundaries may induce regime switching. The Monte Carlo-based framework quantifies the frequency of such threshold crossings under uncertainty. The resulting regime probabilities are conditional on the adopted input uncertainty model and its dependence assumptions; if alternative dependence scenarios yield materially different regime probabilities, this indicates that classification is sensitive to dependence modeling and merits additional justification.

In reporting results, several complementary summaries may be presented. The most probable regime R^* can be identified as

$$R^* = \arg \max_k \hat{P}(R=k), \quad (4.10)$$

which serves as a probabilistic analogue to deterministic assignment. However, unlike conventional classification, this assignment can be accompanied by its associated probability $\hat{P}(R^*)$, thereby communicating the degree of confidence in the classification. Additionally, a single summary measure of classification instability may be reported if desired (e.g., based on the spread of regime probabilities).

In simple terms, probabilistic LUE classification replaces the idea of a single “correct” regime label with a likelihood over regimes. Instead of saying “this city is efficient” based on one fixed set of BUA and Pop values, it is acknowledged that BUA and Pop are uncertain values and thus vary within plausible ranges. When plausible BUA and Pop values are repeatedly drawn (via Monte Carlo), and LCR, PGR, and LCRPGR are recalculated each time, the city can fall into different regimes across draws. The proportion of simulations assigned to each regime can then be interpreted as the probability of membership in that regime under the assumed uncertainty model. This concept is illustrated in Figure 4.2.

By moving from deterministic labels to regime probabilities, the framework provides a formal counterpart to the instability observed in Chapter 3, where dataset substitution alone produced non-trivial reclassification across spatial units. Here, reclassification is quantified explicitly as the likelihood of regime switching under modeled uncertainty in BUA and Pop, establishing a consistent basis for uncertainty-aware interpretation of SDG Indicator 11.3.1 in both retrospective assessments and forward-looking analyses.

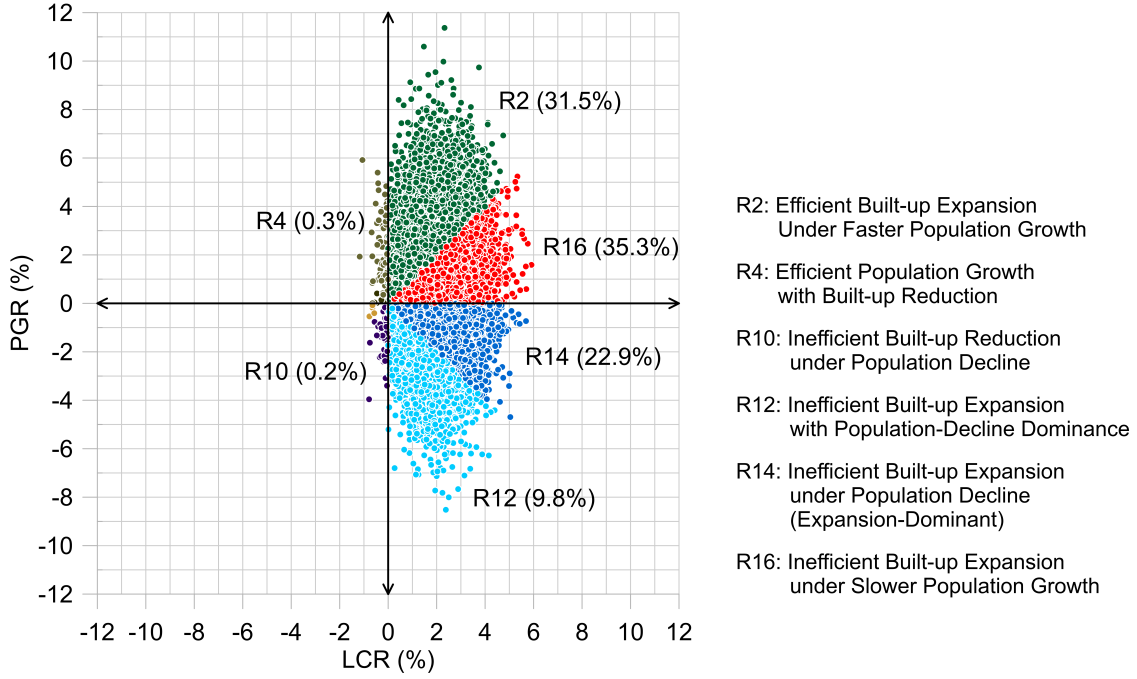


Figure 4.2: Conceptual illustration of probabilistic LUE classification for a hypothetical city with uncertain built-up area (BUA) and population (Pop) inputs. Each point in the LCR–PGR space represents one Monte Carlo realization of LCR and PGR derived from uncertainty-propagated BUA and Pop estimates. Colored clusters indicate LUE regimes defined by the joint signs and relative magnitudes of LCR, PGR, and LCRPGR, following Table 2.1. In this example, the realizations are concentrated mainly in *Inefficient Built-up Expansion under Slower Population Growth* (R16), with substantial shares also falling in R2 and R14. These proportions represent estimated probabilities of regime membership under the assumed uncertainty model, showing that although R16 is the most likely outcome, alternative plausible regime assignments remain substantial.

4.3 Uncertainty Quantification of EO-Derived BUA for SDG Indicator 11.3.1 Monitoring

In the preceding section, the uncertainty propagation framework was formulated under the assumption that uncertainties in BUA and Pop are already available in distributional form, allowing them to be propagated through the SDG Indicator 11.3.1 transformations and used for probabilistic LUE classification. What has not yet been addressed is how such uncertainties, particularly for EO-derived BUA, can be quantified in practice. In this section, a pipeline for uncertainty quantification of EO-derived BUA is introduced, with outputs that can be directly integrated into the general uncertainty propagation framework developed in Section 4.2.

4.3.1 Pipeline Overview

The proposed pipeline begins with an EO-derived built-up change map for the analysis period and a stratified sample of reference observations. The change map serves as the stratification layer for design-based estimation, while the reference sample is used to quantify classification errors through an error matrix of mapped and reference change classes. From this, error-adjusted estimates of

the change-class areas are derived, and the endpoint built-up areas at t_1 and t_2 are reconstructed by summing the relevant change components. Uncertainty in these endpoint estimates is then characterized empirically through stratified bootstrap resampling of the reference sample within each change-map stratum. These procedures yield paired bootstrap replicates of BUA at t_1 and t_2 , which can subsequently be used as empirical input distributions within the uncertainty propagation framework of Section 4.2.1.

4.3.2 Error-Adjusted Area Estimation and Uncertainty Estimation

The first component of the pipeline is the estimation of BUA and its uncertainty from the EO-derived built-up map and the corresponding reference sample. The approach adopted here is the bootstrapped implementation of the post-stratified estimator-based area estimation and uncertainty estimation methodology formalized in Olofsson et al. (2013). This approach forms the basis of widely cited good-practice guidance for estimating area and assessing the accuracy of land cover and land cover change products (Olofsson et al., 2014; Olofsson, 2021), and has been broadly applied in the remote sensing community for error-adjusted area estimation and uncertainty reporting (e.g., Gallaun et al., 2015; FAO, 2016; Marconcini et al., 2020; Pickens et al., 2020; Dalagnol et al., 2023; De Keersmaecker et al., 2026; Pfoch et al., 2026).

Under this approach, a thematic map is treated as an auxiliary stratification layer, and the reference sample (independently interpreted from higher-quality information) provides unbiased information about the true class composition within each mapped stratum (Olofsson et al., 2013). The estimand is the true area (or area proportion) of a target reference class in the study region, while the mapped classes are used only to define strata and area weights.

Consider two binary built-up maps at t_1 and t_2 . Overlaying these two maps yields a built-up change map with four mutually exclusive classes that describe whether a location remained built-up, ceased to be built-up, became built-up, or remained non-built-up. For compact presentation, the four change classes are encoded as numeric labels and used consistently for both the map strata and the reference classes:

- 1 = Stable BUA (BUA at t_1 and t_2),
- 2 = BUA Loss (BUA at t_1 , non-BUA at t_2),
- 3 = BUA Gain (non-BUA at t_1 , BUA at t_2), and
- 4 = Stable Non-BUA (non-BUA at both times).

The built-up change map is treated as the stratification layer for design-based estimation, following the post-stratified, error-matrix⁸-based area estimation framework of Olofsson et al. (2013).

⁸Although confusion matrix is widely used in classification studies, the term error matrix is preferred here to remain consistent with Olofsson et al. (2013) and related land-cover area-estimation literature, where the matrix serves as the basis for map-error quantification, area adjustment, and uncertainty estimation, rather than simply as a summary of agreement between map and reference labels.

Let the four map-based change classes define $q = 4$ strata indexed by $i \in \{1, 2, 3, 4\}$. Denote the total study area by A_{tot} , the map area of stratum i by $A_{\text{mapped},i}$, and its corresponding weight by $W_i = A_{\text{mapped},i}/A_{\text{tot}}$. A probability sample is drawn from these strata. Each sampled location is interpreted using reference information at t_1 and t_2 , producing two reference labels (BUA or non-BUA) per sample. These two reference labels are then converted to a single reference change-class code $j \in \{1, 2, 3, 4\}$ using the same transition rule used to form the built-up change map: BUA at both times $\rightarrow 1$, BUA to non-BUA $\rightarrow 2$, non-BUA to BUA $\rightarrow 3$, and non-BUA at both times $\rightarrow 4$. The sample results are summarized in a 4×4 error matrix of counts n_{ij} , where rows (i) correspond to map-based change classes and columns (j) correspond to reference change classes (Table 4.1a).

Table 4.1: Error matrices of sample counts, n_{ij} , (a) and of estimated area proportions, $\hat{p}_{ij}$, (b). Rows correspond to map-based change classes (i) and columns correspond to reference change classes (j).

(a)	Class	1	2	3	4	Total
	1	n_{11}	n_{12}	n_{13}	n_{14}	$n_{1\cdot}$
	2	n_{21}	n_{22}	n_{23}	n_{24}	$n_{2\cdot}$
	3	n_{31}	n_{32}	n_{33}	n_{34}	$n_{3\cdot}$
	4	n_{41}	n_{42}	n_{43}	n_{44}	$n_{4\cdot}$
	Total	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot 3}$	$n_{\cdot 4}$	n

(b)	Class	1	2	3	4	Total
	1	$\hat{p}_{11}$	$\hat{p}_{12}$	$\hat{p}_{13}$	$\hat{p}_{14}$	$\hat{p}_{1\cdot}$
	2	$\hat{p}_{21}$	$\hat{p}_{22}$	$\hat{p}_{23}$	$\hat{p}_{24}$	$\hat{p}_{2\cdot}$
	3	$\hat{p}_{31}$	$\hat{p}_{32}$	$\hat{p}_{33}$	$\hat{p}_{34}$	$\hat{p}_{3\cdot}$
	4	$\hat{p}_{41}$	$\hat{p}_{42}$	$\hat{p}_{43}$	$\hat{p}_{44}$	$\hat{p}_{4\cdot}$
	Total	$\hat{p}_{\cdot 1}$	$\hat{p}_{\cdot 2}$	$\hat{p}_{\cdot 3}$	$\hat{p}_{\cdot 4}$	1

Following Olofsson et al. (2013), the error matrix is converted from sample counts (Table 4.1a) to a matrix of estimated area proportions (Table 4.1b). An unbiased estimator of the area proportion of area in cell (i, j) of the error matrix is

$$\hat{p}_{ij} = W_i \frac{n_{ij}}{n_{i\cdot}}. \quad (4.11)$$

The term $\frac{n_{ij}}{n_{i\cdot}}$ estimates the probability that a location mapped as class i belongs to reference class j . Summing down column j yields the estimated true area proportion of reference change class j ,

$$\hat{p}_{\cdot j} = \sum_{i=1}^q \hat{p}_{ij}, \quad (4.12)$$

and the corresponding error-adjusted area estimate is

$$\hat{A}_j = A_{\text{tot}} \times \hat{p}_{\cdot j}. \quad (4.13)$$

In simple terms, these steps treat the map as a useful but imperfect first approximation of reality. The reference sample, derived from higher-quality information, indicates how often the map's labels are correct within each mapped class. The error matrix summarizes these matches and mismatches, revealing where the map tends to assign too much area to a class (commission error) or fail to capture it (omission error). The post-stratified estimator (Eq. 4.11) then uses this information to correct map-based area totals: rather than relying on pixel counting, it estimates, for each mapped class i , the share of its area that truly belongs to reference class j , and scales that share by the mapped area weight. Summing these corrected contributions across strata yields an error-adjusted

estimate of the true area of each reference class, i.e., one that effectively adds back missed areas (omissions) and removes wrongly included areas (commissions).

Because the reference classes are change classes, BUA at the endpoint years can be obtained directly by summing the relevant change components. BUA at t_1 corresponds to Stable BUA (class 1) and BUA Loss (class 2). BUA at t_2 consists of Stable BUA (class 1) and BUA Gain (class 3). Therefore,

$$\widehat{\text{BUA}}_{t_1} = \hat{A}_1 + \hat{A}_2, \quad \widehat{\text{BUA}}_{t_2} = \hat{A}_1 + \hat{A}_3. \quad (4.14)$$

Uncertainty in the estimated area proportions arises from the finite reference sample used to estimate the error matrix and therefore reflects the sampling uncertainty of the classification-error-adjusted area estimates. Following Cochran (1977) and Olofsson et al. (2013), this uncertainty is quantified by the estimated standard error⁹ of a column total $\hat{p}_{\cdot j}$:

$$S(\hat{p}_{\cdot j}) = \sqrt{\sum_{i=1}^q W_i^2 \frac{\frac{n_{ij}}{n_{i\cdot}} \left(1 - \frac{n_{ij}}{n_{i\cdot}}\right)}{n_{i\cdot} - 1}}. \quad (4.15)$$

and the corresponding standard error in area units is

$$S(\hat{A}_j) = A_{\text{tot}} \times S(\hat{p}_{\cdot j}). \quad (4.16)$$

From these formulae, the uncertainty of endpoint BUAs can be indirectly estimated because $\widehat{\text{BUA}}_{t_1}$ and $\widehat{\text{BUA}}_{t_2}$ are sums of the estimated change-class areas (Eq. 4.14). In general, the variance of a sum includes both the variances of the individual terms and their covariance. Therefore, the standard errors for $\widehat{\text{BUA}}_{t_1}$ and $\widehat{\text{BUA}}_{t_2}$ can be expressed as

$$S(\widehat{\text{BUA}}_{t_1}) = \sqrt{\left[S(\hat{A}_1)\right]^2 + \left[S(\hat{A}_2)\right]^2 + 2 \text{Cov}(\hat{A}_1, \hat{A}_2)}. \quad (4.17)$$

$$S(\widehat{\text{BUA}}_{t_2}) = \sqrt{\left[S(\hat{A}_1)\right]^2 + \left[S(\hat{A}_3)\right]^2 + 2 \text{Cov}(\hat{A}_1, \hat{A}_3)}. \quad (4.18)$$

While the variance terms $\left[S(\hat{A}_j)\right]^2$ are obtained directly from the stratified estimator (Eqs. 4.15–4.16), the covariance terms are not routinely available from standard error-matrix reporting and are not straightforward to derive without additional assumptions about dependence between the component estimators. Setting these covariances to zero would be an ad hoc simplification that may understate or overstate uncertainty, particularly because the stable class ($\hat{A}_1$) appears in both endpoint sums and shares the same underlying sample. On the other hand, it is also possible to apply the same estimator independently to the built-up map at t_1 and the built-up map at t_2 , but this would typically require two separate samples (one per time point) and would still leave the dependence between $\widehat{\text{BUA}}_{t_1}$ and $\widehat{\text{BUA}}_{t_2}$ difficult to quantify.

⁹Here, the standard error may be interpreted as the estimated standard deviation of the area-proportion estimator, i.e., the expected sampling variability of the estimate across repeated samples drawn under the same design.

To address these limitations in a practical and assumption-light way, this study adopts a stratified bootstrap procedure applied to the change-map sample. The bootstrap generates an empirical distribution of change-class areas and, by extension, an empirical joint distribution of $\widehat{\text{BUA}}_{t_1}$ and $\widehat{\text{BUA}}_{t_2}$, which can be propagated directly into LCR and LCRPGR. The procedure, patterned after Gallaun et al. (2015), is described in the next section.

4.3.3 Bootstrap-based Uncertainty Estimation and Propagation

Bootstrapping is a resampling method that approximates the sampling distribution of an estimator by repeatedly re-drawing samples, with replacement, from the observed data (Efron & Tibshirani, 1986). In the stratified bootstrap used here, resampling is performed within each change-map stratum, preserving the sample allocation across strata. For each bootstrap replicate $b = 1, \dots, B$, the resampled observations are used to reconstruct a replicate error matrix and to calculate replicate area-proportion estimates using the same post-stratified estimator as defined in Eqs. 4.11–4.13. This procedure yields replicate change-class area estimates $\left\{ \hat{A}_j^{(b)} \right\}_{j=1}^q$. Endpoint built-up areas are then obtained replicate-wise using the same component sums as Eq. 4.14, producing paired endpoint estimates $\left\{ \widehat{\text{BUA}}_{t_1}^{(b)}, \widehat{\text{BUA}}_{t_2}^{(b)} \right\}_{b=1}^B$.

This bootstrap-based representation has two practical advantages for the present application. First, it provides uncertainty distributions for $\widehat{\text{BUA}}_{t_1}$ and $\widehat{\text{BUA}}_{t_2}$ without requiring a closed-form covariance term, because the dependence induced by shared sampling variability is reflected automatically in the paired bootstrap replicates. Second, it enables direct uncertainty propagation to the SDG Indicator 11.3.1 components through replicate-wise transformation. For each bootstrap replicate, $\text{LCR}^{(b)}$ is calculated from $\widehat{\text{BUA}}_{t_1}^{(b)}$ and $\widehat{\text{BUA}}_{t_2}^{(b)}$. The corresponding $\text{LCRPGR}^{(b)}$ can then be obtained by dividing $\text{LCR}^{(b)}$ by PGR, whether PGR is treated as a fixed quantity or represented by its own uncertainty distribution. The resulting sets of replicate values form empirical distributions. These distributions can be summarized using medians and percentile-based uncertainty intervals. They can also be used to report probability-based statements about LUE regime membership in line with the probabilistic LUE classification framework.

The same replicate-wise propagation logic also extends to secondary indicators derived from the endpoint built-up areas. Empirical distributions of BU_{UpC} and TC_{BUA} are calculated from the bootstrap endpoint estimates in the same manner as for the primary SDG Indicator 11.3.1 components. They can then be summarized using medians and percentile-based uncertainty intervals, and can be reported through probability statements when needed.

4.4 Illustrative Applications

The following sections present two applications to demonstrate the operational behavior of the uncertainty propagation and probabilistic LUE classification framework developed in Section 4.2, together with the EO-derived built-up area uncertainty-quantification pipeline introduced in Section 4.3. The first application uses a hypothetical City A experiment to fully illustrate the framework

under controlled conditions, treating all four inputs (BUA and Pop, both at t_1 and t_2) as uncertain. The second application uses Baguio City, Philippines, as a real-world case study to demonstrate how the Section 4.3 pipeline can be used to derive empirical BUA uncertainty distributions from EO-derived built-up data and reference samples, and how these distributions can then be propagated through the Section 4.2 framework to produce uncertainty-aware indicator estimates and probabilistic LUE classification outputs.

4.4.1 Application 1: Hypothetical Application Under Alternative Dependence Scenarios

To examine the influence of dependence structure under fully controlled conditions, the framework was applied here to an abstract spatial unit, hereafter referred to as City A, with explicitly specified input values and uncertainty parameters. This hypothetical setup allows defining the joint input uncertainty transparently and systematically varying the assumed dependence structure. In turn, it allows the resulting effects on the propagated distributions of LCR, PGR, and LCRPGR, as well as on probabilistic LUE regime membership, to be examined without the additional complexities of empirical data.

4.4.1.1 Baseline Specification and Marginal Input Uncertainty

City A was defined as a single spatial reporting unit observed at two reference epochs, t_1 and t_2 for a 10-year interval ($\Delta t = 10$). Baseline BUA and Pop values were specified to represent moderate built-up expansion and population change, and epistemic input uncertainty was introduced by treating BUA_{t_1} , BUA_{t_2} , Pop_{t_1} , and Pop_{t_2} as uncertain quantities with marginal distributions centered on their baseline means, consistent with the probabilistic input representation in Section 4.2.1. Marginal uncertainty was parameterized using fixed relative standard deviations applied symmetrically across epochs: 5% for BUA at both t_1 and t_2 , and 10% for Pop at both epochs, with standard deviations calculated as the corresponding proportions of the baseline means. All baseline inputs and marginal uncertainty parameters are summarized in Table 4.2. The deterministic baseline values of LCR, PGR, and LCRPGR are reported in Table 4.3. Pop uncertainty was intentionally set higher than BUA uncertainty to ensure that probabilistic regime switching is observable in this illustrative experiment.

4.4.1.2 Alternative Dependence Structures

Consistent with the dependence possibilities outlined in Section 4.2.1, namely full independence, within-variable temporal dependence without cross-variable dependence, and full joint dependence, three alternative dependence scenarios were specified for the hypothetical application, as summarized in Table 4.4. Here, the correlations refer to dependence in the stochastic input variables used for Monte Carlo sampling, i.e., how uncertainty in BUA and Pop co-varied across epochs and across variables, rather than empirical correlations between the unknown true BUA and Pop levels.

Scenario S1 adopted the fully independent factorization discussed in Section 4.2.1, treating all variables as mutually independent. Under this assumption, uncertainty at one epoch or for one variable provided no information about uncertainty in any other component. Scenario S2 corresponded to the within-variable temporal dependence case. Here, strong positive inter-epoch correlations of 0.8 were specified for both BUA and Pop. Scenario S3 represented the general joint dependence case. In addition to the strong temporal dependence within each variable, with a correlation of 0.8 for both BUA and Pop across epochs, moderate cross-variable dependence was imposed between BUA and Pop components, with cross-correlations set to 0.5. This configuration allowed uncertainty to co-move both across time and across variables, reflecting the possibility of shared inputs, constraints, or harmonization procedures that could link the two quantities.

Across all scenarios, the marginal specifications in Table 4.2 were held constant, and only the dependence structure varied.

Table 4.2: Baseline inputs and marginal uncertainty specification for City A ($\Delta t = 10$ years).

Variable	Units	Baseline mean (μ)	Relative Standard Deviation	Standard Deviation (σ)
BUA_{t_1}	km^2	53.0	0.05	2.650
BUA_{t_2}	km^2	61.5	0.05	3.075
Pop_{t_1}	persons	320,000	0.10	32,000
Pop_{t_2}	persons	330,000	0.10	33,000

Table 4.3: Deterministic baseline indicator values and regime assignment for City A.

Quantity	Value
LCR (% yr^{-1})	1.60
PGR (% yr^{-1})	0.31
LCRPGR	5.19
LUE Regime (Table 2.1)	Inefficient Expansion under Slower Population Growth (R16)

Table 4.4: Alternative dependence structures assumed among BUA and Pop at t_1 and t_2 in the hypothetical City A application. Here, $\rho(\cdot, \cdot)$ denotes the correlation coefficient between the indicated variables or variable components.

Scenario	$\rho(\text{BUA}_{t_1}, \text{BUA}_{t_2})$	$\rho(\text{Pop}_{t_1}, \text{Pop}_{t_2})$	ρ (BUA components, Pop components)
S1: Full independence	0	0	0
S2: Within-variable temporal dependence	0.8	0.8	0
S3: Temporal + cross-variable dependence	0.8	0.8	0.5

4.4.1.3 Joint Distribution Specification and Dependence Modeling

For the hypothetical application, the joint uncertainty in BUA and Pop was represented using a parametric joint distribution. Specifically, the four-dimensional input vector $\mathbf{v}$, with components BUA_{t_1} , BUA_{t_2} , Pop_{t_1} , and Pop_{t_2} , was assumed to follow a multivariate normal distribution, $\mathbf{v} \sim$

$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where the mean vector $\boldsymbol{\mu}$ contained the baseline values in Table 4.2, and $\boldsymbol{\Sigma}$ denoted the covariance matrix.

Under this assumption, dependence between the variables was encoded through a scenario-specific correlation matrix, denoted by $\mathbf{C}$,

$$\mathbf{C} = \begin{bmatrix} 1 & \rho(\text{BUA}_{t_1}, \text{BUA}_{t_2}) & \rho(\text{BUA}_{t_1}, \text{Pop}_{t_1}) & \rho(\text{BUA}_{t_1}, \text{Pop}_{t_2}) \\ \rho(\text{BUA}_{t_1}, \text{BUA}_{t_2}) & 1 & \rho(\text{BUA}_{t_2}, \text{Pop}_{t_1}) & \rho(\text{BUA}_{t_2}, \text{Pop}_{t_2}) \\ \rho(\text{BUA}_{t_1}, \text{Pop}_{t_1}) & \rho(\text{BUA}_{t_2}, \text{Pop}_{t_1}) & 1 & \rho(\text{Pop}_{t_1}, \text{Pop}_{t_2}) \\ \rho(\text{BUA}_{t_1}, \text{Pop}_{t_2}) & \rho(\text{BUA}_{t_2}, \text{Pop}_{t_2}) & \rho(\text{Pop}_{t_1}, \text{Pop}_{t_2}) & 1 \end{bmatrix}. \quad (4.19)$$

This correlation structure was combined with the marginal standard deviations to construct a covariance matrix. Let v_i denote the i -th component of $\mathbf{v}$, and let σ_i denote its marginal standard deviation. The covariance terms are then defined as $\text{Cov}(v_i, v_j) = \rho_{ij}\sigma_i\sigma_j$ for $i \neq j$, while the marginal variances satisfy $\text{Var}(v_i) = \sigma_i^2$. In compact matrix form, the covariance matrix was written as $\boldsymbol{\Sigma} = \mathbf{D}\mathbf{C}\mathbf{D}$, where $\mathbf{D} = \text{diag}(\sigma_{\text{BUA}_{t_1}}, \sigma_{\text{BUA}_{t_2}}, \sigma_{\text{Pop}_{t_1}}, \sigma_{\text{Pop}_{t_2}})$.

The multivariate normal specification was adopted for computational convenience, as it allowed systematic manipulation of the covariance structure while preserving fixed marginal uncertainty.

4.4.1.4 Monte Carlo Implementation

The Monte Carlo uncertainty propagation was executed separately for each dependence scenario in Python 3.10.11. For a given scenario, the scenario-specific correlation structure in Table 4.4 was combined with the fixed marginal standard deviations in Table 4.2 to define the covariance structure of the joint input vector $\mathbf{v}$ under the multivariate normal model described in Section 4.4.1.3.

Random sampling was carried out by first generating independent standard normal random numbers. These numbers were then transformed into correlated realizations using covariance factorization implemented in NumPy's multivariate normal sampling routine¹⁰, for example through matrix decomposition of the covariance matrix $\boldsymbol{\Sigma}$. Because the multivariate normal distribution is unbounded, rare non-physical negative values could occur. To preserve the non-negativity of BUA and Pop and to avoid numerical issues in logarithmic and ratio-based calculations, simulated inputs were constrained to be strictly positive by applying a small lower bound (1×10^{-10}) before indicator evaluation.

For each Monte Carlo realization, the SDG Indicator 11.3.1 components were calculated deterministically as described in Section 4.2.2. When calculating LCRPGR, realizations with PGR equal to zero, or numerically indistinguishable from zero under a tolerance of 1×10^{-10} , were explicitly flagged, and the corresponding LCRPGR value was treated as undefined. This safeguard avoided unstable ratios resulting from division by effectively zero values, while preserving these realizations for subsequent regime assignment under the “undefined LCRPGR” cases in Table 2.1. Each simulated triplet (LCR, PGR, LCRPGR), including cases with undefined LCRPGR, was then mapped to an LUE regime using the regime definitions in Table 2.1.

¹⁰https://numpy.org/doc/stable/reference/random/generated/numpy.random.multivariate_normal.html (accessed February 23, 2026).

For reporting and visualization, realizations classified as R4–R6 were grouped and reported as *Efficient Built-up Reduction under Population Growth*, while realizations classified as R12–R14 were grouped and reported as *Inefficient Built-up Expansion under Population Decline*. All other regimes were retained as individual categories. This grouping was adopted for practical interpretation. Although R4–R6 and R12–R14 are distinguished in the full classification system, the differences among the regimes within each group were not considered especially useful for the purposes of this study. From a planning or policy perspective, it is often more meaningful to identify the broader development pattern itself—such as built-up reduction despite population growth, or built-up expansion despite population decline—than to further distinguish its subtypes. In other words, once the realizations already reflect the same overall development pattern, distinguishing the individual regimes within that group adds complexity but contributes little additional interpretive value. The grouped reporting, therefore, retains the policy-relevant message while making the probabilistic results easier to communicate.

For each dependence scenario, $N = 444,130$ realizations were drawn from the corresponding input distribution. This choice was justified through a convergence or sample-size sensitivity analysis in which increasing Monte Carlo sizes were evaluated using a geometric progression to avoid sensitivity to uneven step sizes. Specifically, candidate sample sizes were defined as $N_{i+1} = \lceil rN_i \rceil$ with a fixed multiplier $r = 1.25$, supplemented by a small set of initial values (e.g., $N = 100, 500, 1000, 5000, 10000$) and capped at $N_{\max} = 500,000$. For each candidate N , the median and 95% interval (2.5th and 97.5th percentiles) of LCR, PGR, and LCRPGR were calculated.

Convergence was assessed using a plateau criterion based on successive increases in N . Specifically, for each statistic $s \in \{\tilde{x}, q_{0.025}, q_{0.975}\}$, where $q_{0.025}$ and $q_{0.975}$ denote the 2.5th and 97.5th percentiles, respectively, stability was evaluated via the successive change $\Delta s = |s_{N_{i+1}} - s_{N_i}|$. The minimum N was defined as the smallest value for which all reported statistics for all three indicators satisfied predefined tolerances for two consecutive sample-size increments, indicating that further increases in N produced negligible changes in the reported summaries. Tolerances were set to require that the relative change in the median was less than 0.5% and that the 2.5th/97.5th percentiles changed by less than 1%. Under this criterion, the most stringent requirement arose for LCRPGR under scenario S1, full independence, yielding $N = 444,130$. The corresponding plateau values for S2 and S3 were $N = 227,394$ and $N = 93,140$, respectively. For consistent precision across all scenarios, $N = 444,130$ was adopted uniformly for all simulations.

To ensure comparability across candidate N , simulations were generated with a fixed random seed so that results for smaller N corresponded to the first N realizations, i.e., a nested subset, of the same underlying simulation run. Despite this relatively large N , the simulations were computationally inexpensive in practice; executing all three dependence scenarios, including indicator calculation and regime assignment, required less than one minute on the laptop used for this experiment.¹¹

¹¹Runtime measured for the City A experiment using Python 3.10.11 on an Acer Predator PT515-51 laptop running Windows 11 Pro 64-bit with an Intel® Core™ i7-8750H CPU @ 2.20 GHz (6 cores / 12 threads) and 16 GB RAM. Actual runtimes may vary with hardware, operating system, and implementation details.

The resulting sets of simulated indicators and regime labels formed empirical distributions, which were summarized in the next subsection using distribution plots, summary statistics, and regime probabilities, calculated as the relative frequency of realizations assigned to each regime.

4.4.1.5 Results and Discussion

4.4.1.5.1 Distributional Behavior of LCR, PGR, and LCRPGR Across Dependence Scenarios

Figure 4.3 summarizes the Monte Carlo-propagated distributions of LCR, PGR, and LCRPGR for City A under three dependence scenarios. Across all scenarios, LCR and PGR exhibited unimodal distributions centered near their expected growth rates, while LCRPGR showed stronger non-Gaussian behavior, reflecting its ratio structure.

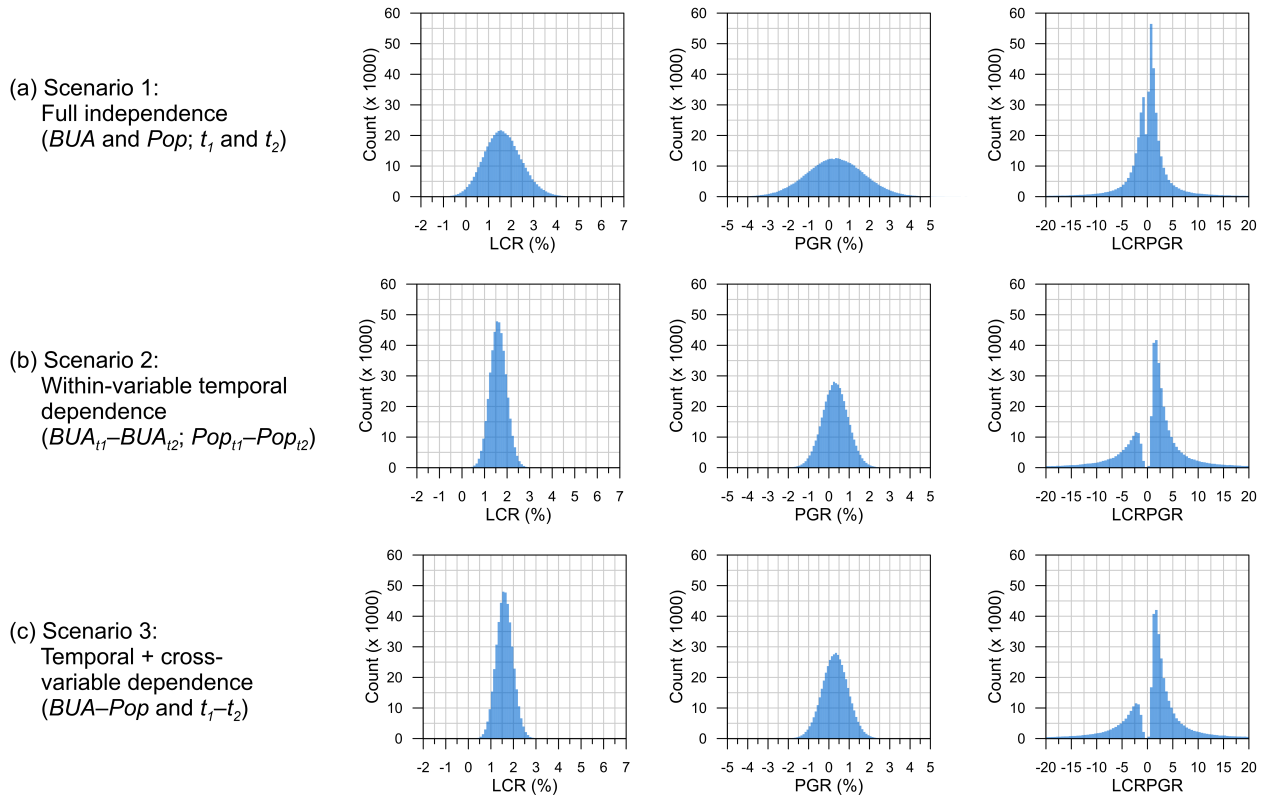


Figure 4.3: Monte Carlo-propagated distributions of LCR, PGR, and LCRPGR for City A under three dependence scenarios. For interpretability, LCRPGR values were truncated to the interval $[-20, 20]$, meaning the plotted histograms represent the central mass of the ratio distribution rather than its full tail extent.

Under S1 (full independence), the LCR distribution is comparatively wider, indicating larger dispersion in land consumption estimates when uncertainties in BUA at t_1 and t_2 are allowed to vary independently (Figure 4.3a). This pattern is consistent with the idea that, without temporal coupling in the error terms, uncertainty in the change term ($BUA_{t2} - BUA_{t1}$) can be larger, which carries through to LCR. The PGR distribution likewise appears relatively broad, reflecting independent variation in Pop uncertainty at t_1 and t_2 . For LCRPGR, the histogram shows a sharp central peak with extended tails; this is in line with ratio behavior, where simulations with PGR

values close to zero can generate very large positive or negative LCRPGR values (Angel et al., 2024; Zhong et al., 2025b).

Under S2 (within-variable temporal dependence), the LCR distribution appears noticeably tighter than in S1, and the PGR distribution shows a similar narrowing (Figure 4.3b). This pattern suggests that when uncertainty components for BUA at t_1 and t_2 are positively correlated, the two epochs tend to co-vary more consistently in the simulations, so the change between them varies less from one draw to the next. In contrast, the LCRPGR histogram appears more uneven and less symmetric, with a visible dip around zero. One way to interpret this is that ratios near zero would require LCR to be very small relative to PGR; however, because LCR is more concentrated away from zero in this scenario, fewer simulations produce LCRPGR values close to zero. At the same time, occasional simulations with PGR close to zero can still create very large positive or negative ratios, which helps explain the heavier tails and the relatively lower counts in the midrange.

Under S3 (temporal plus cross-variable dependence), the LCR and PGR distributions are very similar to those in S2, remaining narrow and clearly unimodal (Figure 4.3c). This similarity suggests that most of the variance reduction is already achieved once strong within-variable temporal dependence is introduced, and that adding cross-variable dependence in the uncertainty terms does not noticeably change the marginal spread of LCR and PGR for this case. The LCRPGR distribution also closely resembles that of Scenario S2, including the low density around zero.

Overall, the distributions demonstrate that dependence assumptions primarily control the magnitude of uncertainty for LCR and PGR, with temporal dependence reducing variance. At the same time, LCRPGR remains inherently more unstable and non-Gaussian because it is a ratio of two uncertain growth rates. This distributional behavior motivates reporting LCRPGR using robust quantiles and probabilistic regime membership rather than relying on single-point estimates.

4.4.1.5.2 Summary Statistics and Deterministic Agreement Table 4.5 complements the distributional patterns in Section 4.4.1.5.1 by summarizing the Monte Carlo medians and 95% intervals (2.5th–97.5th percentiles) of LCR, PGR, and LCRPGR¹² across the three dependence scenarios, alongside the deterministic values reported earlier in Table 4.3. For LCR and PGR, the Monte Carlo medians are nearly identical across scenarios and remain close to the deterministic values, with LCR approximately 1.60–1.61% and PGR approximately 0.31%. This suggests that the dependence assumptions mainly influence uncertainty spread rather than shifting the distributional center. This result is consistent with the scenario design, in which marginal uncertainty magnitudes are held constant and only the dependence structure is varied, as described in Section 4.4.1.2.

Scenario differences are most evident in the widths of the 95% intervals. Compared with full independence (S1), the 95% intervals for both LCR and PGR are noticeably narrower under S2 and S3, reflecting the stabilizing effect of strong inter-epoch dependence ($\rho = 0.8$) within each variable. The near-identical summaries for S2 and S3 for LCR and PGR mirror the histogram similarities noted in Section 4.4.1.5.1 and follow from the fact that cross-variable dependence in S3 does not

¹²The median LCRPGR was calculated directly from the simulated LCRPGR distribution. It is therefore not necessarily equal to the ratio of the median LCR and median PGR.

Table 4.5: Monte Carlo-propagated summaries of LCR, PGR, and LCRPGR for City A under three dependence scenarios (S1–S3), with deterministic values from Table 4.3 shown for comparison. Entries report the Monte Carlo median, with the 95% interval (2.5th–97.5th percentiles) shown in brackets. For LCRPGR, the median was calculated directly from the simulated LCRPGR distribution after applying the ratio transformation to each Monte Carlo realization; therefore, it is not necessarily equal to the ratio of the median LCR and median PGR.

Indicator	Scenario 1	Scenario 2	Scenario 3	Deterministic
LCR (% yr ⁻¹)	1.61 [0.10–3.34]	1.61 [0.90–2.35]	1.60 [0.90–2.35]	1.60
PGR (% yr ⁻¹)	0.31 [–2.50–3.12]	0.31 [–0.96–1.57]	0.31 [–0.96–1.57]	0.31
LCRPGR	0.52 [–18.02–18.31]	1.77 [–35.62–36.01]	1.78 [–35.17–36.70]	5.19

modify the within-variable temporal structure that drives the joint behavior of (BUA_{t_1}, BUA_{t_2}) and (Pop_{t_1}, Pop_{t_2}) , from which LCR and PGR are calculated.

For LCRPGR, Table 4.5 shows much stronger tail effects and larger departures from the deterministic value of 5.19 than for the component rates, with propagated medians markedly lower, namely 0.52 in S1 and approximately 1.77–1.78 in S2–S3, and wide 95% intervals across all scenarios. The LCRPGR 95% intervals are wider under S2 and S3 than under S1. While temporal dependence narrows the marginal uncertainty of LCR and PGR, it also concentrates PGR closer to zero. When PGR is very small, dividing LCR by it can produce extreme positive or negative LCRPGR values, which inflate the tail percentiles and widen the interval. Consistent with the histogram results, the LCRPGR summaries for S2 and S3 are again very similar, indicating that moderate cross-variable dependence ($\rho = 0.5$) adds little beyond the dominant effect of within-variable temporal dependence ($\rho = 0.8$) on the marginal distribution of LCRPGR. These results reinforce that dependence assumptions made in this hypothetical experiment matter most through their effects on uncertainty widths and tail behavior, with the largest practical implications arising for LCRPGR rather than for LCR and PGR considered separately.

4.4.1.5.3 Probabilistic LUE Classification of City A Figure 4.4 summarizes the probabilistic LUE classification by assigning each Monte Carlo draw of LCR, PGR, and LCRPGR to an LUE regime. Under full independence (S1), the most likely outcome for City A is *Inefficient Built-up Expansion under Population Decline* (combined R12–R14), but it accounts for less than half of the simulations. Instead, the probabilities are shared across several regimes: the combined R12–R14 group has the largest share at 40.65%, followed closely by R16 (*Inefficient Built-up Expansion under Slower Population Growth*) at 37.73%, and R2 (*Efficient Built-up Expansion under Faster Population Growth*) at 19.83%. The remaining regimes occur only rarely, with the combined R4–R6 group accounting for 1.05%, R8 for 0.12%, and R10 for 0.62%. This wider spread in regime outcomes is consistent with the earlier S1 distribution results, where LCR and PGR show the broadest ranges (Figure 4.3a), allowing more draws to map to different regimes.

With within-variable temporal dependence (S2), the probabilistic classification becomes much more concentrated. City A is most likely *Inefficient Built-up Expansion under Slower Population*

Growth (R16), accounting for 64.54% of simulations, while *Inefficient Built-up Expansion under Population Decline* (combined R12–R14) remains the clear second-most-common outcome at 31.54%. *Efficient Built-up Expansion under Faster Population Growth* (R2) drops to 3.91%, and the remaining regimes are close to 0%. This concentration is consistent with the earlier S2 results, where the LCR and PGR distributions are narrower than in S1 (Figure 4.3a–b).

Under S3 (temporal plus cross-variable dependence), the classification is almost unchanged from S2. R16 remains dominant at 64.66%, followed by the combined R12–R14 group at 31.43%, with R2 at 3.92% and all other regimes near 0%. In other words, introducing moderate cross-variable dependence ($\rho = 0.5$) does not materially change the regime probabilities for City A compared with S2. This result suggests that, in this hypothetical setup, the regime outcomes are driven mainly by the strong temporal dependence within BUA and within Pop, while the added BUA–Pop co-movement has only a small effect.

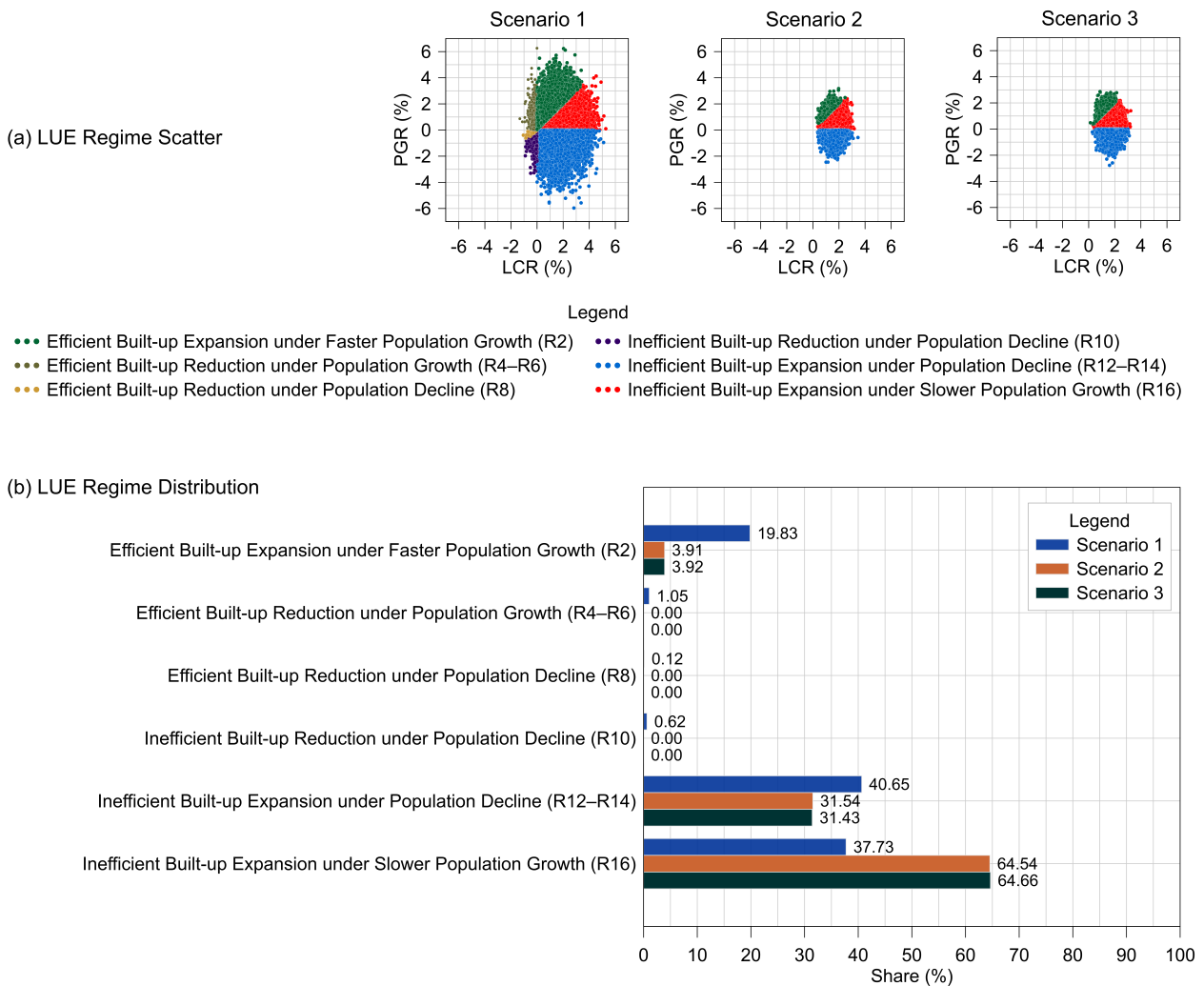


Figure 4.4: Probabilistic LUE regime classification results for City A under three dependence scenarios. Panel (a) shows the Monte Carlo draws in the (LCR, PGR) plane, colored by the assigned LUE regime. Panel (b) shows the share of simulations, in percent, assigned to each regime for each scenario.

4.4.1.6 Application 1 Summary

The hypothetical City A experiment shows how the proposed framework translates input data uncertainty into distributions of LCR, PGR, and LCRPGR, as well as into probabilities across LUE regimes. With marginal uncertainty held constant, changing only the dependence structure mainly affects the spread of the outputs. Under full independence (S1), LCR and PGR exhibited the widest 95% intervals, and the LUE outcomes were spread across several regimes. By contrast, introducing strong within-variable temporal dependence (S2) tightened the LCR and PGR distributions and concentrated the regime probabilities, with R16 and the combined R12–R14 group dominating. When moderate cross-variable dependence was also introduced (S3; $\rho = 0.5$), the results remained very similar to S2. This suggests that, for this configuration, regime assignment was influenced mainly by within-variable temporal dependence. By comparison, additional BUA–Pop co-movement in the uncertainty structure had only a limited effect. Across scenarios, the experiment also highlights that LCRPGR remains the most sensitive quantity, with greater spread and heavier tails than LCR and PGR. This reinforces the value of reporting robust distributional summaries, such as the median, 95% intervals, and regime probabilities, rather than relying only on deterministic point estimates.

These results help contextualize the City A experiment within broader concerns about the robustness of SDG Indicator 11.3.1 derived from EO-based built-up and population datasets. Recent local, regional, and global assessments have shown that LCRPGR is highly sensitive to small discrepancies in its inputs, often leading to unstable or even contradictory interpretations when alternative datasets are used (Cardenas-Ritzert et al., 2024b; Wang et al., 2020; Zhong et al., 2025a). Most responses to this problem have emphasized robustness-oriented strategies, such as alternative indicators, city-size filtering, or robust trend estimation (Zhong et al., 2025a), which address uncertainty only indirectly through analytical design. In contrast, the present study demonstrates the value of explicitly modeling, propagating, and interpreting uncertainty, thereby providing a more transparent and policy-relevant basis for understanding SDG Indicator 11.3.1 outputs and their implications for sustainable urban development.

From a practical reporting perspective, these results suggest two simple takeaways. First, when input uncertainty in BUA and Pop is known, or can be reasonably specified, it should be included in SDG Indicator 11.3.1 calculations, since dependence assumptions, especially temporal dependence, can change uncertainty bounds and regime probabilities. Second, because LCRPGR can exhibit heavy-tail behavior even when LCR and PGR appear stable, decision-oriented reporting can benefit from probability statements, such as the probability of R16 versus the combined R12–R14 group, and interval summaries, which more clearly describe the likelihood of alternative outcomes than a single regime label.

4.4.2 Application 2: Uncertainty-Aware SDG Indicator 11.3.1 Monitoring from EO-Derived Built-Up Data

This application demonstrates the uncertainty propagation and probabilistic LUE classification framework under empirical monitoring conditions. Specifically, it applies the pipeline described in Section 4.3 to derive uncertainty in EO-derived BUA from a reference sample and then propagate

that uncertainty through the SDG Indicator 11.3.1 calculations. Pop is treated as fixed in this implementation because uncertainty information for the census-based population inputs is not available. The application thus illustrates a practical case of retrospective monitoring in which uncertainty can be incorporated into EO-derived BUA, even when Pop must remain deterministic.

4.4.2.1 Study Area Description

Baguio City (Figure 4.5a) is a highly urbanized highland city in northern Luzon, Philippines. Located on the Cordillera Central mountain range, the city serves as the principal regional service center of the Cordillera region, supporting education, tourism, commerce, and government functions (Estoque & Murayama, 2013). Its rugged terrain, with elevations ranging from 950 to 1650 m above sea level, and limited land area, approximately 57.5 km², impose strong constraints on urban development, making Baguio City a recurring case in remote-sensing studies of urban change in environmentally sensitive mountain settings (Estoque & Murayama, 2011, 2012).

Recent analyses continue to document pronounced urbanization dynamics, including substantial built-up expansion and associated environmental signals: built-up reportedly increased from about 27% to 45% of the city's land area between 2003 and 2011 through conversions from vegetation and other covers (Macaspac et al., 2024), while multi-sensor assessments show built-up growth concentrated around the urban core alongside expanding surface-thermal hotspots (Baloloy et al., 2019). Baguio was selected for this real-world application because its combination of strong development pressure, terrain-constrained expansion, and compact administrative area provides a demanding yet realistic context for monitoring SDG Indicator 11.3.1, where small differences in mapped built-up extent—and in how uncertainty is handled—can meaningfully affect LCR and LCRPGR interpretation.

4.4.2.2 Temporal Scope and Built-up and Population Datasets

This application focuses on the 2010–2020 period, which is sufficiently long to capture meaningful changes in built-up extent. Population totals for Baguio City were obtained from the Philippine Statistics Authority (PSA) censuses, with 318,676 residents in 2010 and 366,358 in 2020.¹³ These endpoints imply a PGR of 1.39% per year over the decade. In the subsequent analysis, the population totals—and the resulting PGR—are treated as fixed, meaning that no population uncertainty is assumed.

Built-up data were derived from the 30 m GLCLUC product, which was also included among the EO-derived datasets examined in the Chapter 3 sensitivity analysis (Table 3.1). GLCLUC was selected for this application because, among the discrete built-up datasets considered, it exhibited comparatively high commission error in mapping BUA (Table 3.2). Moreover, the sensitivity analysis results in Chapter 3 showed that in several cities, the use of GLCLUC produced large upward shifts in LCRPGR and reversed the LUE class relative to the baseline. This makes it especially

¹³<https://psa.gov.ph/content/highlights-cordillera-administrative-region-car-population-2020-census-population-and> (accessed January 24, 2026).

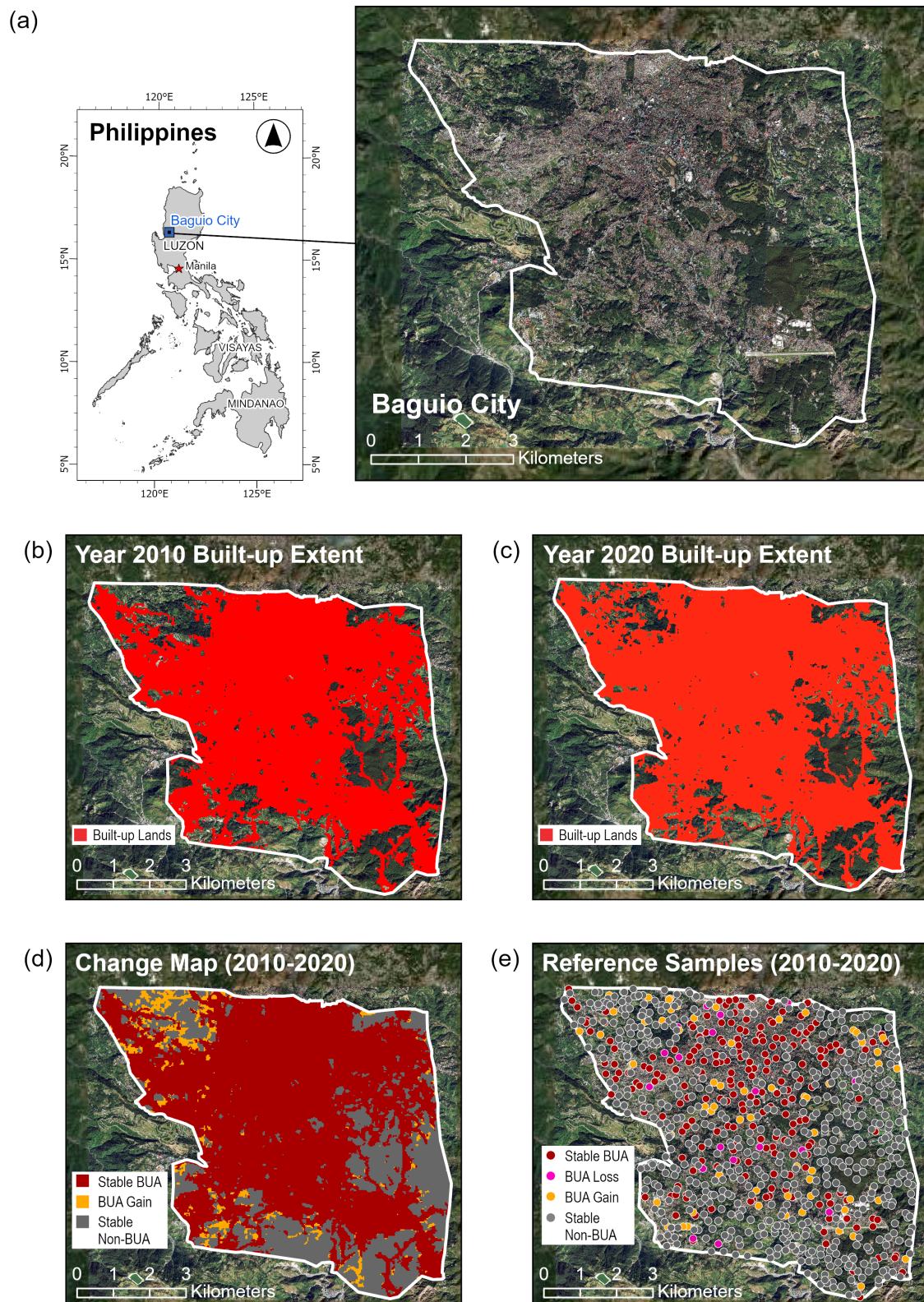


Figure 4.5: Study area and GLCLUC-derived built-up products for Baguio City: (a) location and administrative boundary, (b) built-up extent in 2010, (c) built-up extent in 2020, (d) 2010–2020 change map, and (e) stratified reference samples overlaid on the change map. High-resolution imagery source: Esri, Vantor.

Table 4.6: Mapped 2010–2020 change-map strata for Baguio City based on the GLCLUC dataset, showing pixel counts, stratum weights, and final validation sample allocation used for stratified reference sampling.

Class	Class Name	Number of Pixels	Weight (W_i)	Number of Samples (n_i)
1	Stable BUA	44,652	0.6917	534
2	BUA Loss	0	0.0000	0
3	BUA Gain	2,901	0.0449	100
4	Stable Non-BUA	16,999	0.2633	203
Total		64,552	1	837

suitable for demonstrating the central motivation of this chapter: that classification error in EO-derived built-up products can substantially affect BUA estimates and, in turn, propagate into the derived SDG Indicator 11.3.1 components. Using GLCLUC, therefore, provides an informative test case for showing how uncertainty in built-up classification can alter not only the magnitude of LCR and LCRPGR, but also their interpretation.

In GLCLUC, the built-up class is defined as built-up land, including man-made land surfaces associated with infrastructure and commercial and residential land uses. Accordingly, all BUA estimates used in this application, and the resulting LCR and LCRPGR values, are interpreted strictly according to this GLCLUC built-up definition.

The GLCLUC rasters for the endpoint years 2010 and 2020 were downloaded from the GLAD website.¹⁴ The rasters are provided in a geographic coordinate reference system (WGS 1984); they were reprojected to the same Philippine Albers Equal Area projection used in the sensitivity analysis (Section 3.3.4) and clipped to the Baguio City administrative boundary. Binary built-up/non-built-up maps were then produced (Figure 4.5b–c), followed by a 2010–2020 change map (Figure 4.5d). Raster statistics indicate that the clipped study area contains $N = 64,552$ pixels, corresponding to approximately 58.1 km² of ground area, which is slightly larger than the administrative area reported in Section 4.4.2.1.

4.4.2.3 Sampling Design and Sample Size Planning

Stratified random sampling was employed in accordance with good practice guidelines (Olofsson et al., 2014), with pixels as the sampling unit. Sampling was stratified by the 2010–2020 change-map classes, using the mapped pixel counts and stratum weights W_i summarized in Table 4.6. Initial allocation followed proportional allocation ($n_i \propto W_i$).

For this application, a target margin of error (MOE) of $\pm 5\%$ at the 95% level was adopted for the BUA proportion at each epoch. A stricter target, such as $\pm 1\text{--}2\%$, would require substantially larger reference samples and would shift the emphasis of this application toward intensive reference data collection. The $\pm 5\%$ target provides a practical balance; it keeps sample sizes feasible while still

¹⁴<https://storage.googleapis.com/earthenginepartners-hansen/GLCLU2000-2020/v2/download.html> (accessed January 26, 2026).

producing uncertainty intervals that clearly show how built-up mapping errors translate into SDG Indicator 11.3.1 uncertainty. In area units, the MOE corresponds to $0.05 \times 58.1 \text{ km}^2 \approx \pm 2.9 \text{ km}^2$. To link this MOE to Eqs. 4.17–4.18, the target standard error was set as $S_{\text{target}} = \text{MOE}/1.96 = 0.05/1.96 \approx 0.0255$. The same target was applied to both $\widehat{\text{BUA}}_{t_1}$ and $\widehat{\text{BUA}}_{t_2}$. The total sample size n was chosen so that the epoch-level standard errors would satisfy $S(\widehat{\text{BUA}}_{t_1}) < S_{\text{target}}$ and $S(\widehat{\text{BUA}}_{t_2}) < S_{\text{target}}$ when evaluated through Eqs. 4.17–4.18. For planning purposes, the covariance terms in these equations were set to zero. Because the within-stratum agreement rate between the map and the reference labels was not known in advance for this local case, a conservative value of $n_{ij}/n_i = 0.5$ was used in the planning calculations, which yields the largest possible sample size and helps avoid underestimating n .

From these calculations, a minimum of $n = 772$ samples was determined. The initial proportional allocation produced a small sample for the BUA Gain stratum ($n_3 = 35$). Following common good-practice recommendations for stratified accuracy assessment (Olofsson et al., 2014), this size was increased to $n_3 = 100$ to improve the stability of stratum-specific error estimates for this class. The remaining samples were distributed across the other strata while keeping the design close to proportional allocation (Table 4.6). Because the mapped change product did not contain any BUA Loss pixels for 2010–2020, no samples were assigned to that stratum. The final reference sample size was $n = 837$.

Within each stratum, sample pixels were selected at random using a spatially balanced approach to reduce clustering and improve spatial coverage (Theobald et al., 2007). An eligibility mask ensured that all eligible pixels within a stratum had the same chance of being selected. The selected pixels were then converted to polygons to support reference labeling.

4.4.2.4 Reference Data and Sample Labeling

High-resolution basemap imagery from Esri World Imagery Wayback¹⁵ was used to manually label the validation samples corresponding to 30 m GLCLUC pixels for the 2010 and 2020 epochs. The temporally closest available Wayback imagery for the 2010 epoch was WorldView-2 (WV02) acquired on 15 December 2009, with an approximately 0.5 m source pixel size. The closest imagery for the 2020 epoch was WV02 acquired on 31 December 2019, also with an approximately 0.5 m source pixel size. Because the reference images precede the nominal map years by only a few weeks, the analysis assumed that land-cover change within these short offsets was negligible at the 30 m scale for dominant-cover labeling.

To ensure that inter-epoch misregistration in the Wayback imagery would not affect dominant-cover labeling of 30 m sample footprints, positional agreement between the 2009 and 2019 reference images was assessed using 13 well-distributed check points on stable, clearly interpretable features visible in both epochs, such as major road intersections and bridge approaches. Check-point coordinates were transformed to the Philippine Albers Equal Area CRS, as defined in Section 3.3.4, and residuals were calculated as dX and dY differences between epochs. The resulting errors were

¹⁵<https://livingatlas.arcgis.com/wayback/#mapCenter=120.59460%2C16.41270%2C13&mode=explore&active=64001> (accessed January 28, 2026).

$RMSE_X = 1.91$ m and $RMSE_Y = 1.57$ m, corresponding to a horizontal $RMSE_{2D} = 2.47$ m, with a maximum residual of 4.35 m. Because these offsets were small relative to a 30 m sample footprint, the imagery was deemed suitable for reference labeling without further geometric adjustment.

Reference labeling was carried out by overlaying the sample polygon layers on the corresponding Wayback imagery. Labels were assigned using a dominant-cover rule consistent with the original GLCLUC “built-up land” definition. Specifically, a sample was labeled built-up if built-up land features, including buildings, roads, and other man-made structures, predominantly occupied the 30 m footprint. A sample was labeled non-built-up otherwise, including cases dominated by vegetation, bare soil, water, and other non-built surfaces. Where mixed content occurred within a footprint, the dominant-cover rule was applied to ensure consistent class assignment across samples and epochs. Samples affected by strong shadowing, haze, partial occlusion, or boundary ambiguity were re-examined at higher zoom using additional local context in the Wayback imagery. When the Wayback view remained unclear, temporally close imagery in Google Earth Pro was used as a supplementary reference to support consistent interpretation.

Because reference interpretation produced two epoch-specific labels per sample, namely built-up/non-built-up labels for 2010 and for 2020, these paired labels were subsequently recoded into a single reference change-class label using the same transition rules used to create the mapped change strata: stable BUA, BUA loss, BUA gain, and stable non-BUA (Figure 4.5e).

4.4.2.5 Implementation of Error-Adjusted Area Estimation and Stratified Bootstrap Uncertainty Propagation

The reference samples with their change-class labels were reconciled with their corresponding mapped change classes and were then processed using the error-adjusted area estimation and stratified bootstrap procedures described in Section 4.3. The full workflow, including error matrix construction, area-adjusted estimation of change components, derivation of $\widehat{BUA}_{t_1}$ and $\widehat{BUA}_{t_2}$, and replicate-wise propagation to LCR, LCRPGR, and secondary indicators, was implemented in Python 3.10 using a stratified bootstrap with $B = 5000$ replicates. In all LCRPGR calculations, a fixed value of $PGR = 1.39\%$ was used, as discussed in Section 4.4.2.2. A convergence check, following the same approach used in Section 4.4.1.4 and based on the stability of the bootstrap median and the 95% percentile interval across increasing B , suggested a plateau around $B = 3125$; however, $B = 5000$ was retained as a conservative choice to ensure stable uncertainty summaries.

The bootstrap outputs yielded empirical distributions for BUA, LCR, LCRPGR, BU_{up}C, and TC_{BUA}, which were summarized using the median and the 95% percentile interval. In addition, each replicate was assigned an LUE regime following the probabilistic LUE classification procedures described in Section 4.2.5. As in Section 4.4.1.4, outcomes falling within R4–R6 were consolidated under *Efficient Built-up Reduction under Population Growth*, while those within R12–R14 were combined as *Inefficient Built-up Expansion under Population Decline*; the remaining regimes were kept distinct. The city’s reported regime was taken as the dominant, or most frequent, regime, together with its associated membership probability. Finally, these uncertainty-aware results were

Table 4.7: Comparison of SDG Indicator 11.3.1 metrics for Baguio City (2010–2020) under three reporting approaches: (i) map-based pixel counting using GLCLUC mapped built-up area (treated as fixed), (ii) error-adjusted point estimates from post-stratified area estimation based on the reference sample, and (iii) uncertainty-aware results reported as bootstrap median and 95% percentile interval from stratified resampling of the reference sample. BUA and derived metrics follow the GLCLUC “built-up land” definition; population and PGR are from PSA census and treated as fixed.

Metric	Map-based (pixel count; fixed BUA)	Error-adjusted point estimate (post- stratified)	Uncertainty-aware (bootstrap median [95% percentile interval])
BUA (km ²), 2010	40.19	17.01	17.01 [15.35–18.74]
BUA (km ²), 2020	42.80	19.61	19.60 [17.90–21.36]
TCBUA (%)	6.50	15.29	15.21 [8.26–23.15]
BU _p C (m ² per person), 2010	126.11	53.37	53.37 [48.18–58.80]
BU _p C (m ² per person), 2020	116.82	53.52	53.51 [48.87–58.31]
LCR (% yr ⁻¹)	0.65	1.53	1.52 [0.83–2.32]
PGR (% yr ⁻¹)	1.39	1.39	1.39
LCRPGR	0.47	1.10	1.09 [0.59–1.66]
LUE Regime	Efficient Built-up Expansion under Faster Population Growth (R2)	Inefficient Built-up Expansion under Slower Population Growth (R16)	Inefficient Built-up Expansion under Slower Population Growth (R16)
Dominant Regime Share of Simulations (%)	–	–	63.08%

compared with their deterministic counterparts, calculated directly from the mapped BUA values obtained by pixel counting in the built-up maps.

4.4.2.6 Results and Discussion

4.4.2.6.1 Conventional SDG Indicator 11.3.1 Results from Map-Based Pixel Counting (Fixed BUA) Using the map-based pixel-counting totals from GLCLUC, BUA in Baguio City increased from 40.19 km² in 2010 to 42.80 km² in 2020 (Table 4.7), corresponding to a total built-up area increase (TCBUA) of 6.5% over the decade. Within the study area, built-up land accounted for approximately 69% of the city’s area in 2010 and approximately 74% in 2020, indicating that Baguio City was already highly built-up in both epochs.

4.4.2.6.2 Error-Adjusted SDG Indicator 11.3.1 Results from Post-stratified Estimation (Full Sample) Table 4.8 summarizes the full-sample change-based error matrix and the corre-

sponding post-stratified area-proportion matrix used to derive error-adjusted change-component areas. Using the area proportion matrix as a basis for the accuracy assessment of the change map (Olofsson et al., 2014), the error-adjusted overall accuracy was 54%, with strong class-to-class differences.

Table 4.8: Error matrices for the Baguio City (2010–2020) GLCLUC built-up change map: (a) sample count matrix n_{ij} from the full reference sample and (b) corresponding estimated area-proportion matrix $\hat{p}_{ij}$ from the post-stratified estimator. Rows represent mapped change strata and columns represent reference change classes. Class codes: 1 = Stable BUA, 2 = BUA Loss, 3 = BUA Gain, 4 = Stable Non-BUA. For readability, $\hat{p}_{ij}$ values are rounded to two decimal places in the table; all subsequent area and indicator calculations used the unrounded values.

(a)	Class	1	2	3	4	Total
	1	208	18	47	261	534
	2	0	0	0	0	0
	3	0	0	16	84	100
	4	0	0	0	203	203
	Total	208	18	63	548	837

(b)	Class	1	2	3	4	Total
	1	0.27	0.02	0.06	0.34	0.69
	2	0.00	0.00	0.00	0.00	0.00
	3	0.00	0.00	0.01	0.04	0.05
	4	0.00	0.00	0.00	0.26	0.26
	Total	0.27	0.02	0.07	0.64	1.00

Using the error-adjusted point estimates, BUA was 17.01 km² in 2010 and 19.61 km² in 2020 (Table 4.7), corresponding to approximately 29% and 34% of the 58.10 km² study area, respectively. Relative to these error-adjusted estimates, the map-based totals overestimated BUA by about 136% in 2010 and 118% in 2020. BU_PC was substantially lower than its map-based counterpart in both years, from about 53 m²/person in 2010 to about 54 m²/person in 2020 (Table 4.7). Moreover, the corrected decade change was larger than suggested by pixel counting: TC_BUA increased to 15.3%, yielding an LCR of 1.53% per year. With PGR fixed at 1.39% per year, the corresponding LCRPGR became 1.10, which shifted the city’s LUE regime classification to *Inefficient Built-up Expansion under Slower Population Growth* (R16).

4.4.2.6.3 Uncertainty-Aware SDG Indicator 11.3.1 Results (Bootstrap-Based) Figure 4.6 presents uncertainty-aware results as empirical distributions obtained from the stratified bootstrap. The endpoint built-up area estimates (Figure 4.6a–b) form compact, unimodal distributions, and these uncertainties propagate smoothly into LCR and LCRPGR (Figure 4.6c–d). The LCRPGR distribution spans both sides of the unity threshold, indicating that the city’s regime interpretation is not fixed once built-up uncertainty is acknowledged. Consistent with this, the probabilistic regime summary (Figure 4.6e) shows that two regimes dominate the bootstrap outcomes, with *Inefficient Built-up Expansion under Slower Population Growth* (R16) occurring most often, at 63%, and *Efficient Built-up Expansion under Faster Population Growth* (R2) accounting for the remaining substantial share, at 37%.

The corresponding bootstrap summaries, namely the median and 95% percentile interval, are reported in Table 4.7. Relative to the map-based baseline, the uncertainty-aware results remained aligned with the error-adjusted interpretation. Bootstrapped BUA values were much lower than pixel-counting totals, about 17 km² in 2010 and 20 km² in 2020; therefore, the implied growth

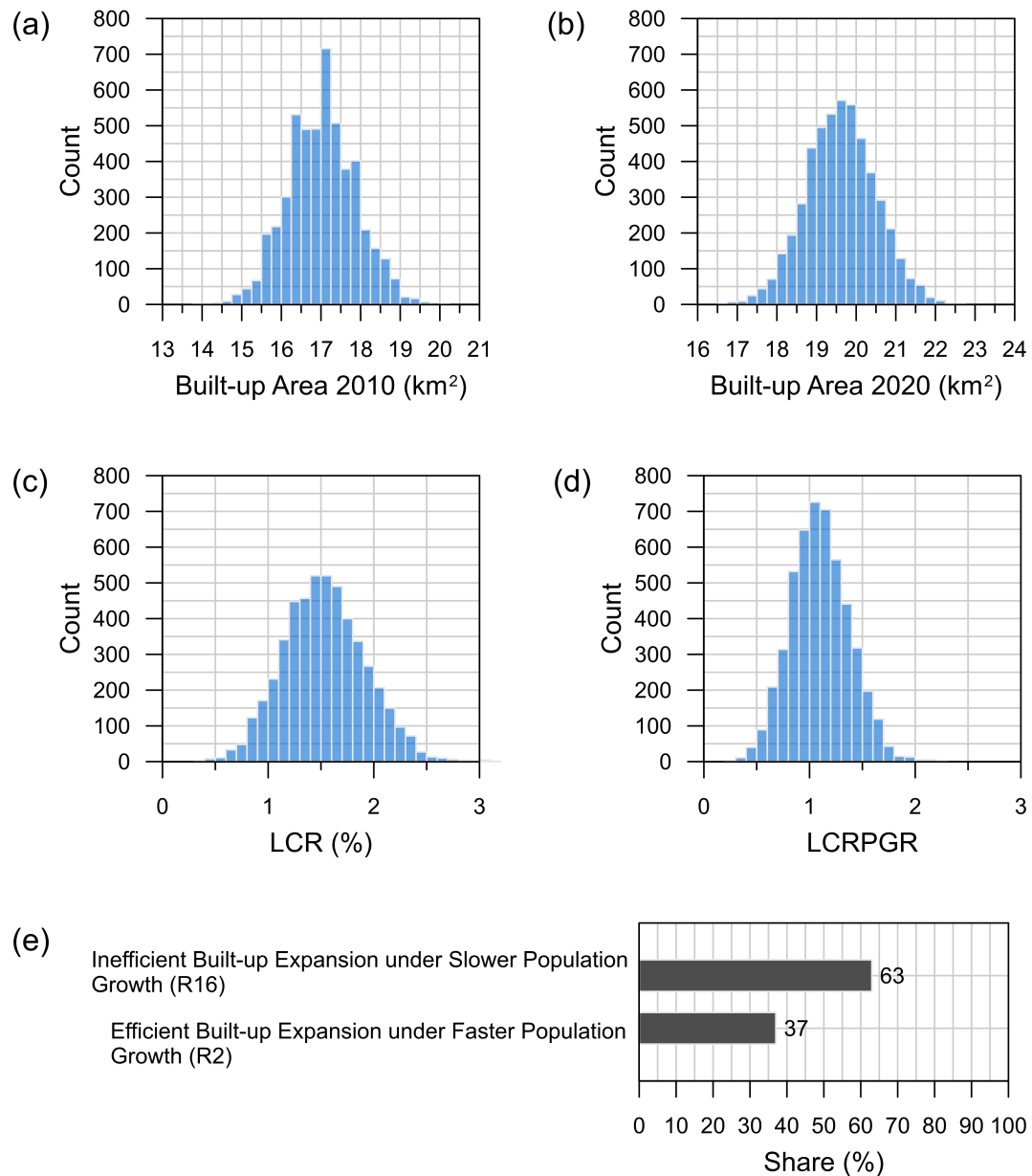


Figure 4.6: Bootstrap distributions of uncertainty-aware SDG Indicator 11.3.1 inputs and outputs for Baguio City (2010–2020): (a) error-adjusted built-up area in 2010, (b) error-adjusted built-up area in 2020, (c) LCR, (d) LCRPGR, and (e) probabilistic LUE regime membership expressed as the share of bootstrap replicates assigned to each regime ($B = 5000$; PGR treated as fixed).

rate in built-up land was higher, with median LCR equal to 1.52% per year, and the resulting LCRPGR is centered above the unity threshold, with median LCRPGR equal to 1.09. At the same time, the uncertainty intervals indicate that LCRPGR can still fall below 1 in a sizable share of realizations, explaining why the city does not map to a single regime deterministically. Overall, the uncertainty-aware assessment identifies *Inefficient Built-up Expansion under Slower Population Growth* (R16) as the most likely regime for Baguio City over 2010–2020, occurring in about 63% of bootstrap replicates, while *Efficient Built-up Expansion under Faster Population Growth* (R2) remains plausible in the remaining 37% (Table 4.7; Figure 4.6e).

4.4.2.7 Application 2 Summary

Using an error-adjusted, bootstrap-based implementation of the uncertainty propagation and probabilistic LUE classification framework, the Baguio City application demonstrates how interpretation of SDG Indicator 11.3.1 can change when BUA is treated as an uncertain, error-prone input. Under map-based pixel counting, Baguio City appeared highly built-up and was classified as *Efficient Built-up Expansion under Faster Population Growth*. In contrast, the reference-based, error-adjusted results indicate that the mapped built-up extent is substantially overestimated, thereby lowering baseline BUA levels and revising the implied built-up growth signal. As a result, the corrected LCR increased and the LCRPGR shifted above the unity threshold, changing the dominant classification to *Inefficient Built-up Expansion under Slower Population Growth*.

The error matrix helps explain why the correction is large. The GLCLUC-derived change map showed only moderate agreement with the reference data, and the weakest performance occurred in the change component, particularly BUA Gain. In addition, the mapped Stable BUA stratum contained a sizable share of off-diagonal entries, indicating substantial commission error relative to the reference interpretation. These patterns are consistent with the difficulty of change mapping at 30 m in heterogeneous urban settings, where mixed pixels and fragmented built-up features can lead to systematic over-assignment of built-up persistence or change. Practically, this means that pixel-counting-based change summaries from the mapped product are not reliable proxies for reference-based area change in this case.

The uncertainty-aware results further show that the city's regime classification is not uniquely determined once uncertainty is acknowledged. The bootstrap distribution of LCRPGR spans both sides of the unity threshold, leading to a split between R16 and R2 outcomes. Reporting the dominant regime along with its membership probability, therefore, provides a more transparent interpretation. That is, Baguio City is most likely characterized by inefficient expansion, but efficient expansion under faster population growth remains a non-negligible possibility.

4.5 Discussion

The two applications presented in the preceding sections provide complementary evidence for evaluating the uncertainty propagation framework developed in this chapter. Considered together, they show that uncertainty-aware assessment of SDG Indicator 11.3.1 is not only conceptually feasible but also practically informative. The hypothetical City A experiment demonstrated the behavior of the framework under controlled conditions, particularly the role of dependence assumptions in shaping indicator distributions and tail behavior. The Baguio City application, in turn, demonstrated that the same framework can be implemented in a real-world setting using empirical BUA uncertainty derived from standard validation and error-adjustment procedures. Although the two applications differ in purpose and level of realism, both point to the same central conclusion: once uncertainty in the inputs is acknowledged and propagated, SDG Indicator 11.3.1 outputs are more appropriately interpreted as distributions with varying levels of classification support rather than as fixed and definitive values.

In relation to RQ2, the results show that explicit uncertainty propagation can alter both the magnitude and stability of SDG Indicator 11.3.1 components compared with deterministic approaches. In City A, propagating uncertain inputs produced indicator distributions that often differed materially from values calculated at central inputs. In Baguio City, treating mapped BUA as fixed led to an Efficient Expansion interpretation, while error adjustment and uncertainty propagation shifted the dominant interpretation toward Inefficient Expansion. These findings confirm that deterministic monitoring can yield regime labels that are sensitive to input error structures, and that distributional reporting is necessary to characterize result stability.

In relation to RQ3, probabilistic LUE classification improves robustness and interpretability by expressing regime membership as likelihoods rather than as a single threshold-based label. This approach also responds to a limitation identified in the literature review. Established interpretation practices for LCRPGR—such as reporting LCR and PGR alongside the ratio, using sign-based interpretation tables, and screening for extreme values (e.g., Estoque et al., 2021; Calka et al., 2022; Cimini et al., 2023)—are primarily deterministic strategies for making ratio-based outcomes more interpretable. While useful, such approaches do not quantify how uncertainty in BUA and Pop propagates into regime assignment. The framework developed in this chapter extends these practices by representing LUE classes probabilistically, thereby conveying not only the dominant regime but also the degree of support for that regime under propagated uncertainty. The City A experiment showed how regime probabilities provide a direct measure of classification sensitivity under competing dependence scenarios. In the Baguio application, regime probabilities indicate that the city does not map uniquely to a single regime when BUA uncertainty is acknowledged; instead, a dominant regime emerges alongside quantified probabilities for alternative regimes. Reporting the dominant regime along with its membership probability, therefore, offers a transparent summary that supports comparability across cities and avoids overconfident conclusions in borderline cases.

From a practical perspective, the chapter offers several reporting takeaways for monitoring SDG Indicator 11.3.1 under uncertain inputs. First, pixel-counting map totals should be treated as descriptive statistics; when EO-derived built-up products are used, reference-based error adjustment is needed to obtain defensible area estimates. Second, uncertainty should be propagated to indicator outputs whenever possible, especially for ratio-based quantities such as LCRPGR, which can be amplified by input errors. Third, to enable routine uncertainty-aware use of EO products, dataset producers should provide validation information in a form that supports design-based inference. At minimum, this includes sampling design and stratum definitions, sample sizes by stratum, and the full confusion/error matrix, ideally with area weights or mapped stratum areas. Providing these elements would allow users to quantify class-area uncertainty transparently and propagate it to SDG metrics using the framework developed in this chapter, rather than relying on pixel counting or on accuracy summaries that are insufficient for uncertainty propagation. Finally, results are best communicated using a combined reporting template: point summaries, such as the median; uncertainty intervals, such as 95% percentile intervals; and probabilistic regime membership. This format retains compatibility with the SDG Indicator 11.3.1 regime taxonomy while adding an explicit measure of confidence.

4.6 Limitations and Further Work

Several limitations should be noted. The uncertainty quantified here is primarily driven by the assumed uncertainty representations for the hypothetical scenarios and by the stratified bootstrap sampling variability for the Baguio City application. Additional sources, such as temporal mismatch between map years and reference imagery, interpreter uncertainty in mixed pixels, and spatial dependence, were minimized through careful design. However, these sources were not explicitly modeled and could further affect the uncertainty intervals. A further limitation in the real-world application arises from the spatial resolution of the mapping and reference framework. Because the reference samples correspond to $30\text{ m} \times 30\text{ m}$ pixels and were assigned labels based on the dominant class, the estimated built-up areas may also be affected by within-pixel heterogeneity. In mixed urban pixels, this can introduce an additional source of uncertainty that is not captured by the bootstrap, which primarily reflects sampling variability under the adopted labeling scheme. Moreover, all interpretations remain conditional on the adopted built-up definition of the input datasets, reinforcing the importance of clear definitional alignment when comparing results across products or studies.

An additional limitation is that Pop uncertainty was not quantified in the real-world application due to insufficient information to construct the corresponding uncertainty distributions. Accordingly, the Baguio City results should be interpreted as demonstrating the practical applicability of the framework under incomplete uncertainty information, rather than as a fully joint uncertainty assessment of SDG Indicator 11.3.1.

For further work, a priority is to extend the framework to real-world settings that incorporate Pop uncertainty, enabling BUA and Pop to be propagated jointly. For census counts, the most direct basis is a post-enumeration survey or other official census coverage assessment. Such data can provide information on undercount, overcount, or net coverage error and can therefore be used to define uncertainty in the reported population totals. For survey-based population estimates, uncertainty may be incorporated using the standard errors or published margins of error released with the estimates. For example, estimates from the American Community Survey (ACS) are accompanied by 90% confidence-level margins of error based on survey variance estimation (US Census Bureau, 2024). For model-based gridded population products, uncertainty can be incorporated directly when the modeling framework provides probabilistic outputs. This may take the form of posterior samples, predictive intervals, or simulation-based uncertainty propagation. For example, recent LandScan HD / LandScan Mosaic work quantifies uncertainty in building use classification, floor counts, and occupancy rates and propagates these through Monte Carlo simulation to produce population-count distributions and associated confidence measures (Adams et al., 2025). By contrast, uncertainty handling remains limited or inconsistently reported across many dasymetric population-mapping workflows and gridded products (Nagle et al., 2014; Leyk et al., 2019). In such cases, a practical alternative is to compare aggregated population estimates against reference data for validation units and use the resulting residual errors to build empirical or bootstrap-based uncertainty distributions. Where a predictive recalibration model is developed, uncertainty may also be quantified through methods such as quantile regression and conformal prediction, which provide interval estimates (Romano et al., 2019; Angelopoulos & Bates, 2021). If suitable population uncertainty information

is available through any of these routes, the same framework can incorporate it directly by treating Pop at t_1 and t_2 as uncertain inputs and propagating them jointly with BUA.

A second direction is to move from the present partial real-world implementation toward fully empirical joint uncertainty propagation in operational SDG Indicator 11.3.1 applications. This would involve not only incorporating uncertainty in both BUA and Pop, but also examining how alternative within-variable and cross-variable dependence structures affect empirical indicator distributions and regime probabilities in real monitoring contexts. Such work would help determine whether the sensitivity to dependence assumptions observed in the controlled City A experiment also emerges systematically in applied settings. This direction is pursued in part in Chapter 5, although in a forecasting rather than a retrospective monitoring setting.

A third direction is to broaden the treatment of uncertainty sources beyond those represented here. While the current framework captures sampling-based uncertainty and illustrates dependence effects, future work could explicitly model additional components such as spatial autocorrelation, temporal misalignment between EO products and reference data, interpreter disagreement in ambiguous or mixed land-cover situations, and uncertainty introduced by dominant-class labeling of coarse pixels in spatially heterogeneous settings. Incorporating these effects would support a more comprehensive representation of the uncertainty budget for SDG Indicator 11.3.1 estimates.

4.7 Chapter Conclusion

This chapter developed and demonstrated an uncertainty-aware formulation of SDG Indicator 11.3.1 in which BUA and Pop are treated as uncertain inputs. Building on the deterministic sensitivity baseline in Chapter 3, the chapter showed how uncertainty in these inputs can be represented and propagated through the SDG Indicator 11.3.1 transformations to obtain empirical distributions of LCR, PGR, and LCRPGR, thereby enabling probabilistic LUE classification. Through the controlled City A experiment and the empirical Baguio City application, the chapter further showed that uncertainty propagation can materially affect indicator magnitudes, widen or reshape their distributions depending on dependence assumptions, and alter the dominant LUE interpretation relative to deterministic results. By shifting from point estimates to distributions, the chapter provides a coherent basis for quantifying how input uncertainty, nonlinear transformations, and dependence structures affect both indicator values and their interpretability.

Overall, Chapter 4 demonstrates that uncertainty-aware SDG Indicator 11.3.1 monitoring is both conceptually straightforward and operationally implementable. By turning uncertain inputs into probability distributions of LCR, PGR, and LCRPGR, and by translating these into probabilistic LUE regime classifications—the chapter moves beyond deterministic interpretation aids toward an explicitly uncertainty-aware basis for LUE assessment. The empirical application also showed that practical implementation may remain partial when uncertainty information is available for only some inputs, particularly BUA, but not Pop. Even under this constraint, however, the framework yielded more informative and transparent outputs than conventional deterministic reporting. In this way, the chapter establishes the methodological foundation for the uncertainty-aware forecasting and forward-looking LUE analyses developed in Chapter 5.

5 Data-Driven, Uncertainty-Aware Deep Learning Forecasting of Built-Up Area and Population for Forward-Looking Probabilistic Assessment of LUE¹⁶

5.1 Chapter Overview

This chapter develops and evaluates a data-driven, uncertainty-aware forecasting workflow for BUA and Pop, and links the resulting predictive uncertainty to forward-looking, probabilistic LUE assessment. Whereas the preceding chapters examined the sensitivity and uncertainty propagation of SDG Indicator 11.3.1 under historical conditions, the focus here shifts to prospective assessment. Future BUA and Pop are treated as uncertain quantities predicted from EO-derived time series and then propagated through the indicator transformations. The resulting predictive distributions are translated into probabilistic LUE assessments and regime probabilities for future intervals using the uncertainty propagation and probabilistic classification framework established in Chapter 4. In doing so, the chapter directly addresses the fourth research question:

- *RQ4. To what extent can data-driven, uncertainty-aware forecasting models accurately and reliably predict built-up area and population time series across diverse localities, while producing predictive uncertainty that enables subsequent SDG Indicator 11.3.1-based LUE assessments to be well-calibrated and decision-relevant?*

In contrast to scenario-based or process-driven projection approaches, the workflow developed here relies on empirical EO-derived time series and supervised deep learning (DL) to learn temporal dynamics directly from observed data. Forecasting is formulated as a probabilistic sequence-learning problem. Thus, the objective is not only to generate accurate point forecasts of future BUA and Pop, but also to quantify the uncertainty surrounding those forecasts through prediction intervals. These uncertainty-aware forecast outputs are then propagated into future LCR, PGR, and LCRPGR outcomes.

The chapter pursues six objectives, all implemented and evaluated through a global-scale application to 8,478 urban centers (UCs) worldwide. First, it formalizes joint BUA–Pop forecasting as a pooled time-series learning task that enables shared representation learning across diverse localities while retaining locality-specific temporal structure. Second, it evaluates multiple DL architectures

¹⁶This chapter builds upon and substantially extends the author’s earlier published work in Santillan et al. (2026), *A deep learning framework for forecasting built-up area and population to support land use efficiency projections under SDG 11.3.1. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10, 385–392, <https://doi.org/10.5194/isprs-annals-X-5-W4-2025-385-2026>, by expanding the methodological formulation, analyses, and dissertation-specific interpretation.

for joint-target forecasting to assess their ability to model coupled land and demographic trajectories accurately and coherently. Third, it implements deep ensemble-based probabilistic forecasting to quantify predictive uncertainty with minimal distributional assumptions. Fourth, it benchmarks the proposed global DL models against a classical exponential smoothing baseline to assess the added value of cross-city learning and modern sequence modeling relative to conventional univariate approaches. Fifth, it examines how forecast uncertainty propagates into future components of SDG Indicator 11.3.1 and probabilistic LUE outcomes. Finally, it demonstrates the application of the uncertainty-aware forecasting workflow to forward-looking SDG Indicator 11.3.1 forecasting for the 2025–2030 period across the global UC dataset.

To clarify the scope of this forecasting analysis, forecasting is carried out conditionally on the available historical BUA–Pop sequences, with the input segment treated as fixed. Under this assumption, the predictive distributions describe uncertainty in future BUA and Pop given the observed historical values. Uncertainty in the historical inputs themselves, such as measurement, definition, or source-data errors, is not explicitly accounted for in the forecasting stage. This choice is made to keep the forecasting analysis focused and to allow a fair comparison of alternative forecasting models using the same input data. It is also consistent with the broader thesis framework, which examines input uncertainty separately. It can later be combined with forecasting uncertainty, as needed, to propagate full uncertainty into the forward-looking assessment of SDG Indicator 11.3.1.

5.2 Related Work

Section 2.2.5 reviewed existing extensions of SDG Indicator 11.3.1 toward forward-looking assessment and showed that many forecasting efforts rely on scenario-based projections or process-driven simulations. These approaches typically employ Shared Socioeconomic Pathway (SSP) scenarios, statistical downscaling frameworks, CA–Markov models, or optimization-based land-use simulations to generate future trajectories of built-up area and population. Such methods are widely used for exploratory planning and policy evaluation under alternative development narratives, and they commonly operationalize future trajectories through scenario assumptions or model-based simulations.

Beyond the SDG Indicator 11.3.1 context, spatial urban growth forecasting has long employed models such as CA (e.g., Al-shalabi et al., 2013; Liu et al., 2017; Santé et al., 2010), agent-based models (ABMs) (e.g., Jokar Arsanjani et al., 2013; Li & Liu, 2007), and various machine learning and deep learning frameworks (e.g., Gómez et al., 2019; Kim et al., 2022; Robi & George, 2025; Yaagoubi et al., 2024). These models are often powerful tools for simulating where growth may occur, but they pose important limitations for the temporal forecasting problem central to SDG Indicator 11.3.1. Their calibration commonly demands extensive, region-specific data and intensive parameter tuning, making large-scale generalization difficult (Li & Liu, 2007; Liu et al., 2017; Santé et al., 2010; Xu et al., 2019; Yaagoubi et al., 2024). More fundamentally, their primary objective is spatial allocation (i.e., predicting the location of new urban areas) rather than learning how aggregate urban metrics evolve. For SDG Indicator 11.3.1, however, the required inputs are

aggregated quantities—BUA and Pop—whose future trajectories must be forecast coherently to support prospective calculation of LCR, PGR, and LCRPGR.

To model changes in aggregate indicators such as BUA and Pop over time, a time-series forecasting approach is more appropriate. 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; Mendis et al., 2024). Joint forecasting of BUA and Pop falls into this latter category because the two processes are coupled through shared socio-economic and spatial drivers (Schiavina et al., 2022a).

Univariate statistical techniques such as exponential smoothing and autoregressive integrated moving-average (ARIMA) models have long been used in urban time series forecasting due to their simplicity, interpretability, and modest data requirements (Hyndman & Athanasopoulos, 2018; Ostertagová & Ostertag, 2012). However, these univariate approaches typically treat each time series independently, limiting their ability to exploit shared temporal dynamics across multiple time series. They also rely on predefined functional forms and assumptions of linearity and stationarity, which restrict their capacity to represent complex, nonlinear growth patterns (Arumugam & Natarajan, 2023; Kim et al., 2025). Conventional machine learning models such as gradient-boosted trees (e.g., XGBoost, LightGBM, CatBoost) can partially address these limitations by capturing nonlinear relationships without strict parametric assumptions (Hall & Rasheed, 2025; Kim et al., 2025). Nonetheless, these machine learning models are not inherently designed to learn temporal dependencies and therefore typically require substantial feature engineering (e.g., lag variables, moving averages, and exogenous predictors) to represent time-related structure (Hall & Rasheed, 2025).

These limitations of both statistical and conventional machine learning approaches are particularly consequential for SDG Indicator 11.3.1 forecasting, which often relies on EO-derived time series for BUA and on gridded or census-based Pop inputs that may be infrequent, short, and of uneven quality. In addition, the forecasting task must generalize across thousands of UCs worldwide, many of which exhibit heterogeneous growth regimes and limited time-series length. These constraints reduce the feasibility of city-specific parameterization and extensive covariate engineering, underscoring the need for scalable, data-driven forecasting workflows that enable models to learn directly from EO time-series data and support coherent joint prediction of BUA and Pop for forward-looking SDG Indicator 11.3.1 assessment.

Recent advances in DL-based time-series forecasting offer a promising path forward. Architectures such as one-dimensional convolutional networks (1D CNNs) (Bai et al., 2018), recurrent neural networks including Long Short-Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997) and Gated Recurrent Units (GRU) (Cho et al., 2014), and Transformer-based models (Vaswani et al., 2017) can learn nonlinear temporal dependencies and extract shared representations from large collections of related time series. Graph Neural Networks (GNNs) enable forecasting in spatiotemporal contexts by learning temporal evolution alongside spatial dependencies among interconnected locations (Jin et al., 2024b). These models have achieved strong performance across diverse forecasting domains, including energy, finance, health, and mobility (Lara-Benítez et al., 2021; Benidis et al., 2023; Li & Law, 2024; Song et al., 2025; Kim et al., 2025). However, in urban growth forecasting, applications

of DL models are comparatively limited and have often focused on population prediction (e.g., Geiß et al., 2023; Grossman et al., 2023; Alghanmi et al., 2024; Budsaba et al., 2025).

A further methodological development, especially relevant for SDG Indicator 11.3.1, is the emergence of global forecasting paradigms, in which a single model is trained across a collection of related time series rather than fitting separate models for each series (Rabanser et al., 2020; Montero-Manso & Hyndman, 2021). Global DL models, in particular, can leverage pooled training across many series to learn shared temporal representations that improve generalization, while still allowing series-specific variability through learned latent structure and/or covariate conditioning (Salinas et al., 2020; Semenoglou et al., 2021; Lim et al., 2021).

Finally, the SDG Indicator 11.3.1 context emphasizes not only predictive accuracy but also uncertainty quantification. As established in the preceding chapters, LCRPGR is highly sensitive to small fluctuations in its components, such that point forecasts alone provide insufficient support for risk-aware interpretation of future outcomes. Probabilistic forecasting methods that yield predictive distributions or calibrated prediction intervals are, therefore, necessary to enable formal uncertainty propagation into future SDG Indicator 11.3.1 calculations. This requirement can be supported by DL models, which can be designed or adapted to produce distributional outputs in several ways. One approach is likelihood-based distributional regression, where the DL model predicts the parameters of a chosen probability distribution for the target (e.g., the mean and variance of a Gaussian) and is trained by minimizing the negative log-likelihood (Salinas et al., 2020). Another approach is direct quantile prediction, where the model is trained to output specific quantiles of the target distribution (e.g., the 5th, 50th, and 95th percentiles) conditional on the input features. This approach provides prediction intervals directly from the estimated quantiles and does not require assuming a particular parametric likelihood or distributional form for the forecast errors (Lim et al., 2021). Additional widely used strategies include Monte Carlo dropout and deep ensembles. Monte Carlo dropout uses the network’s dropout layers to estimate uncertainty by keeping dropout active during inference. Each forward pass randomly deactivates different neurons, producing a slightly different sub-network and a different output. Repeating this multiple times yields a set of predictions whose spread provides an empirical measure of predictive uncertainty (Gal & Ghahramani, 2016; Kumaraka & Srisuradetchai, 2025). Deep ensembles, in contrast, train multiple models independently (e.g., with different random initializations and data shuffles) and quantify uncertainty through the diversity of their predictions; aggregating across ensemble members yields a robust predictive distribution that often improves both accuracy and uncertainty estimation (Lakshminarayanan et al., 2017; Sakib et al., 2025).

Despite these developments, forward-looking extensions of SDG Indicator 11.3.1 remain limited in two respects: first, many approaches do not prioritize direct forecasting of the aggregate inputs required by the indicator; and second, uncertainty is rarely quantified in a form that can be formally propagated into LCR, PGR, and LCRPGR for probabilistic interpretation. Moreover, existing applications of DL forecasting in urban studies have often focused on single targets (e.g., population) rather than coherent joint modeling of land–population dynamics. It remains unclear whether joint-target DL architectures can leverage coupled land–population processes to improve both forecast accuracy and cross-variable coherence, particularly in large-scale settings where thousands of UCs

must be modeled consistently. Such coherence is critical because inconsistencies between independently forecast trajectories can propagate through indicator transformations, leading to unstable or misleading LUE interpretations.

Building on advances in probabilistic DL time-series forecasting and global modeling, particularly quantile-based prediction and ensemble-based uncertainty estimation, this chapter develops and evaluates an uncertainty-aware, data-driven forecasting workflow for joint BUA–Pop prediction across diverse localities. Multiple DL architectures are compared within this workflow. The resulting distributional outputs are explicitly designed for downstream propagation using the uncertainty-propagation and probabilistic LUE classification framework developed in Chapter 4.

5.3 Methods

5.3.1 Overview

The uncertainty-aware, data-driven forecasting workflow for joint BUA–Pop prediction is formulated as a global time-series learning approach trained across a large and diverse set of localities for which historical BUA and Pop time series are available. Instead of fitting separate locality-specific models, a single forecasting model is learned from the pooled collection of BUA–Pop trajectories, enabling shared representation learning of temporal growth dynamics while accommodating heterogeneity across localities. The workflow is designed to integrate directly with the uncertainty propagation and probabilistic LUE classification framework established in Chapter 4. The methodological contribution lies in the specification of forward-looking input uncertainty, in which predictive distributions of BUA and Pop at future epochs are generated by the global forecasting model. These distributional forecasts are then treated as uncertain inputs for SDG Indicator 11.3.1 calculations, enabling prospective distributions of LCR, PGR, and LCRPGR, as well as probabilistic LUE regime outcomes under forecast uncertainty.

5.3.2 Notation and Data Representation for Multiple Localities

Following Benidis et al. (2023), matrices, vectors, and scalars are denoted by bold uppercase (e.g., $\mathbf{Z}$), bold lowercase (e.g., $\mathbf{z}$), and italic symbols (e.g., z), respectively. Calligraphic symbols denote sets/collections (e.g., $\mathcal{Z}$, $\mathcal{X}$).

Let $i \in \{1, \dots, N\}$ index localities and let $t \in \{1, \dots, T\}$ index time on a harmonized temporal grid. The dataset comprising multiple localities is represented as a set of locality time series $\mathcal{Z} = \{\mathbf{z}_{i,1:T_i}\}_{i=1}^N$, with associated covariates $\mathcal{X} = \{\mathbf{X}_{i,1:T_i}\}_{i=1}^N$.

For each locality i , the observed state of BUA and Pop at time t is represented as a paired target vector

$$\mathbf{z}_{i,t} = (\text{BUA}_{i,t}, \text{Pop}_{i,t}) \in \mathbb{R}^2, \quad (5.1)$$

yielding a multivariate time series $[\mathbf{z}_{i,1}, \dots, \mathbf{z}_{i,T}]$ per locality. These values are consolidated into the locality sequence matrix $\mathbf{Z}_{i,1:T} \in \mathbb{R}^{T \times 2}$, which stacks the target vectors as rows over the temporal

grid. Optional covariates are denoted by $\mathbf{x}_{i,t} \in \mathbb{R}^{d_x}$ and are similarly stacked to form the covariate matrix $\mathbf{X}_{i,1:T} \in \mathbb{R}^{T \times d_x}$, which may include time-varying inputs and static locality attributes.

5.3.3 Forecasting Problem Formulation

The general object of interest in probabilistic forecasting is the conditional distribution of future values over a horizon length h , conditioned on the observed history and available covariates (Benidis et al., 2023). Using the set-based notation of Section 5.3.2, this objective can be expressed as

$$p(\mathcal{Z}_{t+1:t+h} \mid \mathcal{Z}_{1:t}, \mathcal{X}_{t+1:t+h}; \boldsymbol{\theta}), \quad (5.2)$$

where $\mathcal{Z}_{t_1:t_2}$ denotes the values of all N locality series over the time slice $[t_1, t_2]$, and $\boldsymbol{\theta}$ denotes the parameters that specify the predictive distribution, i.e., the parameters of the probabilistic model (Benidis et al., 2023). In the present setting, each series is multivariate because $\mathbf{z}_{i,t} \in \mathbb{R}^2$ represents the paired evolution of BUA and Pop. It follows that the conditional distribution in Eq. 5.2 characterizes joint future trajectories of both quantities.

For locality-level forecasting, Eq. 5.2 can be written equivalently for each locality i as

$$p(\mathbf{Z}_{i,t+1:t+h} \mid \mathbf{Z}_{i,1:t}, \mathbf{X}_{i,t+1:t+h}; \boldsymbol{\theta}_{i,t}), \quad i = 1, \dots, N, \quad (5.3)$$

where $\mathbf{Z}_{i,t_1:t_2} \in \mathbb{R}^{(t_2-t_1+1) \times 2}$ denotes the paired BUA–Pop target slice for locality i over $[t_1, t_2]$, and $\boldsymbol{\theta}_{i,t}$ denotes the predictive-distribution parameters for locality i at forecast origin t .

A global modeling strategy is adopted by learning a shared set of model parameters, $\boldsymbol{\Phi}$, from a pooled collection of locality histories. In the formulation of Benidis et al. (2023), the predictive-distribution parameters are produced by a shared mapping

$$\boldsymbol{\theta}_{i,t} = \Psi(\mathbf{Z}_{i,1:t}, \mathbf{X}_{i,t+1:t+h}; \boldsymbol{\Phi}), \quad (5.4)$$

where $\Psi(\cdot)$ denotes the global forecasting model and $\boldsymbol{\Phi}$ denotes its learnable parameters. This global multivariate formulation supports joint BUA–Pop forecasting across diverse localities within a single model, while producing predictive distributions that serve as uncertainty-bearing inputs for forward-looking assessments of SDG Indicator 11.3.1.

5.3.4 Global Joint-Target Sequence-to-Sequence Learning Setup

To operationalize the probabilistic objective in Eqs. 5.2–5.4 as a supervised learning task, $\Psi(\cdot)$ is implemented in a sequence-to-sequence setting with a fixed historical context length L and forecast horizon H . For each locality i and each eligible forecast origin t , an input segment is defined by the most recent L observations,

$$\mathbf{Z}_{i,t}^{\text{in}} = \mathbf{Z}_{i,t-L+1:t} \in \mathbb{R}^{L \times 2}, \quad (5.5)$$

and the corresponding output target segment is defined as the subsequent H steps,

$$\mathbf{Z}_{i,t}^{\text{out}} = \mathbf{Z}_{i,t+1:t+H} \in \mathbb{R}^{H \times 2}. \quad (5.6)$$

Optionally, covariates are aligned to these segments using the analogous slices $\mathbf{X}_{i,t-L+1:t}$ for the historical context and, where required by the forecasting formulation, $\mathbf{X}_{i,t+1:t+H}$ for the forecast horizon.

A global joint-target model with learnable parameters Φ is trained to map the input segment, and optionally the aligned covariates, to the parameters of the predictive distribution of the future segment:

$$\theta_{i,t} = \Psi(\mathbf{Z}_{i,t}^{\text{in}}, \mathbf{X}_{i,t-L+1:t+H}; \Phi), \quad (5.7)$$

thereby defining the conditional distribution $p(\mathbf{Z}_{i,t}^{\text{out}} | \mathbf{Z}_{i,t}^{\text{in}}, \mathbf{X}_{i,t-L+1:t+H}; \theta_{i,t})$. Training examples are created by sliding the $(L \rightarrow H)$ window along each locality's time series and repeating the same procedure for every locality. All resulting input-output pairs are then combined into one training dataset, and a single global model is learned from this pooled set. In this way, the model learns patterns that recur across many localities while still using each locality's historical sequence to produce its forecasts, with BUA and Pop jointly predicted under the same objective.

5.3.5 Probabilistic Forecasting via Quantiles

To enable probabilistic forecasting without imposing a specific parametric distributional form, the global model $\Psi(\cdot)$ is formulated to predict conditional quantiles of the future multi-horizon joint BUA-Pop sequence (Wen et al., 2017; Wang et al., 2019).

Let $\mathcal{Q} = \{q_1, \dots, q_M\} \subset (0, 1)$ denote a prespecified set of quantile levels, e.g., $q \in \{0.025, 0.5, 0.975\}$, which corresponds to the median and the 95% prediction intervals, and let the future segment for locality i at forecast origin t be $\mathbf{Z}_{i,t}^{\text{out}} \in \mathbb{R}^{H \times 2}$, with horizon index $h \in \{1, \dots, H\}$. Conditioned on the historical windows $\mathbf{Z}_{i,t}^{\text{in}}$ and aligned covariates $\mathbf{X}_{i,t}^{\text{in}}$, the model outputs a set of quantile forecasts for each horizon step and each target variable, i.e.,

$$\theta_{(i,t)} = \left\{ \hat{z}_{i,t+h,v}^{(q)} : q \in \mathcal{Q}, h = 1, \dots, H, v \in \{\text{BUA}, \text{Pop}\} \right\} = \Psi(\mathbf{Z}_{i,t}^{\text{in}}, \mathbf{X}_{i,t}^{\text{in}}; \Phi), \quad (5.8)$$

where $\hat{z}_{i,t+h,v}^{(q)}$ is the model-predicted q -quantile of variable v at horizon h . This formulation yields a nonparametric characterization of predictive uncertainty for both targets, directly supporting interval forecasts and probability statements derived from the predicted quantiles (Wen et al., 2017; Wang et al., 2019; Lim et al., 2021).

5.3.6 Deep Learning Architectures for Global Joint BUA-Pop Forecasting

The global model $\Psi(\cdot)$ is instantiated using several DL architectures that share the same sequence-to-sequence forecasting principle but differ in how temporal dependence is represented. In all cases, the model learns a nonlinear mapping from a historical input window, and optional covariates, to a latent representation that summarizes salient temporal structure, and a forecasting head maps this representation to the parameters of the predictive distribution for the future segment. Architectures considered include 1D CNNs, recurrent models (LSTM and GRU), hybrid models (1D CNN+LSTM and 1D CNN+GRU), and Transformer-based models. Each architecture maps the historical input segment $\mathbf{Z}_{i,t}^{\text{in}}$ to a shared latent representation, which is then used to parameterize the predictive

distribution for the future segment $\mathbf{Z}_{i,t}^{\text{out}}$. Multiple architectures are considered to test how different temporal representation mechanisms affect the model’s ability to learn shared BUA–Pop dynamics from the pooled set of locality time series under a single global learning formulation.

Each architecture is described in the following subsections. To ensure consistency with the global forecasting formulation introduced earlier, the original equations from the respective sources are re-expressed using the notation adopted in this chapter. While the overarching notation for the global model is retained, architecture-specific recursions are written using a within-window index $\tau \in \{t - L + 1, \dots, t\}$. For locality i , the input window $\mathbf{Z}_{i,t}^{\text{in}} = \mathbf{Z}_{i,t-L+1:t} \in \mathbb{R}^{L \times 2}$ consists of the sequence of paired vectors $\mathbf{z}_{i,\tau} = (\text{BUA}_{i,\tau}, \text{Pop}_{i,\tau})$. Hidden states are denoted $\mathbf{h}_{i,\tau}$. This additional notation is used only to define the temporal representation mechanism. All architectures implement the same mapping from $\mathbf{Z}_{i,t}^{\text{in}}$, and aligned covariates, to predictive-distribution parameters $\boldsymbol{\theta}_{i,t}$.

5.3.6.1 1D CNN

A 1D CNN (one-dimensional convolutional neural network) models sequential data by applying learnable filters along the temporal dimension to extract local temporal patterns and short-range dependencies. In forecasting applications, these temporal convolutions are typically implemented as 1D convolutional (Conv1D) layers. They are often organized into temporal convolutional networks (TCNs), which incorporate design choices such as causal and dilated convolutions, often with residual blocks, to preserve temporal order and expand the receptive field for sequence modeling (Bai et al., 2018). Each filter performs a weighted aggregation over a fixed-length window of past observations, producing a sequence of latent feature vectors that can highlight temporal structures such as rapid changes, periodicities, or gradual transitions. In its simplest form, a temporal convolution layer (Conv1D) calculates a feature vector at position τ as a weighted sum over a local window of length k ,

$$\mathbf{h}_{i,\tau} = \phi \left(\sum_{r=0}^{k-1} \mathbf{W}_r \mathbf{z}_{i,\tau-r} + \mathbf{b} \right), \quad (5.9)$$

where r is the kernel offset, or lag index, $\mathbf{W}_r$ are learnable filter weights, $\mathbf{b}$ is a bias term, and $\phi(\cdot)$ is an element-wise nonlinear activation function. A 1D CNN forecasting model is then formed by stacking such layers, optionally with dilation or residual connections, to produce a sequence of latent features, which is subsequently mapped—via a decoder or output head—to the predictive-distribution parameters $\boldsymbol{\theta}_{i,t}$ for the joint BUA–Pop outputs $\mathbf{Z}_{i,t}^{\text{out}}$. Stacking multiple convolutional layers increases the abstraction of temporal representations and enables the extraction of both fine-scale and coarse-scale variations in the input sequence (Bai et al., 2018; Cacciari & Ranfagni, 2024).

The temporal sensitivity of 1D CNN models is controlled by architectural parameters such as kernel size, stride, and dilation. Kernel size determines the temporal span processed at each position, while stride controls the step size with which the filter moves along the sequence. Dilation enlarges the receptive field by spacing filter applications, allowing the model to capture longer-range dependencies without increasing the number of parameters (Cacciari & Ranfagni, 2024; Yu & Koltun, 2015). For sequence forecasting, causal convolutions, commonly implemented via causal padding, preserve temporal order by ensuring that each output depends only on current and past inputs,

thereby preventing information from future time steps from entering the representation (Bai et al., 2018).

5.3.6.2 LSTM

LSTM (Long Short-Term Memory) is a type of recurrent neural network (RNN) designed to capture long-term temporal dependencies in sequential data (Hochreiter & Schmidhuber, 1997). Unlike conventional RNNs, which suffer from vanishing or exploding gradients when trained on long sequences, LSTMs introduce a gated memory architecture that regulates how information is stored, updated, and retrieved across time steps (Gers et al., 2000; Goodfellow et al., 2016).

An LSTM unit maintains (i) a cell state $\mathbf{c}_{i,\tau}$ that acts as persistent memory and (ii) a hidden state $\mathbf{h}_{i,\tau}$ that serves as the output representation at within-window position τ . Three multiplicative gates control information flow: the input gate, how much new information is written; the forget gate, how much prior memory is retained; and the output gate, how much memory is exposed as output. These mechanisms allow the model to preserve context over long temporal spans while dynamically adjusting to new patterns (Hochreiter & Schmidhuber, 1997; Gers et al., 2000; Krichen & Mihoub, 2025).

Using the within-window index τ and the joint input vector $\mathbf{z}_{i,\tau}$, the LSTM update equations are (Hochreiter & Schmidhuber, 1997; Gers et al., 2000)¹⁷:

$$\mathbf{f}_{i,\tau} = \sigma(\mathbf{W}_f \mathbf{z}_{i,\tau} + \mathbf{U}_f \mathbf{h}_{i,\tau-1} + \mathbf{b}_f), \quad (5.10)$$

$$\mathbf{g}_{i,\tau} = \sigma(\mathbf{W}_g \mathbf{z}_{i,\tau} + \mathbf{U}_g \mathbf{h}_{i,\tau-1} + \mathbf{b}_g), \quad (5.11)$$

$$\tilde{\mathbf{c}}_{i,\tau} = \tanh(\mathbf{W}_c \mathbf{z}_{i,\tau} + \mathbf{U}_c \mathbf{h}_{i,\tau-1} + \mathbf{b}_c), \quad (5.12)$$

$$\mathbf{c}_{i,\tau} = \mathbf{f}_{i,\tau} \odot \mathbf{c}_{i,\tau-1} + \mathbf{g}_{i,\tau} \odot \tilde{\mathbf{c}}_{i,\tau}, \quad (5.13)$$

$$\mathbf{o}_{i,\tau} = \sigma(\mathbf{W}_o \mathbf{z}_{i,\tau} + \mathbf{U}_o \mathbf{h}_{i,\tau-1} + \mathbf{b}_o), \quad (5.14)$$

$$\mathbf{h}_{i,\tau} = \mathbf{o}_{i,\tau} \odot \tanh(\mathbf{c}_{i,\tau}), \quad (5.15)$$

where $\mathbf{f}_{i,\tau}$, $\mathbf{g}_{i,\tau}$, and $\mathbf{o}_{i,\tau}$ are the forget, input, and output gate vectors, respectively, with each gate belonging to $(0, 1)^{d_h}$. The quantities $\tilde{\mathbf{c}}_{i,\tau}$, $\mathbf{c}_{i,\tau}$, and $\mathbf{h}_{i,\tau}$ denote the candidate cell state, current cell state, and hidden state, respectively, and each belongs to $\mathbb{R}^{d_h}$. The functions $\sigma(\cdot)$ and $\tanh(\cdot)$ are the sigmoid and hyperbolic tangent nonlinear activation functions, respectively. The operator $\odot$ denotes the Hadamard product, i.e., element-wise multiplication. Subscripts f , g , o , and c denote gate-specific parameter sets for the forget, input, output, and candidate states.

An LSTM forecasting model is formed by unrolling this recurrence over the input window $\mathbf{Z}_{i,t-L+1:t}^{\text{in}} = \mathbf{Z}_{i,t-L+1:t}$, thereby producing a sequence of hidden states $\{\mathbf{h}_{i,\tau}\}$ that encodes progressively richer summaries of the joint BUA–Pop history. The final hidden state, or an aggregation over hidden states, provides a fixed-dimensional latent representation of the input segment, which is then

¹⁷To avoid notational conflict with the locality index i , the input gate is denoted $\mathbf{g}_{i,\tau}$ rather than $\mathbf{i}_{i,\tau}$.

mapped—via a decoder or output head—to the predictive-distribution parameters $\theta_{i,t}$ for the future segment $\mathbf{Z}_{i,t}^{\text{out}}$ under the global sequence-to-sequence formulation.

5.3.6.3 GRU

GRU (Gated Recurrent Unit) is an RNN architecture introduced by Cho et al. (2014) as a simplified alternative to the LSTM network for modeling sequential dependencies. Like LSTM, the GRU is designed to mitigate the vanishing-gradient problem that limits standard RNNs when learning long-range temporal relationships. However, the GRU achieves this goal with a more compact structure that merges the memory cell and hidden state into a single representation and replaces the three-gate system of the LSTM, namely input, forget, and output gates, with only two gates: the update gate and the reset gate (Cho et al., 2014; Chung et al., 2014).

The update gate controls how much of the previous hidden state is carried forward to the next time step, thereby governing long-term information retention. The reset gate determines how much past information is forgotten when integrating new input, allowing the network to adapt quickly to changing patterns. This gating mechanism enables the GRU to balance memory persistence and adaptability while using fewer parameters than an equivalent LSTM, resulting in faster training and reduced risk of overfitting on small or medium-sized datasets (Chung et al., 2014; Lara-Benítez et al., 2021).

The GRU calculates the reset and update gates as

$$\mathbf{r}_{i,\tau} = \sigma(\mathbf{W}_r \mathbf{z}_{i,\tau} + \mathbf{U}_r \mathbf{h}_{i,\tau-1} + \mathbf{b}_r), \quad (5.16)$$

$$\mathbf{u}_{i,\tau} = \sigma(\mathbf{W}_u \mathbf{z}_{i,\tau} + \mathbf{U}_u \mathbf{h}_{i,\tau-1} + \mathbf{b}_u), \quad (5.17)$$

and forms a candidate hidden state as

$$\tilde{\mathbf{h}}_{i,\tau} = \tanh(\mathbf{W}_h \mathbf{z}_{i,\tau} + \mathbf{U}_h (\mathbf{r}_{i,\tau} \odot \mathbf{h}_{i,\tau-1}) + \mathbf{b}_h). \quad (5.18)$$

The updated hidden state is then obtained by interpolating between the previous hidden state and the candidate state,

$$\mathbf{h}_{i,\tau} = (1 - \mathbf{u}_{i,\tau}) \odot \mathbf{h}_{i,\tau-1} + \mathbf{u}_{i,\tau} \odot \tilde{\mathbf{h}}_{i,\tau}. \quad (5.19)$$

In the above formulations, $\mathbf{r}_{i,\tau} \in (0, 1)^{d_h}$ is the reset gate, $\mathbf{u}_{i,\tau} \in (0, 1)^{d_h}$ is the update gate, $\sigma(\cdot)$ is the logistic sigmoid, $\tanh(\cdot)$ is the hyperbolic tangent, and $\odot$ denotes element-wise multiplication (Cho et al., 2014; Chung et al., 2014).

Following the same sequence-to-sequence construction described for the LSTM architecture, the GRU recurrence in Eq. 5.19 is applied step-by-step across the historical input window to produce a latent representation that summarizes the joint BUA–Pop history. The shared forecasting head then maps this representation to the predictive-distribution parameters for the future joint BUA–Pop outputs under the global formulation.

5.3.6.4 Hybrid Architectures

Hybrid architectures that combine temporal convolution and recurrence integrate the complementary strengths of both model families for sequence forecasting. In these designs, a 1D CNN, implemented via stacked Conv1D layers, first acts as a temporal feature extractor, learning local patterns and short-range dependencies through parallel filtering across the input window, while the subsequent recurrent layer, LSTM or GRU, serves as a temporal aggregator that captures longer-range dynamics by integrating information across the extracted feature sequence (Kim et al., 2025b).

The convolutional block first transforms the joint BUA–Pop input window into a sequence of latent feature vectors. These features correspond to multiple learned channels, or filters, and may have slightly different sequence lengths depending on the padding or stride choices. The transformed representation emphasizes local temporal structure, such as rapid changes, periodicities, and short-term interactions between BUA and Pop, before being passed as sequential input to the recurrent block. The subsequent LSTM/GRU layer then processes this feature sequence step by step, producing a final hidden representation that summarizes both the locally extracted patterns and their longer-range temporal organization (Huang et al., 2021; Khatun et al., 2024; Kim et al., 2025; Sajad et al., 2020). This final representation is then mapped, via the same forecasting head used across architectures, to the parameters of the predictive distribution for the joint BUA–Pop outputs under the global sequence-to-sequence formulation.

5.3.6.5 Transformer

The Transformer (Vaswani et al., 2017) models temporal dependence using self-attention, replacing recurrent updates with a parallel mechanism that relates all time steps within the input window. This feature enables efficient learning of both short- and long-range dependencies because each time step can directly attend to every other time step when forming contextualized representations.

Given the historical input window $\mathbf{Z}_{i,t}^{\text{in}} \in \mathbb{R}^{L \times 2}$, the sequence is first mapped to a d -dimensional embedding and augmented with positional information to preserve temporal order,

$$\mathbf{E}_{i,t} = \mathbf{Z}_{i,t}^{\text{in}} \mathbf{W}_e + \mathbf{1} \mathbf{b}_e^T + \mathbf{P}_L, \quad (5.20)$$

where $\mathbf{W}_e \in \mathbb{R}^{2 \times d}$, $\mathbf{b}_e \in \mathbb{R}^d$, and $\mathbf{P}_L \in \mathbb{R}^{L \times d}$ denotes a fixed or learnable positional encoding. In a self-attention layer, queries ($\mathbf{Q}$), keys ($\mathbf{K}$), and values ($\mathbf{V}$) are obtained by linear projections,

$$\mathbf{Q} = \mathbf{E} \mathbf{W}^Q, \quad \mathbf{K} = \mathbf{E} \mathbf{W}^K, \quad \mathbf{V} = \mathbf{E} \mathbf{W}^V, \quad (5.21)$$

and the resulting contextual representation is calculated via scaled dot-product attention,

$$\text{Attn}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax} \left(\frac{\mathbf{Q} \mathbf{K}^T}{\sqrt{d_k}} \right) \mathbf{V}. \quad (5.22)$$

Multi-head attention applies Eq. 5.22 in parallel across multiple heads and concatenates the results, followed by a linear projection (Vaswani et al., 2017). Each Transformer block then applies a position-wise feed-forward transformation to refine the representation. Stacking such blocks yields an encoded sequence that summarizes the full temporal context of the input window.

For forecasting, Transformer models are commonly used in encoder–decoder or encoder-only configurations (Shen et al., 2025; Wen et al., 2022). In either case, the encoded representation is mapped by a forecasting head to the predictive-distribution parameters $\theta_{i,t}$ for the future segment $\mathbf{Z}_{i,t}^{\text{out}}$. Autoregressive generation can be enforced when needed through causal masking in the attention mechanism, ensuring that representations and predictions at each step depend only on available past information (Lim et al., 2021; Vaswani et al., 2017; Zhou et al., 2020).

5.3.7 Model Training Objective: Multi-Horizon Quantile Loss

Recall from Section 5.3.5 that the adopted probabilistic specification represents the predictive distribution of the future joint BUA–Pop segment through a set of conditional quantiles $\mathcal{Q} = \{q_1, \dots, q_M\}$. Accordingly, for each locality i , forecast origin t , horizon step $h \in \{1, \dots, H\}$, and target variable $v \in \{\text{BUA}, \text{Pop}\}$, the global model $\Psi(\cdot)$ outputs the quantile forecast $\hat{z}_{i,t+h,v}^{(q)}$ for every $q \in \mathcal{Q}$. The DL instantiations of Ψ are trained to estimate these quantiles by minimizing the pinball, or quantile, loss (Wen et al., 2017; Wang et al., 2019).

For an observed value z and a predicted q -quantile $\hat{z}^{(q)}$, the pinball loss is written in piecewise form as (Wang et al., 2019)

$$L_q(z, \hat{z}^{(q)}) = \begin{cases} (1-q)(\hat{z}^{(q)} - z), & \hat{z}^{(q)} \geq z, \\ q(z - \hat{z}^{(q)}), & \hat{z}^{(q)} < z. \end{cases} \quad (5.23)$$

Eq. 5.23 shows an asymmetric penalty, where under-prediction is penalized proportionally to q , while over-prediction is penalized proportionally to $(1-q)$. Minimizing L_q yields $\hat{z}^{(q)}$ as an estimator of the conditional q -quantile of z .

The overall training objective aggregates this loss across the multi-horizon output segment, the two joint targets, and all quantile levels. Let $\mathcal{T}_i$ denote the set of admissible forecast origins for locality i , i.e., times for which both an input window of length L and an output window of length H are available, and let $\mathcal{V} = \{\text{BUA}, \text{Pop}\}$. The minimization objective used for training is expressed as the average pinball loss

$$\mathcal{L}(\Psi) = \frac{1}{\sum_i |\mathcal{T}_i|} \sum_i \sum_{t \in \mathcal{T}_i} \left[\frac{1}{H|\mathcal{V}||\mathcal{Q}|} \sum_{h=1}^H \sum_{v \in \mathcal{V}} \sum_{q \in \mathcal{Q}} L_q(z_{i,t+h,v}, \hat{z}_{i,t+h,v}^{(q)}) \right]. \quad (5.24)$$

This objective function enforces multi-horizon learning by penalizing quantile errors at every lead time within the forecast segment, while simultaneously learning joint-target quantile forecasts for BUA and Pop under a single global model shared across localities.

5.3.8 Deep Ensemble Learning

Deep ensemble learning is used to improve the stability and reliability of probabilistic forecasts from each DL architecture. Ensemble learning combines multiple independently trained base models rather than relying on a single estimator, with the final prediction obtained through an aggregation

strategy (Huang et al., 2009). Deep ensembles extend this principle by leveraging the expressive capacity of deep neural networks while benefiting from the stability and variance-reduction properties of ensemble methods (Lakshminarayanan et al., 2017). By combining these strengths, deep ensembles typically generalize better than individual models and deliver more reliable predictive performance (Dong et al., 2020; Ganaie et al., 2022). This feature makes them well-suited for applications that require both accurate forecasts and credible uncertainty estimates.

In the present workflow, each DL architecture produces direct multi-horizon quantile forecasts for the joint BUA–Pop outputs (Section 5.3.5) and is trained using the quantile loss objective described in Section 5.3.7. However, stochastic elements of DL training, such as random initialization, minibatch ordering, and optimizer dynamics, can introduce run-to-run variability in the resulting quantile estimates. Deep ensembles provide a principled way to consolidate these independent training realizations into a single, more stable predictive representation. For a given architecture, R networks are trained independently under the same data partitioning but different random seeds, producing replicate-wise quantile forecasts $\hat{z}_{i,t+h,v}^{(q,r)}$, for $r = 1, \dots, R$. The ensemble quantile forecast is then obtained by aggregating replicate predictions at each quantile level, horizon, and target variable using a fusion operator consistent with quantile outputs, such as the mean or median across replicates:

$$\hat{z}_{i,t+h,v}^{(q,\text{ensemble})} = \text{Agg} \left(\hat{z}_{i,t+h,v}^{(q,1)}, \dots, \hat{z}_{i,t+h,v}^{(q,R)} \right), \quad (5.25)$$

where $\text{Agg}(\cdot)$ denotes the chosen aggregation rule.

Although the base DL models already produce quantile forecasts, ensemble aggregation reduces sensitivity to stochasticity in training. It helps ensure that the predictive distributions used in subsequent analyses reflect the architecture’s systematic behavior rather than variation attributable to a single training realization. This sensitivity reduction is especially important in uncertainty-aware forecasting, where stable, reproducible quantile estimates are required for downstream tasks, such as propagating BUA and Pop forecast uncertainty into SDG Indicator 11.3.1 components. The ensemble therefore serves both as a performance-enhancing mechanism and as a diagnostic check on whether random initialization and optimization variability materially influence the probabilistic forecasts used for forward-looking LUE assessment.

5.3.9 Baseline Model: Exponential Smoothing

Exponential Smoothing (ES) is used as a classical local baseline to benchmark the proposed global, DL-based joint BUA–Pop forecasting workflow. ES models temporal evolution as an exponentially weighted average of past observations, assigning larger weight to more recent observations than to older ones (Hyndman & Athanasopoulos, 2018). This recursive updating mechanism provides a transparent reference for what can be achieved using only each locality’s own history, without cross-locality learning, feature sharing, or deep representations. ES methods are widely adopted for short- to medium-term forecasting due to their simplicity, interpretability, and robustness to noise (e.g., Taylor, 2003, 2012; Ferbar Tratar et al., 2016). In this study, ES is fit separately for each locality i and separately for each target series $v \in \{\text{BUA}, \text{Pop}\}$, yielding univariate baseline forecasts for BUA and Pop.

Let $z_{i,t,v}$ denote the observed value of variable v for locality i at time t . ES models summarize each series using components such as level, trend, and, when present, seasonality (Hyndman & Athanasopoulos, 2018). The term $\ell_{i,t,v}$ denotes the model’s smoothed estimate of the underlying baseline magnitude of the series at time t , i.e., the local “current value” after filtering short-term fluctuations, which serves as the reference point for forecasting. The trend component $b_{i,t,v}$ captures sustained increase or decrease, i.e., the slope of the series, and the seasonal component $s_{i,t,v}$ captures repeating, periodic, and stable fluctuations within the series. Candidate ES models could also include a damped trend, in which the projected increase or decrease is gradually moderated at longer forecast horizons. Although the simplest ES form, Simple Exponential Smoothing, updates only the level, the ES baseline implemented in this study allows for level + trend and, where relevant, seasonality, with additive or multiplicative forms. Additive forms are appropriate when trend or seasonal effects are roughly constant in magnitude, whereas multiplicative forms are appropriate when these effects scale with the level of the series, i.e., are proportional (Gardner, 2006; Hyndman & Athanasopoulos, 2018).

The exponential smoothing principle for the level component can be written as

$$\ell_{i,t,v} = \alpha_v z_{i,t,v} + (1 - \alpha_v) \ell_{i,t-1,v}, \quad 0 \leq \alpha_v \leq 1, \quad (5.26)$$

where α_v is the smoothing parameter that controls the weight placed on the most recent observation (Hyndman & Athanasopoulos, 2018). Forecasts are generated by extrapolating the estimated components into the prediction horizon. In the common additive-trend form, Holt’s Linear Trend Method, the h -step-ahead forecast given information up to time t is

$$\hat{z}_{i,t+h,v} = \ell_{i,t,v} + h b_{i,t,v}, \quad h = 1, \dots, H, \quad (5.27)$$

where $b_{i,t,v}$ denotes an estimate of the trend, or slope, of the series at time t (Hyndman & Athanasopoulos, 2018). When seasonality is present, Holt–Winters variants extend Eq. 5.27 by incorporating a seasonal component, additive or multiplicative, so that forecasts reflect both the extrapolated level–trend structure and the expected seasonal effect at lead time (Hyndman & Athanasopoulos, 2018). Smoothing parameters and any required initial component values are estimated from the available historical series using standard likelihood-based or error-minimization procedures, and the fitted ES specification is used to generate baseline forecasts over the target horizon (Gardner, 2006).

Forecast uncertainty is obtained from the statistical formulation of ES models, which yields analytical multi-step forecast variances for the selected specification (Hyndman et al., 2005; Hyndman & Athanasopoulos, 2018). Let $\sigma_{i,t+h,v|t}^2$ denote the forecast variance corresponding to $\hat{z}_{i,t+h,v}$. Assuming approximately normally distributed forecast errors, a 95% prediction interval is calculated as (Hyndman & Athanasopoulos, 2018)

$$\hat{z}_{i,t+h,v|t} \pm 1.96 \sqrt{\sigma_{i,t+h,v|t}^2}, \quad h = 1, \dots, H. \quad (5.28)$$

5.3.10 Model Performance Evaluation Strategy and Metrics

Model performance is evaluated to (i) compare alternative deep learning (DL) architectures instantiated within the same global forecasting formulation and (ii) contextualize their performance

relative to a classical baseline benchmark. The primary performance comparison is therefore conducted across the set of candidate DL architectures (1D CNN, LSTM, GRU, hybrid variants, and Transformer), while ES is evaluated under the same design as a reference, indicating the performance achievable with a widely used classical time-series approach.

Evaluation combines point-forecast accuracy metrics and probabilistic uncertainty metrics (Table 5.1). Point-forecast accuracy is assessed using mean absolute error (MAE) and mean absolute percentage error (MAPE), which compare the model's point prediction with the observed value at each forecast horizon (Hyndman & Koehler, 2006). The point forecast is taken from the central tendency of the predictive output. For example, in a quantile-based model that outputs the 2.5th, 50th, and 97.5th percentiles, the median is used to calculate MAE and MAPE, whereas in a model that outputs a predictive mean and an interval, the mean is used.

Table 5.1: Model performance evaluation metrics. In the formulae, n is the number of observations and j is the case index; z_j is the observed value; $\hat{z}_j$ is the predicted value; and $\hat{z}_j^L$ and $\hat{z}_j^U$ are lower and upper bounds of the prediction interval. For the Winkler Score, α denotes the miscoverage level of the central prediction interval, such that $1 - \alpha$ is the nominal coverage probability. Sources: Hyndman and Koehler (2006), Jensen et al. (2024), and Hyndman and Athanasopoulos (2018).

Metric	Formula
Mean Absolute Error (MAE)	$\text{MAE} = \frac{1}{n} \sum_{j=1}^n z_j - \hat{z}_j $
Mean Absolute Percentage Error (MAPE)	$\text{MAPE} = \frac{100}{n} \sum_{j=1}^n \left \frac{z_j - \hat{z}_j}{z_j} \right $
Prediction Interval Coverage Probability (PICP)	$\text{PICP} = \frac{1}{n} \sum_{j=1}^n c_j, \quad c_j = \begin{cases} 1, & \text{if } \hat{z}_j^L \leq z_j \leq \hat{z}_j^U, \\ 0, & \text{otherwise.} \end{cases}$
Prediction Interval Average Width (PIAW)	$\text{PIAW} = \frac{1}{n} \sum_{j=1}^n (\hat{z}_j^U - \hat{z}_j^L)$
Winkler Score (WS)	$\text{WS}_j = \begin{cases} \hat{z}_j^U - \hat{z}_j^L, & \text{if } \hat{z}_j^L \leq z_j \leq \hat{z}_j^U, \\ (\hat{z}_j^U - \hat{z}_j^L) + \frac{2}{\alpha} (\hat{z}_j^L - z_j), & \text{if } z_j < \hat{z}_j^L, \\ (\hat{z}_j^U - \hat{z}_j^L) + \frac{2}{\alpha} (z_j - \hat{z}_j^U), & \text{if } z_j > \hat{z}_j^U. \end{cases}$

Probabilistic performance is assessed using prediction interval coverage probability (PICP), prediction interval average width (PIAW), and Winkler Score (WS) (Jensen et al., 2024; Khosravi

et al., 2011; Hyndman & Athanasopoulos, 2018). These metrics are calculated from the model’s prediction interval. PICP measures the proportion of observed values falling within the predicted interval and is interpreted relative to its nominal coverage level. For example, for a 95% interval, the nominal PICP is 0.95. Values below this level indicate under-coverage, whereas values above it indicate over-coverage. PIAW measures the average width of the same interval and reflects sharpness, with narrower intervals indicating better precision, provided that coverage remains adequate. Because PICP and PIAW should be interpreted jointly, WS is also used as a single-interval performance metric that rewards narrow intervals while penalizing those that fail to contain the observed value.

To preserve the spatial independence of evaluation units, performance summaries are constructed at the locality level. For each model, target variable, and locality, absolute error, absolute percentage error, prediction interval width, and WS are first calculated separately for each evaluated forecast horizon. The locality-level MAE is then obtained by averaging the absolute errors across the evaluated horizons. The same within-locality averaging is applied to MAPE, prediction interval width, and WS, so that each locality contributes a single value for each metric per model. These locality-level values are then summarized across all localities using the median and median absolute deviation (MAD). This procedure avoids treating multiple forecast horizons derived from the same input sequence as independent observations.

In contrast, PICP is evaluated as a pooled empirical coverage measure across all locality–horizon combinations. For each model and target variable, all forecast cases are checked to determine whether the observed value falls within the corresponding prediction interval, and the proportion of covered cases is then calculated across the full set of localities and evaluation horizons. Thus, MAE, MAPE, PIAW, and WS are first averaged within locality and then summarized across localities, whereas PICP is calculated from the pooled set of interval inclusion outcomes.

5.3.11 Model Selection for Downstream SDG Indicator 11.3.1 Propagation

The model selection step uses the locality-level performance summaries, calculated using the procedures described in Section 5.3.10, to rank and identify the “best” forecasting architecture whose predictive distributions are propagated through the SDG Indicator 11.3.1 transformations. Because the forecasting task is joint-target, ranking is based on a unified performance quantity that reflects errors in both BUA and Pop.

To rank joint-target architectures using a single accuracy metric, the BUA and Pop multi-horizon MAE values are first placed on a common, unitless scale before being combined. This standardization is necessary because BUA and Pop are expressed in different units and typically differ substantially in magnitude. Without rescaling, a joint error criterion would be dominated by the variable with larger numerical errors, rather than reflecting balanced forecasting skill in both targets. Accordingly, for each target variable, the full collection of locality-level MAE values, across all evaluated localities and candidate models, is transformed using a robust location-and-scale standardization. In this case, a representative center of the MAE distribution, the median, is subtracted, and the result is divided by a robust measure of spread, the MAD. Robust standardization is used

because forecast error distributions across heterogeneous localities can be strongly right-skewed and heavy-tailed, so classical mean–standard deviation scaling may be unduly influenced by extreme errors (Huber & Ronchetti, 2009). After standardization is performed separately for BUA and Pop, the two standardized error values are combined with equal weight, e.g., by averaging, to obtain a single joint standardized error per locality and model, which is then used as the input for rank-based model comparison and selection.

To determine whether observed differences reflect systematic performance advantages rather than random variation across localities, a two-stage nonparametric comparison framework is applied (Demšar, 2006). First, a Friedman test (Friedman, 1937) is performed on the locality-wise joint standardized errors to assess whether there are statistically significant differences among candidate architectures. The Friedman procedure ranks models within each locality and aggregates the ranks across localities to yield a mean rank per model, with lower values indicating better overall joint performance. If the Friedman test indicates significant differences, post-hoc pairwise comparisons are conducted using the Wilcoxon signed-rank test (Wilcoxon, 1945) on paired locality-wise joint errors. Because multiple pairwise tests are performed, Holm’s sequential correction is applied to control the family-wise error rate, i.e., to minimize false positives (Holm, 1979). Effect sizes are summarized using rank-biserial correlation (Kerby, 2014; Tomczak & Tomczak, 2014) to quantify the magnitude and direction of pairwise differences and to distinguish practically meaningful improvements from statistically detectable but negligible gaps.

Final model selection prioritizes the architecture with the most favorable overall ranking while ensuring robustness under post-hoc statistical testing. Specifically, the selected model is the top-ranked architecture, i.e., the one with the lowest Friedman mean rank, that is not statistically inferior to competing candidates under Holm-adjusted Wilcoxon comparisons and exhibits non-negligible effect-size advantages where differences are significant. When the top-ranked candidates are statistically indistinguishable, selection is resolved by parsimony, favoring the architecture with lower structural complexity while maintaining comparable forecasting skill. Finally, the probabilistic diagnostics from Section 5.3.10, namely PICP and WS, are used as an additional eligibility check to confirm that the chosen model’s predictive intervals remain appropriately calibrated and sufficiently sharp for reliable propagation through the SDG Indicator 11.3.1 transformations. The selected joint-target model provides the predictive distributions of BUA and Pop, which are propagated in Section 5.3.12 and used for probabilistic LUE classification.

5.3.12 Forecast Uncertainty Propagation to SDG Indicator 11.3.1

The selected joint-target forecasting model produces predictive quantiles for future BUA and Pop at the required endpoint years. These quantile outputs are converted into continuous empirical predictive distributions via monotone piecewise linear interpolation, with the predicted quantiles serving as anchor points. This interpolation defines an empirical cumulative distribution function (CDF) for each forecasted variable, allowing the predictive distribution to retain the asymmetry implied by the forecast output without imposing a parametric form.

These empirical predictive distributions are then used as inputs for uncertainty propagation through the deterministic SDG Indicator 11.3.1 transformations. Using Monte Carlo simulation, repeated draws are taken from the predictive distributions of the endpoint BUA and Pop values, and each draw is transformed through the SDG Indicator 11.3.1 formulae to obtain corresponding realizations of LCR, PGR, and LCRPGR for the forecast interval. Repeating this process yields empirical predictive distributions for the derived indicator components.

The propagation procedure follows the same replicate-wise logic established in Chapter 4. That is, uncertain endpoint inputs are repeatedly propagated through the SDG Indicator 11.3.1 formulae to generate empirical distributions for each derived quantity. The main difference is that, in the present chapter, uncertainty arises from forecast predictive distributions rather than from retrospective sampling-based uncertainty. For each Monte Carlo replicate, one joint draw of endpoint BUA and Pop is generated for a given locality and forecast horizon, transformed through the SDG Indicator 11.3.1 formulae, and stored. The resulting empirical distributions are then summarized using median estimates and percentile-based uncertainty intervals.

As emphasized in Chapter 4, the generation of Monte Carlo inputs must account not only for marginal predictive uncertainty but also for the dependence structure between the forecast uncertainties of the input variables. Sampling independently from the marginal predictive distributions would ignore potential co-variation in forecast uncertainty and could therefore distort the propagated uncertainty of LCR, PGR, and LCRPGR. The dependence modeling step used for this purpose is described in the following section.

5.3.13 Dependence Modeling of Forecast Uncertainties

The predictive uncertainty produced by the joint-target forecasting model is expressed in marginal form through separate predictive quantiles for each forecast variable. Accordingly, an additional dependence model is required to reconstruct a plausible joint distribution for uncertainty propagation. Copulas provide a suitable framework because they link univariate marginal distributions to a joint distribution while preserving the marginals themselves (Nelsen, 2006). The use of copulas is advantageous here because the predictive distributions of BUA and Pop are represented nonparametrically via empirical quantiles and monotone interpolation, and their joint behavior cannot be assumed to follow a convenient parametric form, such as multivariate normality. In this setting, the empirical CDFs define the predictive uncertainty of the endpoint variables individually, while the copula specifies how these uncertainties are coupled for joint sampling, preserving the co-occurrence structure implied by the forecast residuals.

In this study, the dependence structure required for copula modeling is inferred from forecast residuals obtained under out-of-sample predictive conditions, on the premise that correlated forecast errors provide empirical evidence of how uncertainty in BUA and Pop covaries under realistic prediction settings. This approach rests on the assumption of dependence stationarity; specifically, the copula structure captured from out-of-sample residuals is assumed to remain time-invariant over the forecast horizon.

For the relevant forecast horizons and endpoint years, the residual series are first assembled into paired observations in the bivariate case, or into higher-dimensional tuples when more than two forecast inputs are modeled jointly. The empirical CDF of each residual series is then used to transform the residuals to the unit interval, thereby replacing each residual value by its empirical percentile position. This transformation removes scale differences across variables and produces pseudo-observations in copula space for copula fitting. By contrast, the empirical predictive CDFs constructed from the forecast quantiles are used later during Monte Carlo propagation to transform copula-generated uniform samples into joint draws of BUA and Pop on their original scales.

Candidate copula models are then fitted to the pseudo-observations. Their parameters are estimated by maximum pseudo-likelihood, with the marginals treated as fixed and represented by their empirical CDFs, so that estimation is confined to the parameters governing dependence, including its strength, asymmetry, and, where relevant, tail dependence (Genest et al., 1995; Joe, 2014). For each fitted candidate, the maximized pseudo-log-likelihood is used to summarize the goodness of fit. The Akaike Information Criterion (AIC) is then calculated from this fit measure, together with a penalty for the number of estimated dependence parameters (Akaike, 1974). In this way, model selection balances fit to the observed dependence structure against model complexity, so that more flexible copulas are preferred only when they provide a sufficiently better description of the pseudo-observations.

Several candidate copula specifications are considered, including elliptical copulas, specifically the Gaussian and Student- t copulas, and vine copulas for higher-dimensional settings. Elliptical copulas provide a parsimonious representation of symmetric dependence. In contrast, vine copulas represent multivariate dependence through a sequence of linked bivariate copulas and allow different dependence forms to be assigned to different variable pairs (Aas et al., 2009; Bedford & Cooke, 2001, 2002; Jiang et al., 2015). Candidate pair-copula families within the vine construction include Gaussian, Student- t , Clayton, Gumbel, and Frank copulas. To balance flexibility and parsimony, truncated vine specifications are also considered, in which higher-order conditional dependencies above a specified tree level are replaced by the independence copula (Nagler, 2025). Truncated vine constructions have been shown to provide a practical balance between flexibility and parsimony in applications involving finite samples and Monte Carlo-based uncertainty propagation (e.g., Huang et al., 2020; Jiang et al., 2015; Yang et al., 2026).

Among the admissible candidate models, the copula specification with the lowest AIC is selected. The selected copula is then used to generate dependent Monte Carlo draws at the forecast endpoints. Specifically, in each replicate, a vector of dependent uniform variates is first sampled from the fitted copula. These dependent uniform draws are subsequently mapped through the inverse empirical CDFs of the corresponding forecast variables, yielding joint samples on the original BUA and Pop scales. In this way, the Monte Carlo inputs preserve both the marginal predictive uncertainty of each endpoint variable and the dependence structure inferred from out-of-sample forecast residuals. These joint samples are then propagated through the SDG Indicator 11.3.1 formulae to obtain predictive distributions of LCR, PGR, and LCRPGR.

5.3.14 Probabilistic LUE Classification

Probabilistic LUE classification is performed by mapping the propagated LCR, PGR, and LCRPGR distributions to LUE classes using the same probabilistic classification workflow introduced in Chapter 4. To support the interpretation of regime stability under forecast uncertainty, the empirical probability of the dominant LUE class is interpreted as a classification confidence measure. This continuous probability is discretized into five ordinal confidence categories, namely very low (< 0.50), low ($0.50\text{--}0.65$), moderate ($0.65\text{--}0.80$), high ($0.80\text{--}0.95$), and very high (> 0.95). These categories distinguish cases with no clear majority support from those with weak, moderate, strong, and near-deterministic regime dominance, and provide an interpretable indication of how sensitive LUE classification is to uncertainty in forecasted BUA and Pop.

To evaluate whether forecast-based probabilistic LUE classification remains decision-relevant, its agreement with observed LUE classification is assessed on the evaluation set where reference outcomes are available. Here, the predicted class is defined as the dominant regime, i.e., the class with the highest empirical probability mass under the propagated indicator distribution. The agreement between predicted and reference classes is evaluated using a confusion matrix approach, which cross-tabulates predicted and reference classes to quantify classification accuracy (Congalton, 1991). From each confusion matrix, producer's accuracy (PA), user's accuracy (UA), and overall accuracy (OA) are derived to assess class-wise recall, precision, and total agreement, respectively.

5.4 Global Application to Urban Centers

This section operationalizes the workflow described in Section 5.3 through a global-scale application to 8,478 urban centers (UCs) worldwide. The workflow was implemented both for out-of-sample evaluation on observed later time steps and for forward forecasting of BUA and Pop to 2025 and 2030.

In this application, the forecasting models were trained on the available historical BUA–Pop trajectories and used to generate uncertainty-aware forecasts of future BUA and Pop. The uncertainty considered here, therefore, pertains to the forecasted values themselves, rather than to uncertainty already embedded in the historical input series. These forecast uncertainties were subsequently propagated through the SDG Indicator 11.3.1 transformations using the framework developed in Chapter 4. In this way, the forecasting analysis remained methodologically consistent across architectures while reusing the same uncertainty propagation logic in a forward-looking setting.

5.4.1 Data Source

The global application of the forecasting framework utilized historical BUA and Pop time series derived from the Global Human Settlement–Urban Centre Database (GHS-UCDB 2025) (European Commission et al., 2024). The spreadsheet version of the GHS-UCDB 2025 was downloaded from the European Commission Joint Research Centre Data Catalogue.¹⁸ GHS-UCDB 2025 provides globally

¹⁸<https://data.jrc.ec.europa.eu/dataset/1a338be6-7eaf-480c-9664-3a8ade88cbcd> (accessed December 30, 2024).

harmonized, multitemporal information for UCs worldwide, making it well-suited to large-scale forecasting for SDG Indicator 11.3.1 applications. UCs in the database were delineated according to the Degree of Urbanisation (DEGURBA) methodology, which defines a UC as a contiguous built-up cluster with at least 50,000 inhabitants and a population density of at least 1,500 inhabitants per km² (European Commission et al., 2024). The multitemporal dataset comprises ten observation epochs at 5-year intervals from 1975 to 2020 for each UC. In addition to BUA and Pop, each UC record includes contextual attributes such as country, SDG region, and World Bank (WB) income classification. The BUA and Pop attributes were derived from the corresponding 100 m GHS-BUILT-S and GHS-POP layers of the GHSL Data Package 2023 (European Commission, 2023), from which the UC-level values reported in GHS-UCDB 2025 were aggregated.

The original dataset contained 11,422 UCs. Before model development, a preliminary quality check was performed to improve the temporal consistency of the time series used for forecasting. Specifically, annual change rates over the 5-year interval for BUA and Pop were calculated for each UC from 1975–2020. Inspection of the resulting frequency distributions revealed strong skewness and extreme values, which could distort model training and increase the risk of overfitting. To reduce the influence of such anomalous trajectories, mild to extreme outliers were identified and removed using the modified IQR method proposed by Barbato et al. (2011). Any UC with at least one BUA or Pop change rate falling outside the corresponding acceptance range was excluded from the forecasting dataset. This filtering reduced the sample from 11,422 to 8,478 UCs.

It should be noted that the underlying GHS-BUILT-S and GHS-POP products used as sources for the BUA and Pop estimates have been extensively validated in prior GHSL documentation and related studies (e.g., European Commission, 2023; Pesaresi et al., 2024), supporting their suitability for large-scale comparative analysis. However, any residual uncertainty in these historical inputs was not explicitly modeled within the forecasting workflow considered here.

Additionally, it is important to note that the BUA values reported in GHS-BUILT-S, and inherited by GHS-UCDB 2025, correspond to the area occupied by buildings and other roofed structures (European Commission, 2023). This definition is consistent with the built-up concept adopted in the metadata for SDG Indicator 11.3.1. Consequently, the LCR and LCRPGR values derived later in the chapter should be interpreted within this specific BUA definition. This clarification is important because, as shown in Chapter 3, SDG Indicator 11.3.1 outcomes are sensitive to the definition of the built-up input. Explicitly stating the adopted built-up definition helps ensure that the results presented in this study can be properly evaluated and compared by other researchers.

5.4.2 Data Partitioning and Normalization

A geographical train–validation–test split was applied to the filtered dataset to evaluate model generalization across spatially distinct UCs. The split followed a stratified random sampling design based on paired UN SDG region and WB income classification to ensure balanced representation of diverse regional and socioeconomic contexts in each partition. Each UC was assigned to one of these combined strata, from which approximately 70% were randomly allocated to the training set (6,016 UCs), and the remaining UCs were divided approximately equally between the validation (1,263

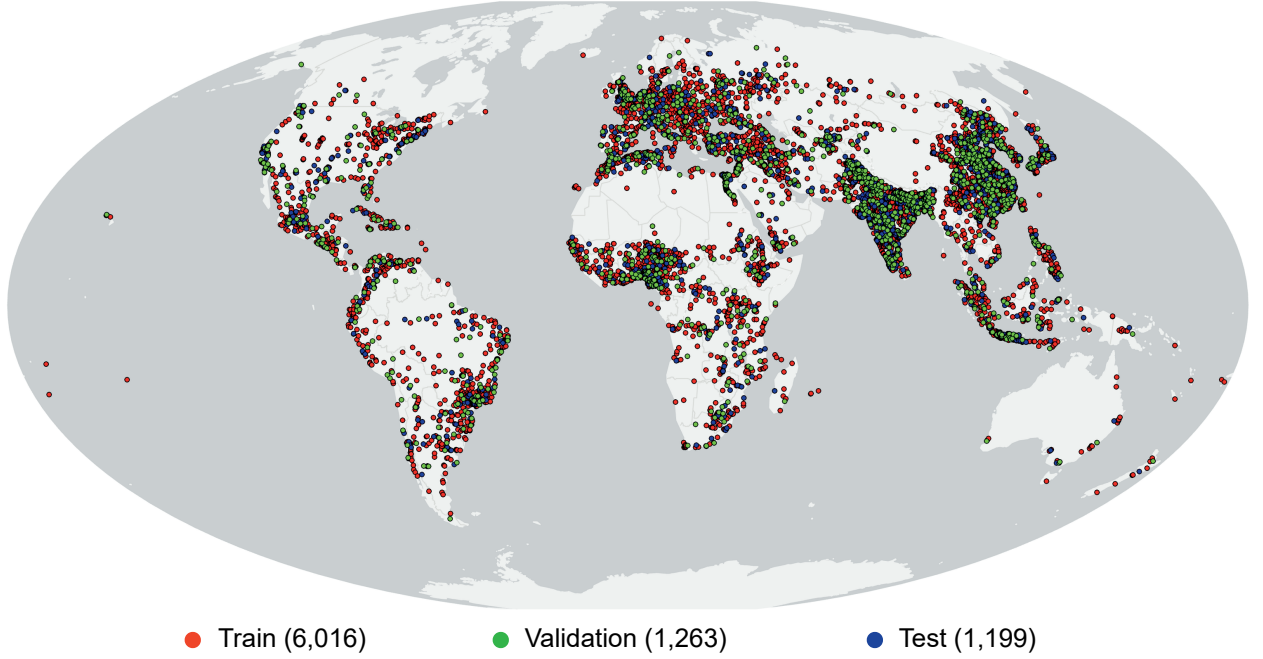


Figure 5.1: Geographical distribution of urban centers included in the study, showing the spatial allocation of the train, validation, and test sets used in model development. Basemap credit: Esri.

and test sets (1,199) (Figure 5.1). For strata containing fewer than three UCs, at least one case was assigned to the training set to ensure that all region–income group combinations were represented during model development.

The empirical implementation used an input length of $L = 8$ and a forecast horizon of $H = 2$. This implementation is consistent with jointly forecasting BUA and Pop for the 2025–2030 period later. Accordingly, for model training and evaluation, the historical input segment $\mathbf{Z}_{i,t}^{\text{in}}$ spanned the eight time steps from 1975 to 2010, while the future output segment $\mathbf{Z}_{i,t}^{\text{out}}$ spanned the two subsequent time steps, corresponding to 2015 and 2020. These latter two epochs formed the out-of-sample evaluation horizon for assessing forecasting performance.

A temporal split was not adopted beyond this input–output construction because the available time series are short. Further dividing the ten-step trajectories into separate temporal subsets for training, validation, and testing, as mostly done in the DL forecasting literature, would substantially reduce the historical context available for learning and weaken the basis for multi-step forecasting. Instead, partitioning was performed geographically at the UC level, while the sequence construction itself preserved the chronological order of prediction from earlier to later epochs.

Within each partition, the BUA–Pop time series were converted into supervised input–output pairs following the global joint-target sequence-to-sequence formulation. Thus, the training, validation, and test sets remained mutually exclusive with respect to UCs, while each UC contributed one sequence pair of the form $L \rightarrow H$, i.e., eight historical time steps used to predict the next two time steps. This design preserved the geographic independence of the evaluation units while enabling pooled learning across multiple localities.

For the DL models, BUA and Pop values were transformed and normalized to improve numerical stability and comparability across UCs. Because both variables exhibited strong positive skewness

and large magnitude differences, a logarithmic transformation was first applied to reduce scale disparity and stabilize variance. The transformed values were then standardized using z-score normalization. To prevent leakage from the forecast horizon, the normalization parameters were estimated exclusively from the training set and using only the historical input portion of the sequences, i.e., the $L = 8$ time steps spanning 1975–2010. These same parameters were then applied unchanged to normalize the corresponding future values in the $H = 2$ evaluation horizon, 2015 and 2020, as well as the validation and test sequences. In this way, the scaling procedure relied solely on information available at the model development stage and excludes distributional information from the future targets.

In contrast, the ES baseline was fitted using the original untransformed BUA and Pop series, consistent with its classical statistical formulation.

5.4.3 DL Model Design and Training

This section describes how the DL architectures introduced in Section 5.3.6 were instantiated for the global joint BUA–Pop forecasting task. Given the very short length of the input sequences, the candidate architectures were deliberately designed with relatively shallow structures. This approach was chosen to mitigate the high risk of overfitting, as a model with excessive capacity and a large number of parameters would be poorly constrained by the limited temporal data and would struggle to generalize (Goodfellow et al., 2016; Hastie et al., 2009).

All DL models were implemented and optimized in Python 3.10 using the TensorFlow (v2.15.1) and Keras (v2.15.0) framework, and were trained using the Adam optimizer (Kingma & Ba, 2015) with an initial learning rate of 1×10^{-3} on a Central Processing Unit (CPU)-only setup.

All architectures were configured to receive as input a fixed-length sequence of eight historical observations of BUA and Pop and to produce 12 outputs corresponding to the 2.5th, 50th, and 97.5th quantile forecasts for both target variables at each of the two forecast years. Hyperparameter configuration and selection were determined through systematic hyperparameter optimization with Optuna (Akiba et al., 2019) and the Hyperband pruning algorithm (Li et al., 2018) to efficiently explore architectural and training configurations. This pruning strategy automatically halted poorly performing trials based on intermediate validation results, allowing computational resources to focus on the most promising candidates. Tuning for each architecture followed a two-stage process. In Stage A, a broad search across 120 trials, with ≤ 200 epochs and early stopping after 10 epochs, was conducted to identify promising configurations. In Stage B, the top 10 candidate configurations were retrained more extensively, with ≤ 400 epochs and early stopping after 20 epochs, to refine convergence.

To ensure stable optimization, gradient clipping (Pascanu et al., 2013) with a threshold of 1 and an adaptive learning rate scheduler, reduction factor = 0.5, minimum rate = 1×10^{-6} , were applied. Gradient clipping is a technique used to prevent the exploding gradient problem by limiting the magnitude of gradients during training/optimization (Goodfellow et al., 2016). If a gradient’s norm exceeds a predefined threshold, it is rescaled to that threshold, ensuring that model weight updates remain manageable and the training process is more stable. The scheduler then dynamically adjusts

the learning rate, reducing it by a factor of 0.5 whenever a performance metric stops improving, to help the model settle into a precise minimum while ensuring the rate does not fall below 1×10^{-6} , thereby stabilizing convergence.

All models were trained by minimizing the multi-horizon quantile, or pinball, loss (Eq. 5.24) calculated across forecast steps, target variables, and quantile levels. Model fitting and tuning were performed using the training set. The validation set was utilized to monitor generalization, control early stopping, adjust learning rates, and select the best configuration based on the lowest average pinball loss. Among the models retained in Stage B, the configuration that achieved the best validation performance was retained as the final model for subsequent analyses. The hyperparameter search space for each model architecture is outlined in Table A1 (Appendix). The final tuned designs of the individual architectures are described in the following subsections. The batch size for all models was set to 32.

5.4.3.1 1D CNN

The 1D CNN architecture followed a linear stack design composed of two sequential Conv1D layers. The first layer used 128 filters, a kernel size of 3, a stride of 1, a dilation rate of 1, and L2 regularization with a weight of 1×10^{-6} . Rectified Linear Unit (ReLU) activation and causal padding were applied. The second Conv1D layer used the same number of filters, kernel size, stride, activation function, padding, and regularization, but with a dilation rate of 2. The resulting feature maps were then flattened into a vector and passed to a fully connected hidden layer with 128 units, using ReLU activation and L2 regularization with a weight decay of 1×10^{-6} . The network ended with a linear fully connected output layer containing 12 neurons, corresponding to the multi-horizon quantile forecasts defined for the global joint BUA–Pop task.

5.4.3.2 LSTM

The LSTM architecture began with a single unidirectional LSTM layer with 128 units. This recurrent layer processed the full input sequence and returned only the final hidden state, thereby encoding the historical BUA–Pop sequence into a fixed-length representation. L2 regularization of 1×10^{-6} was applied to the kernel weights. The resulting vector was then passed to a fully connected hidden layer with 50 units, using ReLU activation and L2 regularization with a weight decay of 1×10^{-6} . The network ended with a linear fully connected output layer that produced the required multi-horizon quantile forecasts for the joint BUA–Pop task.

5.4.3.3 GRU

The GRU architecture was structurally similar to the LSTM architecture but replaced the LSTM layer with a single unidirectional GRU layer with 128 units. This recurrent layer processed the full input sequence and returned only the final hidden state, thereby encoding the historical BUA–Pop sequence into a fixed-length representation. L2 regularization of 1×10^{-6} was applied to the kernel weights. The resulting vector was then passed to a fully connected hidden layer with 64 units, using

ReLU activation and L2 regularization with a weight decay of 1×10^{-6} . The network ended with a linear fully connected output layer that produced the required multi-horizon quantile forecasts for the joint BUA–Pop task.

5.4.3.4 1D CNN+LSTM

The 1D CNN+LSTM architecture was designed with a temporal feature extraction block composed of two sequential Conv1D layers. Each convolutional layer used 128 filters, a kernel size of 5, a stride of 1, a dilation rate of 1, ReLU activation, causal padding, and L2 regularization of 1×10^{-6} . The resulting sequence of feature maps was then passed to a single unidirectional LSTM layer with 32 units, which processed the extracted temporal features and returned only the final hidden state. L2 regularization of 1×10^{-6} was applied to the kernel weights of the LSTM layer. The resulting vector was passed to a fully connected hidden layer with 128 units, using ReLU activation and L2 regularization with a weight decay of 1×10^{-6} . The network ended with a linear fully connected output layer that produced the required multi-horizon quantile forecasts for the joint BUA–Pop task.

5.4.3.5 1D CNN+GRU

The 1D CNN+GRU architecture began with a temporal feature-extraction block composed of two sequential Conv1D layers. Each convolutional layer used 64 filters, a kernel size of 3, a stride of 1, a dilation rate of 1, ReLU activation, causal padding, and L2 regularization of 1×10^{-6} . The resulting sequence of feature maps was then passed to a single unidirectional GRU layer with 128 units, which processed the extracted temporal features and returned only the final hidden state. L2 regularization of 1×10^{-6} was applied to the kernel weights of the GRU layer. The resulting vector was then passed to a fully connected hidden layer with 32 units, using ReLU activation and L2 regularization with a weight decay of 1×10^{-6} . The network ended with a linear fully connected output layer that produced the required multi-horizon quantile forecasts for the joint BUA–Pop task.

5.4.3.6 Transformer

A standard encoder–decoder Transformer was employed, in which the encoder processed the historical BUA–Pop input sequence. During training, teacher forcing was used in the decoder: the observed forecast sequence was supplied in shifted form, so that the prediction for the second forecast step was learned using the observed first forecast-step BUA–Pop values as decoder input. This allowed the decoder to learn the conditional structure across the forecast horizon while producing multi-step outputs. During inference, however, true future values were unavailable; therefore, the decoder relied on the implemented inference-time decoding strategy rather than observed forecast targets.

Both encoder and decoder inputs were first projected to a model dimension of 128 through fully connected layers with ReLU activation and L2 regularization of 1×10^{-6} . Learned positional encodings were then added to preserve temporal order. The encoder consisted of a single Transformer block with pre-normalization, 8 attention heads, and a feedforward dimension of 128. No projection dropout or attention dropout was applied, while the feedforward sublayer used a dropout rate of 0.1. The decoder had the same one-block structure and included causal self-attention, cross-attention to the encoder output, and a feedforward sublayer. The resulting decoder output sequence was then passed through a prediction head comprising a hidden fully connected layer with 128 units and ReLU activation, followed by a final linear output layer applied at each decoder step. The per-step outputs were subsequently flattened into a single prediction vector corresponding to the required multi-horizon quantile forecasts for the joint BUA–Pop task.

5.4.4 Deep Ensembling

Using the tuned hyperparameters, each DL model was retrained 30 times, producing 30 independent and diverse model replicates. The choice of 30 replicates balanced computational efficiency with ensemble diversity, ensuring that the ensemble forecast remained robust against rare stochastic convergence failures. This choice was also feasible given the moderate size of the model architectures.¹⁹

Each replicate was applied to the test set to generate forecasts for 2015 and 2020 for every UC, yielding its own predictive distribution represented by the 2.5th, 50th, and 97.5th percentiles. The final ensemble forecasts were obtained by taking the median across the 30 replicates at each quantile level. Since the models were trained to estimate specific quantiles, this aggregation method aligned well with the training objective. In this median-of-quantiles scheme, the median of the replicate medians served as the ensemble point estimate, while the medians of the replicate 2.5th and 97.5th percentiles defined the 95% prediction interval. This aggregation approach provided robust point and interval estimates, reduced sensitivity to outlier replicates, and formed the predictive distributions used for downstream uncertainty quantification.

5.4.5 Baseline Modeling Using ES

In this study, BUA and Pop for each UC were modeled independently as separate univariate forecasting problems within the exponential smoothing (ES) family, specifically using the error–trend–seasonal (ETS) framework (Hyndman & Athanasopoulos, 2018). The ETS formulation expresses ES models in a unified likelihood-based framework, allowing the underlying model structure, parameter values, and forecast uncertainty to be estimated directly from the data. Rather than adopting a single fixed specification, 10 ETS variants were developed for each variable, representing different combinations of error, additive or multiplicative; trend, none, additive, or multiplicative;

¹⁹For example, the Transformer model, which had the largest number of trainable parameters (283,526), required only about 10 minutes to train a single replicate in CPU-only mode (2.2 GHz, 12 CPUs, 16 GB RAM), and roughly 15 seconds for a single replicate to perform inference on the test set, equivalent to approximately 12.5 ms per UC.

and damping, with or without. Seasonality was excluded since both BUA and Pop exhibit long-term monotonic trends rather than recurring cyclical patterns. Model fitting was implemented in Python 3.10 using the `ETSModel` class from the `statsmodels` library, which performs full maximum-likelihood estimation of the level, trend, and damping parameters and provides analytical forecast variance calculations. To maintain comparability with the DL workflow, historical observations from 1975 to 2010 were likewise used for model fitting, while 2015 and 2020 served as evaluation years.

To objectively select the best model variant for each variable for each UC, a multi-criteria model selection procedure was applied following Burnham and Anderson’s (2004) model selection framework. For each fitted model, the Akaike Information Criterion corrected for small-sample bias (AICc) (Hurvich & Tsai, 1989) was calculated to quantify goodness of fit and model parsimony. The relative likelihood of each model was then assessed through its ΔAICc value, defined as the difference between a model’s AICc and the minimum AICc among all variants for the same UC and variable. Models with $\Delta\text{AICc} \leq 2$ were considered statistically equivalent and therefore retained for further comparison. Among these top models, the MAE over the 2015–2020 evaluation period was used as the secondary selection criterion to identify the most accurate forecasting model. In cases where models exhibited both equivalent ΔAICc and identical MAE values, the simplest specification was preferred following the principle of parsimony. Here, the simplest specification is defined as the one with the fewest estimated components, for example, without damping or with additive rather than multiplicative error or trend.

5.4.6 Model Evaluation and Selection

All forecasting approaches were evaluated on the held-out test set (Section 5.4.2) using the 2015 and 2020 forecasts for each UC. For each candidate DL architecture, BUA and Pop forecasts for these two years were generated from the 30 ensemble replicates described in Section 5.4.4. Ensemble-level quantile forecasts were then obtained using the median-of-quantiles aggregation rule, that is, the median across replicate predictions at each quantile. Accordingly, for each UC, variable, and forecast year, the ensemble point forecast was defined by the aggregated median prediction. The corresponding lower and upper uncertainty bounds were defined by the aggregated 2.5th- and 97.5th-percentile predictions, respectively. For the ES baseline, BUA and Pop forecasts were generated directly from the fitted univariate models together with their associated 95% prediction intervals.

Model evaluation followed the procedure described in Section 5.3.10. For each model, target variable, and UC, horizon-specific forecast errors were first calculated for 2015 and 2020. These were then summarized within each UC, yielding one locality-level value per model and variable for MAE, MAPE, PIAW, and Winkler Score, obtained by averaging across the two forecast years. In contrast, PICP was calculated as an overall coverage measure across all UC–horizon combinations for each model and variable. The resulting performance summaries were then used in the statistical comparison, model ranking, and selection procedure described in Section 5.3.11 to identify the forecasting model for downstream uncertainty propagation in SDG Indicator 11.3.1.

5.4.7 Forecast Uncertainty Propagation to SDG Indicator 11.3.1 and Probabilistic LUE Classification on the Test Set

To implement the procedure described in Section 5.3.12 under known historical conditions, the selected joint-target forecasting model was retrained using the same tuned architecture and hyperparameter configuration, but with an expanded set of predictive quantile outputs. Consistent with the deep-ensembling strategy described in Section 5.4.4, this retraining was repeated 30 times to generate an ensemble of independent model replicates. While the original evaluation configuration produced only the 2.5th, 50th, and 97.5th percentiles for each target variable and forecast step, this was considered too coarse for reconstructing the predictive distributions required for Monte Carlo propagation. Therefore, the retrained ensemble models were configured to predict the quantile set 0.025, 0.10, 0.25, 0.50, 0.75, 0.90, and 0.975 for BUA and Pop at both forecast steps. The final ensemble quantile forecasts were obtained by taking the median across the 30 replicate predictions at each quantile. These denser quantile outputs enabled the construction of more realistic empirical cumulative distribution functions (CDFs) that better preserved asymmetry and uncertainty variation across the distribution. Using these ensemble-aggregated quantile forecasts, empirical predictive distributions of BUA and Pop were constructed for the test-set endpoint years 2015 and 2020.

Dependence between forecast uncertainties in the endpoint variables was then incorporated following the copula-based procedure defined in Section 5.3.13. For this purpose, forecast residuals for BUA_{2015} , BUA_{2020} , Pop_{2015} , and Pop_{2020} were extracted from the validation set, which served as the basis for dependence estimation. Although the validation set was also used during model development for training, monitoring, and hyperparameter selection, it remained the most appropriate source of residuals for copula fitting because it reflects out-of-sample predictive behavior under conditions comparable to those encountered during model development while remaining independent of the final test-set evaluation. Training-set residuals were not used because they would tend to understate predictive uncertainty, whereas test-set residuals were excluded to avoid information leakage into the subsequent SDG Indicator 11.3.1 consistency assessment.

Residuals were organized as multivariate observations over the endpoint variables and transformed into pseudo-observations on the unit interval. Copula fitting, dependence estimation, and model selection were then carried out following the procedure described in Section 5.3.13. The selected copula model was subsequently used to couple the empirical predictive distributions of BUA and Pop at the forecast endpoints for Monte Carlo uncertainty propagation.

Because the forecast-based uncertainty propagation in this chapter differs from the Monte Carlo setup used in Chapter 4, the simulation size was reassessed rather than adopted directly. Convergence was evaluated on a representative high-uncertainty UC in the test set, which was selected to reflect three stress conditions: small absolute projected PGR, wide forecast uncertainty bounds for BUA and/or Pop, and cases combining both. Based on this analysis, $N = 250,000$ was found sufficient to obtain stable estimates of the reported indicator summaries and probabilistic regime assignments.

For each UC in the test set, joint random samples were drawn from the coupled BUA–Pop predictive distributions. LCR, PGR, and LCRPGR were then calculated for each Monte Carlo replicate

using the forecasted endpoint values, producing empirical distributions for the three indicators. The median and 95% prediction interval for each distribution were extracted as uncertainty-aware projected estimates, which were subsequently mapped to probabilistic LUE classes using the workflow defined in Sections 4.2.5 and 5.3.14.

5.4.8 Evaluation of Projected SDG Indicator 11.3.1 Estimates and Probabilistic LUE Classification

The projected indicator distributions were evaluated against the observed deterministic indicator values derived from the actual 2015 and 2020 endpoint data. Interval adequacy was assessed by determining whether the observed values fell within the corresponding 95% prediction intervals. Predictive agreement was then assessed by comparing projected medians of LCR, PGR, and LCRPGR with their observed values across UCs using median error, median absolute error, and Spearman rank correlation. These metrics were chosen because they provide robust summaries of central predictive agreement and cross-UC association, while being less sensitive than mean-based measures to skewed error distributions and extreme values, particularly for the ratio-based LCRPGR indicator. To further examine how forecast uncertainty was transmitted into the derived SDG Indicator 11.3.1 components, Spearman rank correlations were additionally calculated between UC-specific input prediction-interval widths and the propagated interval widths of LCR, PGR, and LCRPGR. Finally, the dominant predicted LUE class was compared with the reference class derived from the observed endpoint values. As in Chapter 4, realizations assigned to R4–R6 were grouped as *Efficient Built-up Reduction under Population Growth*, and those assigned to R12–R14 were grouped as *Inefficient Built-up Expansion under Population Decline*; all other regimes were retained individually. This grouping preserves the shared directional interpretation of each set while simplifying probabilistic reporting. Classification agreement was evaluated using the confusion-matrix framework described earlier in Section 5.3.14, from which producer’s accuracy, user’s accuracy, and overall accuracy were derived.

5.4.9 Forecast-Based Projection of SDG Indicator 11.3.1 for 2025–2030

The selected forecasting model was subsequently applied to generate probabilistic forecasts of BUA and Pop for 2025 and 2030. BUA and Pop covering 1985–2020, i.e., eight historical time steps before 2025 from the GHS-UCDB 2025, were used as inputs for inference to obtain multi-quantile probabilistic forecasts for the target years. The dependence between BUA and Pop forecast uncertainties was represented using the same copula specification as in Section 5.4.7. It is assumed that the dependence structure inferred from historical forecast residuals remains stable over the short-term projection horizon and therefore provides a reasonable approximation of how forecast uncertainties in BUA and Pop are likely to co-vary in future periods.

These joint predictive distributions were then propagated to SDG Indicator 11.3.1 components following the same Monte Carlo-based uncertainty propagation framework described earlier, with $N = 250,000$. For the 2020–2025 interval, the observed 2020 BUA and Pop values were treated as fixed inputs and combined with simulated 2025 values, whereas for the 2025–2030 interval,

simulated values were used for both endpoints. The resulting indicator distributions were then used for probabilistic LUE classification and confidence assignment following the same procedure defined earlier, including the grouped regime reporting adopted above.

5.5 Results

5.5.1 Forecasting Performance on the 2015–2020 Evaluation Horizon

The following subsections present the out-of-sample test set forecasting results for the 2015–2020 evaluation horizon based on the ensemble predictions of the candidate models. Results are reported separately for BUA and Pop, focusing on both point-forecast accuracy and prediction-interval performance, before the formal statistical comparison and final model selection are presented in Section 5.5.3.

5.5.1.1 BUA Point Forecast Accuracy and Prediction-Interval Performance

Table 5.2 shows that, for BUA, the global DL models substantially outperformed the ES baseline in both point forecast accuracy and interval quality. Across the 1,199 UCs, the DL models achieved very similar and consistently low point forecast errors, with median two-horizon MAE ranging from 0.04 to 0.05 km² and median MAPE ranging from 0.98% to 1.12%. In particular, five of the six DL architectures attained the same rounded median MAE of 0.04 km², while 1D CNN was only marginally higher at 0.05 km². This indicates that differences in central absolute-error performance among the DL models were minimal. GRU nevertheless emerged as the strongest point-forecast model overall because, while sharing the same median MAE as most of the other DL models, it attained the lowest median MAPE at $0.98 \pm 0.63\%$. By contrast, ES performs markedly worse, with a median MAE of 0.23 ± 0.17 km² and a median MAPE of $4.87 \pm 2.63\%$, confirming that the classical baseline was much less effective for the present two-horizon, multi-locality forecasting task.

In terms of prediction-interval performance, the Transformer produced the narrowest intervals, with the lowest median PIAW of 0.28 km² and the lowest median Winkler score of 0.30. However, its PICP of 0.94 fell slightly below the nominal 0.95 level, indicating mild under-coverage. By contrast, 1D CNN attained a PICP of 0.95, matching the nominal level, whereas 1D CNN+GRU yielded the highest coverage at 0.97, but with wider intervals, as reflected in a median PIAW equal to 0.33 km², and a higher median Winkler score of 0.34. This pattern indicates a calibration–sharpness trade-off across the DL models, with sharper intervals generally associated with slightly reduced coverage.

These findings suggest that the global DL models provided a clear performance advantage over the ES baseline for BUA forecasting. GRU showed slightly better point-forecast performance overall, whereas interval results were differentiated mainly by the balance between sharpness and calibration rather than by a single uniformly superior model. The small spread in MAE, MAPE, PIAW, and Winkler score across the DL models suggests that all global DL architectures considered here were broadly effective for forecasting BUA, with differences emerging mainly in whether performance

was judged more strongly in terms of percentage error, interval sharpness, or calibration. The poor performance of ES across MAE, MAPE, PIAW, PICP, and Winkler score further indicates that a simple classical extrapolation approach was insufficient for capturing the diversity of BUA trajectories across the global set of UCs.

Table 5.2: Two-horizon (2015 and 2020) point forecast accuracy and prediction-interval performance for BUA, comparing global DL models with the Exponential Smoothing baseline across 1,199 UCs. Point forecast accuracy is summarized by per-UC two-horizon MAE and MAPE (median $\pm$ MAD), where lower values indicate better performance. Median PIAW gives the typical 95% interval width, PICP gives empirical coverage across both horizons, and median Winkler Score summarizes overall interval quality, with lower values indicating better performance.

Model	Median MAE $\pm$ MAD (km ²)	Median MAPE $\pm$ MAD (%)	Median PIAW (km ²)	PICP	Median Winkler Score
1D CNN	0.05 $\pm$ 0.03	1.12 $\pm$ 0.68	0.32	0.95	0.33
LSTM	0.04 $\pm$ 0.03	1.03 $\pm$ 0.64	0.32	0.96	0.34
GRU	0.04 $\pm$ 0.03	0.98 $\pm$ 0.63	0.32	0.96	0.33
1D CNN+LSTM	0.04 $\pm$ 0.03	1.10 $\pm$ 0.65	0.33	0.96	0.34
1D CNN+GRU	0.04 $\pm$ 0.03	1.01 $\pm$ 0.61	0.33	0.97	0.34
Transformer	0.04 $\pm$ 0.03	1.07 $\pm$ 0.64	0.28	0.94	0.30
Exponential Smoothing	0.23 $\pm$ 0.17	4.87 $\pm$ 2.63	0.51	0.59	2.51

5.5.1.2 Pop Point Forecast Accuracy and Prediction-Interval Performance

For Pop, Table 5.3 shows that the global DL models again outperformed the ES baseline across both point forecast accuracy and overall interval performance. In terms of point forecasts, 1D CNN+LSTM performed most favorably, with the lowest median two-horizon MAE, 2450 ± 1772 persons, and the lowest median MAPE, $2.08 \pm 1.24\%$. The remaining DL models were close behind, with median MAE values ranging from 2482 to 2691 persons and median MAPE values from 2.09% to 2.22%, indicating that differences in central point-forecast accuracy among the DL architectures were modest overall. By contrast, ES performed substantially worse, with a median MAE of 4450 ± 3069 persons and a median MAPE of $3.52 \pm 2.21\%$, confirming that the classical baseline was much less effective for capturing population trajectories across the global evaluation set.

The probabilistic results again indicated a calibration–sharpness trade-off. The Transformer produced the narrowest intervals, with the smallest median PIAW of 13,016 persons and the lowest median Winkler score of 17,227. However, this advantage was offset by a PICP of only 0.86, well below the nominal 0.95 level and indicating substantial under-coverage. In contrast, LSTM, GRU, and 1D CNN+GRU each attained a PICP of 0.95, but with wider intervals and higher Winkler scores. This indicates that the Transformer favored sharper intervals at the expense of empirical coverage, whereas the other DL models provided better-calibrated but less sharp predictive intervals.

These findings suggest that the global DL models provided a clear performance advantage over the ES baseline for Pop forecasting. Among the DL models, differences in point-forecast performance

Table 5.3: Two-horizon (2015 and 2020) point forecast accuracy and prediction-interval performance for Pop, comparing global DL models with the Exponential Smoothing baseline across 1,199 UCs. Point forecast accuracy is summarized by per-UC two-horizon MAE and MAPE (median $\pm$ MAD), where lower values indicate better performance. Median PIAW gives the typical 95% interval width, PICP gives empirical coverage across both horizons, and median Winkler Score summarizes overall interval quality, with lower values indicating better performance.

Model	Median MAE $\pm$ MAD (no. of persons)	Median MAPE $\pm$ MAD (%)	Median PIAW (no. of persons)	PICP	Median Winkler Score
1D CNN	2,482 $\pm$ 1,754	2.09 $\pm$ 1.31	19,641	0.94	20,854
LSTM	2,691 $\pm$ 1,997	2.22 $\pm$ 1.34	21,940	0.95	23,366
GRU	2,645 $\pm$ 1,986	2.16 $\pm$ 1.29	21,988	0.95	23,685
1D CNN+LSTM	2,450 $\pm$ 1,772	2.08 $\pm$ 1.24	20,935	0.94	22,628
1D CNN+GRU	2,568 $\pm$ 1,836	2.12 $\pm$ 1.29	21,301	0.95	23,105
Transformer	2,535 $\pm$ 1,850	2.13 $\pm$ 1.35	13,016	0.86	17,227
Exponential Smoothing	4,450 $\pm$ 3,069	3.52 $\pm$ 2.21	20,498	0.80	31,556

were modest, although 1D CNN+LSTM performed slightly more favorably overall. For prediction intervals, no single model was uniformly optimal across all criteria. The Transformer yielded the narrowest intervals and lowest Winkler score, but its under-coverage limited its reliability. Several other DL models achieved nominal coverage with somewhat wider intervals. Compared with the DL models, ES performed poorly across all metrics, including coverage of 0.80 and a Winkler score of 31,556, indicating that it was not competitive for global population forecasting in this setting. Overall, the results suggest that, as with BUA, the main performance distinction lay between the DL models collectively and the ES baseline, rather than among the individual DL architectures themselves. The relatively small spread in MAE and MAPE across the DL models suggests that model choice for Pop depends less on large differences in point accuracy than on the preferred balance between slightly better central accuracy and better-calibrated prediction intervals.

5.5.1.3 Short- and Medium-Term Forecast Performance

To supplement the two-horizon results, Tables A2–A3 (Appendix) report year-specific performance metrics for 2015, representing the short-term horizon, and 2020, representing the medium-term horizon. These horizon-specific metrics show that all models generally exhibited larger point forecast errors and wider prediction intervals for 2020 than for 2015, reflecting the expected decline in forecast accuracy with increasing lead time. Nevertheless, the relative performance patterns remained broadly consistent with the two-horizon rankings, with the leading models retaining their comparative advantage across both forecast horizons.

5.5.2 Illustrative UC-Level Forecast Examples

Figures 5.2 and 5.3 present representative out-of-sample BUA and Pop forecasts for 2015 and 2020 and provide visual context for the aggregated results in Tables 5.2 and 5.3. They illustrate why the global DL models generally outperformed the ES baseline, while also showing that Pop was more challenging to forecast than BUA.

For BUA, the global DL models generally preserved the direction and overall shape of the historical trajectories more consistently than the ES baseline (Figure 5.2). Across most representative UCs, the DL point forecasts for 2015 and 2020 extended the pre-2010 upward trend and remained visually close to the observed values, particularly in Wuhan, Manchester, and Istanbul. In São Paulo and Istanbul, the contrast with ES was especially evident. Whereas the DL models continued the observed upward trajectory, the ES baseline produced markedly flatter central forecasts that underrepresented the post-2010 increase. Lahore provided a comparatively less clear-cut case, as ES remained competitive as a point forecaster, while some DL models underpredicted the magnitude of BUA growth by 2020. These examples are consistent with the aggregated BUA results, which showed that the DL models achieved lower point-forecast errors than ES across the test set.

The 95% prediction intervals for BUA also exhibited a size-related effect, with narrower bands in UCs with smaller built-up footprints, such as Manchester and Adelaide, and wider bands in UCs with larger built-up footprints, such as São Paulo, New York, and Johannesburg. In several cases, the DL intervals were asymmetric, expanding more strongly in the upper tail. By contrast, the ES intervals in cases such as São Paulo and Istanbul included sharply declining lower bounds that appeared less consistent with the observed continuation of BUA growth.

For Pop (Figure 5.3), the representative examples showed larger variation in forecast behavior across UCs and across models than was observed for BUA. Although the DL models still tended to preserve the overall direction of change better than ES, the point forecasts showed larger dispersion across architectures and more frequent mismatches in the magnitude of continuation. These characteristics were evident in Lahore and Johannesburg, where several DL models underpredicted the observed increase by 2020, and in Wuhan and Istanbul, where some models projected stronger growth than observed. Manchester, New York, and Adelaide exhibited closer agreement between DL point forecasts and observed values. Overall, these examples are consistent with the aggregated results indicating that Pop was more difficult to forecast accurately than BUA.

The 95% prediction intervals from the DL models often fanned out strongly by 2020, especially in Wuhan, São Paulo, Istanbul, New York, and Adelaide, indicating larger uncertainty and reduced sharpness in population forecasts. In several panels, the lower bounds also indicated nearly stagnant or slightly declining paths despite the generally upward historical pattern. On the other hand, the ES baseline again tended to generate flatter central forecasts, particularly in Wuhan, São Paulo, New York, and Adelaide, where the observed Pop series continued to increase after 2010. The ES intervals often appeared less plausible, particularly where the central forecast became overly flat and the lower bound declined markedly despite continued observed growth, as seen most clearly in Wuhan, São Paulo, and Adelaide.

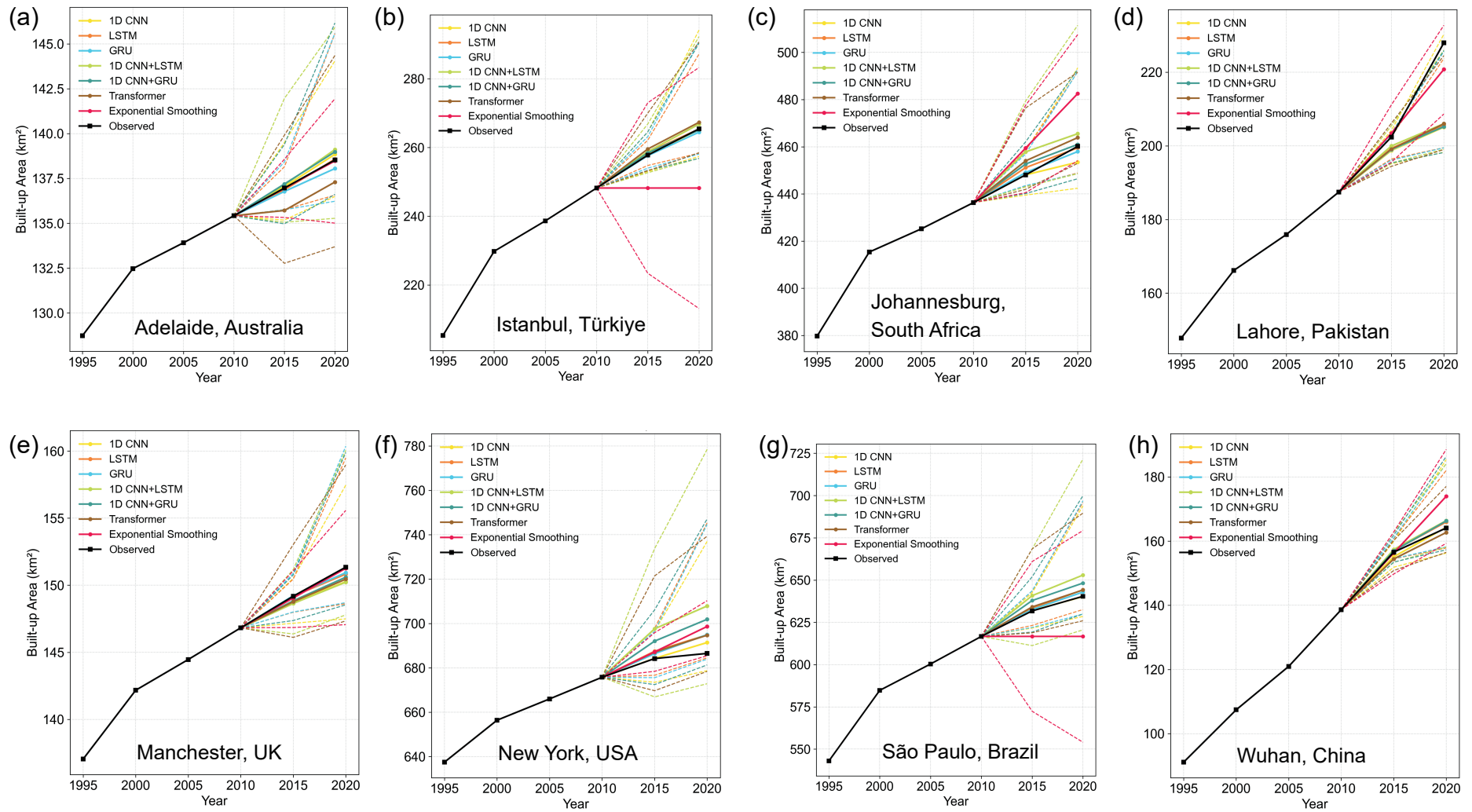


Figure 5.2: Example BUA predictions for 2015 and 2020 from the global deep learning models for representative urban centers in the test set. Deep learning forecasts are shown as medians with corresponding 95% prediction intervals, indicated by dashed lines in matching colors, alongside observed values and the Exponential Smoothing baseline. Models were trained on 1975–2010 data.

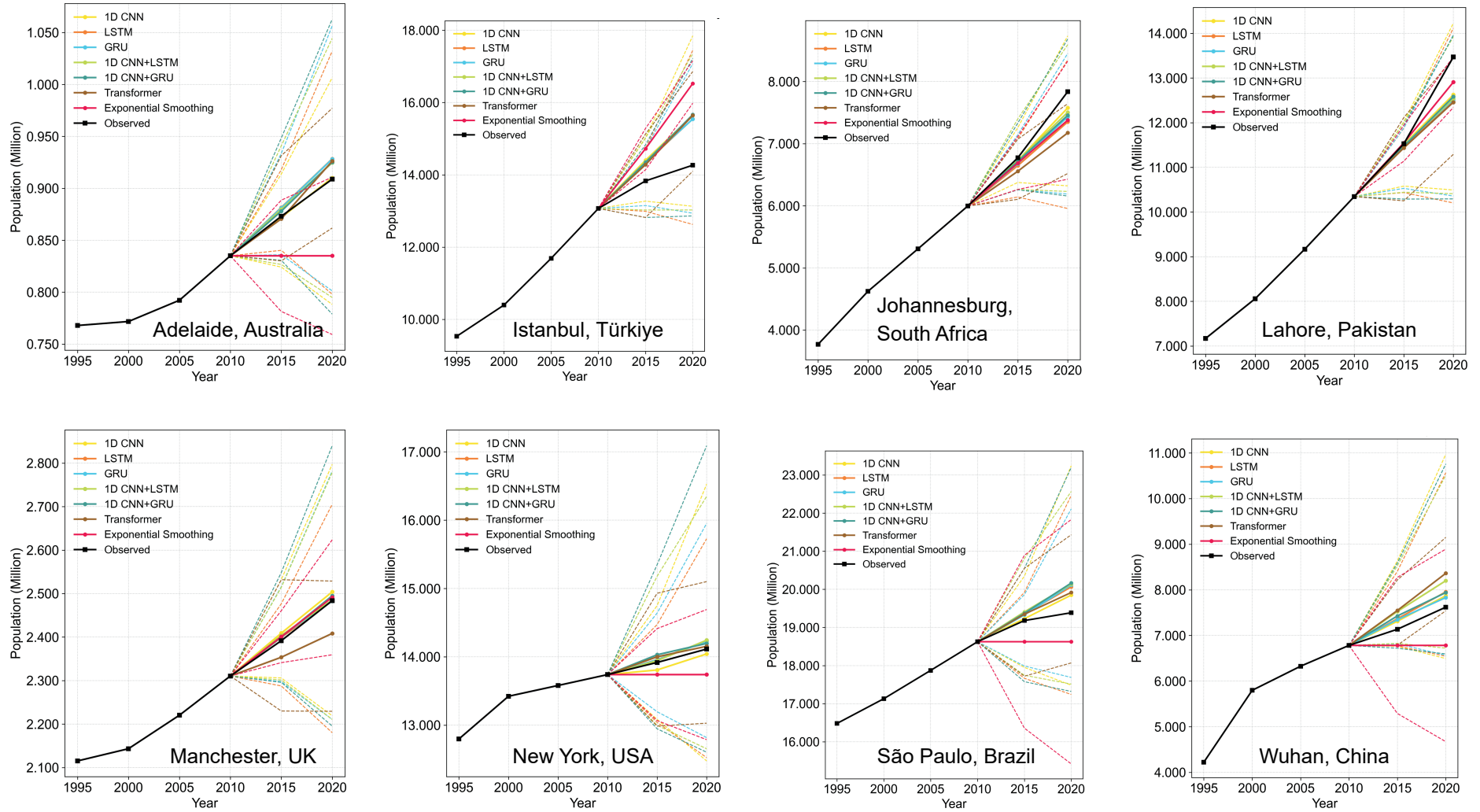


Figure 5.3: Example Pop predictions for 2015 and 2020 from the global deep learning models for representative urban centers in the test set. Deep learning forecasts are shown as medians with corresponding 95% prediction intervals, indicated dashed lines in matching colors, alongside observed values and the Exponential Smoothing baseline. Models were trained on 1975–2010 data.

Overall, the UC-level examples reinforce the performance results by showing that global DL models more often preserve the continuity of observed BUA and Pop trajectories than ES, while also revealing larger uncertainty and heterogeneity in Pop forecasts.

5.5.3 Statistical Comparison and Final Model Selection

For the joint two-horizon point forecasting task for BUA and Pop, the Friedman test indicated that overall model performance was not identical across architectures ($\chi^2 = 16.65, p < 0.05$; Table 5.4). The corresponding mean ranks nevertheless showed a very close ordering among the leading models, with GRU obtaining the best rank (3.376), followed almost identically by 1D CNN+GRU (3.377). This near-equality suggests that the top-performing architectures were broadly similar in overall point-forecast performance. However, the pairwise Holm-adjusted Wilcoxon signed-rank tests (Table 5.4) revealed a selective pattern of significant differences. The top two models, GRU and 1D CNN+GRU, were not significantly different, indicating that the hybrid architecture did not provide a statistically demonstrable improvement over the simpler recurrent formulation. GRU was significantly better than LSTM and 1D CNN+LSTM, while 1D CNN+GRU was significantly better than 1D CNN, Transformer, and 1D CNN+LSTM. The corresponding rank-biserial effect sizes were small, ranging from -0.11 to 0.15 , indicating that although some differences were statistically detectable, the practical separation among the strongest DL candidates remained modest.

Table 5.4: Results of the Friedman test and Holm-adjusted Wilcoxon signed-rank tests for pairwise comparisons of per-urban-center (UC) two-horizon joint standardized errors derived from BUA and Pop MAE across all DL forecasting models. Cells marked with a check (✓) indicate statistically significant differences between the two models at $p < 0.05$ (Holm-adjusted), whereas “×” denotes no significant difference. Models are listed in order of their Friedman mean ranks, from best to worst.

Friedman Mean Rank	Model	GRU	1D CNN+GRU	LSTM	1D CNN	Transformer	1D CNN+LSTM
3.376	GRU	—	×	✓	×	×	✓
3.377	1D CNN+GRU	×	—	×	✓	✓	✓
3.517	LSTM	✓	×	—	×	×	×
3.565	1D CNN	×	✓	×	—	×	×
3.580	Transformer	×	✓	×	×	—	×
3.585	1D CNN+LSTM	✓	✓	×	×	×	—

Given the variable-specific forecasting results reported in Section 5.5.1, the rank-based comparisons supported selecting GRU as the final model for the joint BUA–Pop forecasting task. Although no single architecture was uniformly best across all variables and performance metrics, GRU combined the best overall Friedman rank with strong point-forecast accuracy, including the best BUA point-forecast performance, competitive Pop accuracy, and overall acceptable interval behavior. The Transformer, while favorable in terms of Winkler score, exhibited sub-nominal coverage for Pop, whereas 1D CNN+GRU, despite its nearly identical mean rank, did not significantly outperform

GRU and required a more complex hybrid structure. On this basis, GRU was selected as the final forecasting architecture, as it provided the most balanced and robust overall performance across both BUA and Pop, without evidence that a more complex alternative offered a consistent statistical advantage.

5.5.4 Deep Ensemble Behavior and Stability

With GRU selected as the final architecture for all subsequent analyses, the behavior of its deep ensemble was further examined to assess the stability and reproducibility of predictions across the 1,199 UCs in the test set. This analysis was intended to determine the extent to which random initialization and stochastic optimization affected predictive performance across the 30 independently trained GRU replicates, and to quantify the benefit of aggregating these replicates into a single ensemble forecast. For each replicate, two-horizon performance was evaluated using MAE, PICP, and PIAW, representing point forecast accuracy, empirical interval coverage, and interval width, respectively, over the 2015 and 2020 horizons. Figure 5.4 summarizes the resulting replicate-specific performance for both BUA and Pop, together with the corresponding ensemble result.

Overall, Figure 5.4 indicates that replicate-to-replicate variability was limited, suggesting that the selected GRU architecture was inherently stable across repeated training runs. For both BUA and Pop, the replicate-specific values clustered within relatively narrow ranges, indicating that the model's forecasting behavior was not strongly dependent on a particular random seed or optimization path. Moreover, the ensemble consistently yielded performance equal to or better than that of a typical single replicate. For BUA, ensemble aggregation reduced the median MAE by approximately 6%, narrowed the median PIAW by about 1%, and increased PICP by roughly 2% relative to the average performance of the 30 individual replicates. For Pop, the same pattern was observed, although the magnitude of improvement was smaller: median MAE decreased by around 2%, median PIAW narrowed by approximately 1%, and PICP increased by less than 1%.

These results show that the value of deep ensembling in this study lay less in correcting severe instability and more in refining an already stable forecasting architecture. The individual GRU replicates were broadly consistent, but their aggregation further suppressed residual stochastic variation and produced forecasts that were, on average, more accurate, slightly sharper, and better calibrated than those from any single model instance. These characteristics provide an additional justification for carrying the ensemble-based GRU forecasts forward into subsequent probabilistic SDG Indicator 11.3.1 projection and uncertainty-propagation analyses.

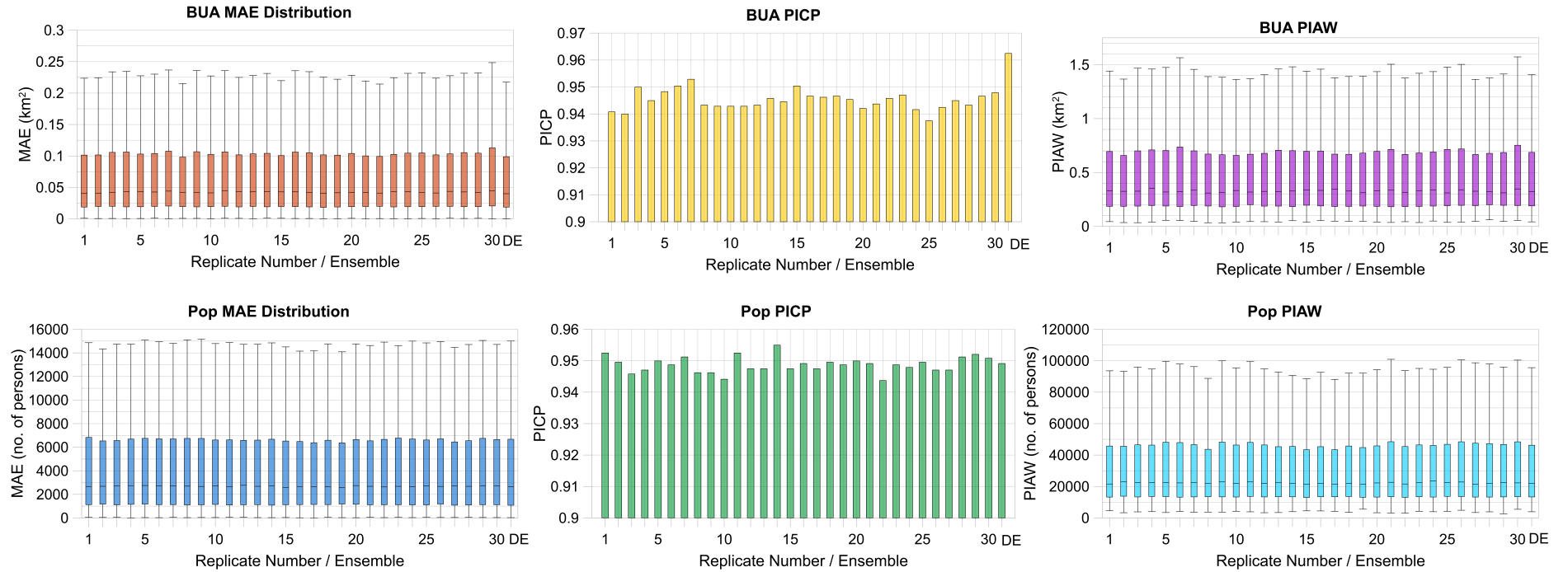


Figure 5.4: Two-horizon performance metrics (MAE, PIAW, PICP) for 30 independently trained replicates of the GRU model and the deep ensemble (DE), evaluated over the 1,199 urban centers (UCs) in the test set. Boxplots for MAE and PIAW represent the distribution of per-UC values after excluding outliers, about 120–140 UCs, using the $1.5 \times \text{IQR}$ rule to enhance visual clarity; all statistical calculations use the full test set. PICP bars report empirical coverage across all UCs for 2015 and 2020.

5.5.5 Forecast Uncertainty Dependence Modeling

The dependence structure used for joint uncertainty propagation was estimated from out-of-sample forecast residuals from the validation set. Among the candidate models, the specification with the lowest AIC was a truncated vine copula with truncation at tree level 2, and all retained pair-copula links belonged to the Student- t family. The fitted dependence structure showed that joint uncertainty was characterized mainly by strong temporal dependence within each variable across the forecast horizons, with model-implied Kendall's tau values of 0.69 for BUA_{2015} – BUA_{2020} and 0.72 for Pop_{2015} – Pop_{2020} . In contrast, residual dependence between BUA and Pop was weak. The same-year link between BUA_{2015} and Pop_{2015} had a Kendall's tau of only 0.04. The second-tree links were also close to zero: BUA_{2020} – Pop_{2015} conditional on BUA_{2015} had a Kendall's tau of -0.03 , while BUA_{2015} – Pop_{2020} conditional on Pop_{2015} had a Kendall's tau of -0.06 . Dependence beyond tree level 2 was truncated to independence; hence, no third-tree pair-copula was retained, and the corresponding higher-order conditional dependence was treated as zero.

This result indicates that the copula mainly captured temporal co-movement of forecast errors within BUA and within Pop, while adding only limited cross-variable dependence. In the propagation step, dependent random samples drawn from this fitted copula were mapped through the inverse empirical predictive CDFs of the forecast variables to obtain joint Monte Carlo samples on the original scales of BUA and Pop.

5.5.6 Effects of Forecast Uncertainty on SDG Indicator 11.3.1 Estimates

Figure 5.5 summarizes the distribution of 95% prediction-interval widths (PIWs) across the 1,199 test-set UCs for forecasted BUA and Pop at both forecast horizons, as well as for the projected SDG Indicator 11.3.1 components derived from them. At the input level, forecast uncertainty in BUA was generally modest and right-skewed, with median PIWs of 0.12 km^2 in 2015 and 0.53 km^2 in 2020 (Figure 5.5a–b). This result indicated a clear widening of uncertainty at the longer forecast horizon, while remaining relatively limited for most UCs. In contrast, forecast uncertainty in Pop was more heterogeneous across UCs. Median PIWs increased from about 11,000 persons in 2015 to about 31,000 persons in 2020; the distributions showed long right tails extending to very large values for a subset of UCs, indicating substantial variation in population forecast uncertainty across the test set, especially at the later horizon (Figure 5.5e–f).

When these forecast uncertainties were propagated through the SDG Indicator 11.3.1 formulae, the resulting uncertainty profiles differed markedly across the three indicators. LCR exhibited the narrowest and most homogeneous projected uncertainty across UCs, with a median PIW of 2.08% and a relatively compact distribution (Figure 5.5c). Its corresponding PICP, calculated as the proportion of UCs for which the observed LCR value fell within the propagated 95% interval, was 0.91. This PICP was below the nominal 0.95 level, indicating slight under-coverage, but still suggested comparatively stable projected intervals. PGR showed broader and more heterogeneous projected uncertainty, with a median PIW of 3.53%, consistent with the more uneven uncertainty already observed in the population forecasts (Figure 5.5f). Its PICP, again evaluated against the observed PGR values, was 0.92, indicating mild under-coverage.

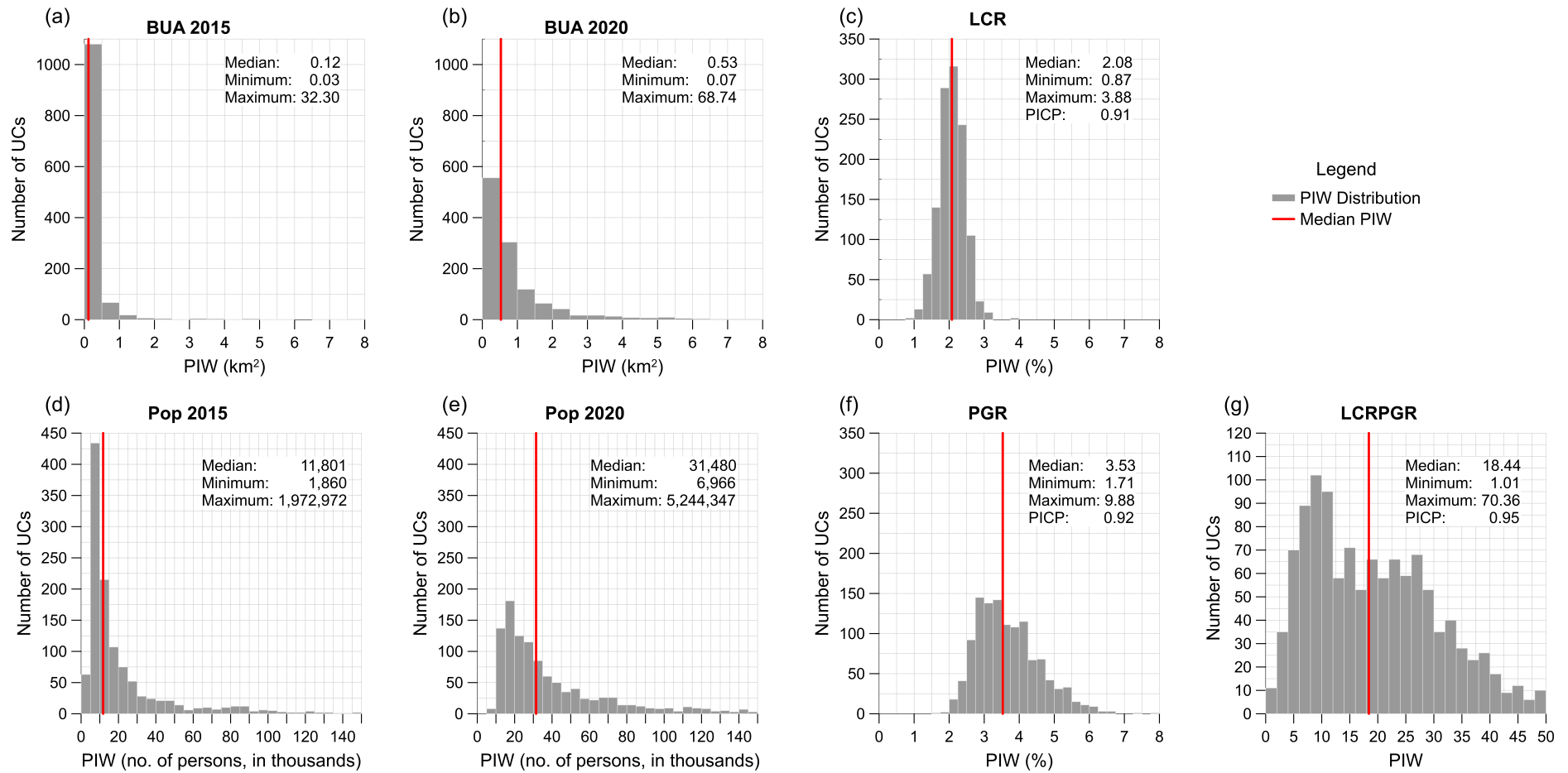


Figure 5.5: Distribution of 95% prediction interval widths (PIW) across the 1,199 test-set urban centers for forecasted built-up area (BUA) and population (Pop) in 2015 and 2020, and for the corresponding uncertainty-aware SDG Indicator 11.3.1 components (LCR, PGR, and LCRPGR), together with their prediction interval coverage probabilities (PICPs) over the 2015–2020 period. For BUA and Pop, PIWs are derived directly from the forecasting model’s 95% prediction intervals, representing input uncertainty at each time step. For LCR, PGR, and LCRPGR, PIWs and PICPs are calculated from the full Monte Carlo predictive distributions obtained through joint uncertainty propagation using the truncated level-2 vine copula. For visualization clarity, the histograms are truncated to the ranges shown.

The strongest uncertainty amplification was observed for LCRPGR (Figure 5.5g). Its median PIW reached 18.44, and the distribution was strongly right-skewed, with very large widths for a non-negligible subset of UCs. The histogram in Figure 5.5g was truncated for visualization clarity. Despite this substantial spread, the corresponding PICP, calculated against the observed LCRPGR values, was 0.95. This PICP suggests that, although LCRPGR exhibited the largest propagated uncertainty, the corresponding intervals provided the strongest empirical coverage among the three propagated outputs. Overall, the combined PIW and PICP results showed a clear pattern in forecast-uncertainty propagation across the test set: projected uncertainty was narrowest and most stable for LCR, broader for PGR, and most strongly amplified for LCRPGR.

To further examine how forecast uncertainty was transmitted from the inputs to the derived indicators, Spearman rank correlations (ρ) were calculated using UC-specific PIWs (Table 5.5). Across the test set, uncertainty width appeared highly persistent within each forecast target across horizons, with correlations of 0.98 for both BUA and Pop between 2015 and 2020. BUA and Pop interval widths were also positively associated across UCs ($\rho = 0.71$ – 0.77), suggesting that some UCs tended to show comparatively wider intervals for both forecast targets. Among the propagated indicators, PGR interval width was more closely associated with population forecast uncertainty ($\rho = 0.57$ – 0.62) than with BUA forecast uncertainty. By contrast, LCR interval width showed only a weak association with the input forecast interval widths. LCRPGR interval width was close to uncorrelated with both BUA and Pop forecast interval widths ($\rho = -0.13$ to -0.02). These results suggest that the absolute sizes of the forecast intervals in the inputs did not fully account for the large uncertainty observed in LCRPGR.

This interpretation is supported by Figure 5.6, which shows that the width of the propagated LCRPGR interval was related to the absolute projected median PGR. The figure shows that the propagated LCRPGR interval widened sharply as the absolute value of the projected PGR approached zero. This pattern could be attributed to the denominator sensitivity of the ratio formulation and suggests that instability in forecast-based LCRPGR was influenced not only by uncertainty in BUA and Pop, but also by the indicator’s structural sensitivity when projected population growth was weak.

Table 5.5: Spearman rank-correlation matrix of UC-specific 95% prediction-interval widths for forecasted BUA and Pop at the 2015 and 2020 horizons, and for the propagated SDG Indicator 11.3.1 components across the 1,199 test-set UCs.

	BUA 2015	BUA 2020	Pop 2015	Pop 2020	LCR	PGR	LCRPGR
BUA ₂₀₁₅	1.00						
BUA ₂₀₂₀	0.98	1.00					
Pop ₂₀₁₅	0.75	0.77	1.00				
Pop ₂₀₂₀	0.71	0.73	0.98	1.00			
LCR	−0.28	−0.32	0.22	0.26	1.00		
PGR	0.06	0.03	0.57	0.62	0.78	1.00	
LCRPGR	−0.12	−0.13	−0.03	−0.02	0.28	0.07	1.00

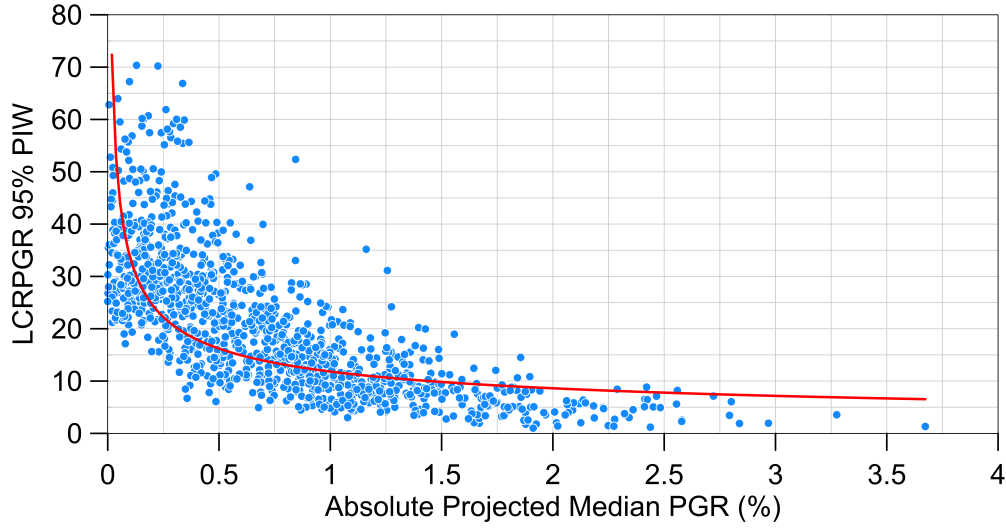


Figure 5.6: Relationship between the absolute projected median PGR and the 95% prediction-interval width (PIW) of projected LCRPGR across the 1,199 test-set urban centers (UCs). Each point represents one UC, and the red curve shows a fitted power function summarizing the overall trend. The Spearman rank-correlation coefficient (ρ) is -0.81 .

5.5.7 Predictive Agreement of Projected LCR, PGR, and LCRPGR with Observed Values

Table 5.6 summarizes the agreement between projected median indicator values and the corresponding observed values across the test-set UCs. The median error was close to zero across all three indicators, indicating little systematic bias. LCR had the smallest median absolute error, showing the closest agreement in terms of typical error size. PGR had the highest Spearman’s ρ , indicating the strongest rank-based agreement; that is, UCs with higher observed PGR also tended to have higher projected PGR, and vice versa. LCRPGR showed the weakest rank-based agreement and a median absolute error similar to that of PGR. These results suggest that the BUA–Pop forecasting model produced broadly reliable projected inputs for downstream indicator estimation, although agreement weakened for LCRPGR due to its more pronounced sensitivity to the ratio-based formulation.

Table 5.6: Agreement statistics between projected median indicator values and observed values across the 1,199 test-set UCs. Median Error and Median Absolute Error are expressed in the same units as the corresponding indicator component. Accordingly, LCR and PGR errors are reported in percentage points, whereas LCRPGR errors are unitless.

Indicator Component	Median Error	Median Absolute Error	Spearman’s ρ
LCR	-0.01	0.28	0.53
PGR	-0.03	0.48	0.66
LCRPGR	0.04	0.49	0.40

5.5.8 Characteristics and Accuracy of Projected Probabilistic LUE Classification

Figure 5.7 and Table 5.7 provide complementary views of the projected probabilistic LUE regime classification performance across the 1,199 test-set UCs. Figure 5.7 summarizes the aggregate class distribution, showing that, relative to the observed classification, the projected probabilistic results assigned more UCs to *Efficient Built-up Expansion under Faster Population Growth* (R2) and *Inefficient Built-up Expansion under Population Decline* (R12–R14), and fewer to *Inefficient Built-up Expansion under Slower Population Growth* (R16). This indicates that, once forecast uncertainty in BUA and Pop was propagated through the SDG Indicator 11.3.1 transformations, the resulting probabilistic projection did not reproduce the observed class frequencies exactly, but instead tended to reallocate some UCs away from the sprawl-driven inefficient class toward either efficient-growth or decline-related regimes.

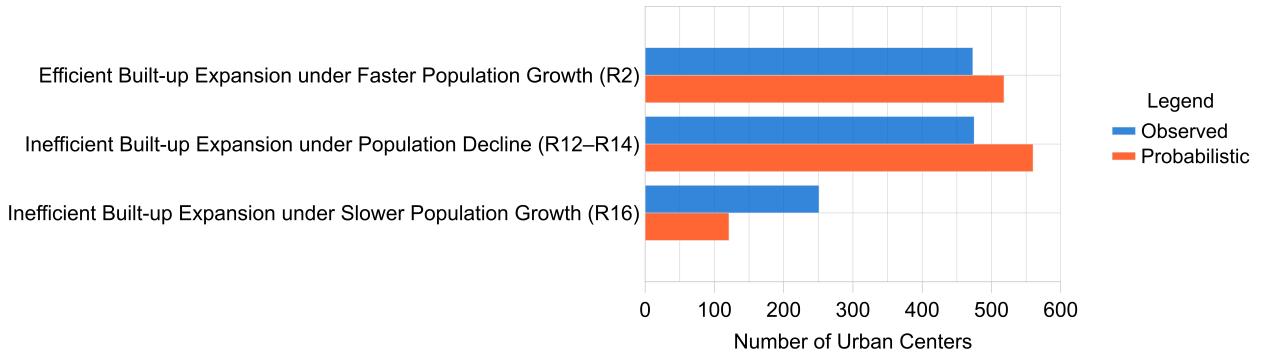


Figure 5.7: Distribution of LUE regimes across the test set (1,199 urban centers) based on observed indicators and probabilistic, uncertainty-aware classification derived from Monte Carlo simulations.

Table 5.7 examines this pattern at the UC level through a confusion-matrix assessment. The projected probabilistic classification achieved an overall accuracy of 63.6%, indicating moderate agreement with the observed classes. Agreement was strongest for *Efficient Built-up Expansion under Faster Population Growth* and *Inefficient Built-up Expansion under Population Decline*, with producer’s accuracies of 73.4% and 77.7%, respectively. In contrast, *Inefficient Built-up Expansion under Slower Population Growth* showed a substantially lower producer’s accuracy of 18.7%. User’s accuracies showed the same pattern, with higher reliability for R2 and the combined R12–R14 group than for R16. These results suggest that forecast uncertainty in BUA and Pop did not affect all regimes equally. Instead, the projected probabilistic classification was more effective at correctly identifying the two dominant LUE regimes than at distinguishing inefficient expansion under slower population growth, which was frequently reassigned to neighboring regimes.

Table 5.8 further shows that correctly projected probabilistic classifications tended to have higher dominant regime probabilities than incorrectly projected ones. The median dominant probability was 0.69 for correct classifications, compared with 0.50 for incorrect classifications, with the interquartile range likewise shifted upward for the correct group. This result is consistent with the earlier findings from Figure 5.7 and Table 5.7. UCs that were correctly classified generally retained a more clearly dominant projected regime, even after forecast uncertainty in BUA and Pop was propagated into LCR, PGR, and LCRPGR. In contrast, misclassified UCs tended to show weaker dominance and larger ambiguity among competing classes.

Table 5.7: Confusion matrix showing agreement between observed LUE classes and probabilistic LUE classifications derived from Monte Carlo-simulated distributions of LCR, PGR, and LCRPGR across 1,199 UCs. Rows represent predicted dominant probabilistic classes and columns represent observed classes.

Predicted / Observed	Efficient Built-up Expansion Under Faster Population Growth (R2)	Inefficient Built-up Expansion under Population Decline (R12–R14)	Inefficient Built-up Expansion under Slower Population Growth (R16)	Total	User's Accuracy (%)
Efficient Built-up Expansion under Faster Population Growth (R2)	347	69	102	518	67.0
Inefficient Built-up Expansion under Population Decline (R12–R14)	89	369	102	560	65.9
Inefficient Built-up Expansion under Slower Population Growth (R16)	37	37	47	121	38.8
Total	473	475	251	1,199	
Producer's Accuracy (%)	73.4	77.7	18.7		
Overall Accuracy (%)	63.6				

Table 5.8: Summary statistics of dominant LUE regime probabilities for projected probabilistic LUE classifications across the 1,199 test-set UCs, grouped by whether the dominant predicted class matches the observed class.

Type	Median	IQR	Min–Max
Correct Classifications	0.69	0.52–0.80	0.35–1.00
Incorrect Classifications	0.50	0.42–0.62	0.33–0.94

5.5.9 SDG Indicator 11.3.1 Projections for 2025–2030

Figure 5.8 summarizes the distributions of observed and forecasted BUA, Pop, and the derived SDG Indicator 11.3.1 components across the 8,478 UCs, together with forecast uncertainty expressed as 95% PIW. The 2020 values are observed, whereas the 2025 and 2030 values are forecast medians. These summaries are complemented by the global maps in Figure 5.9, which show the spatial patterns of LCR, PGR, and LCRPGR for 2020–2025 and 2025–2030 and reveal marked regional differences in growth rates and efficiency outcomes. These figures show how projected central tendencies, spatial structure, and uncertainty evolve, while Tables 5.9–5.10 provide the corresponding aggregated global projections from the Monte Carlo simulations.

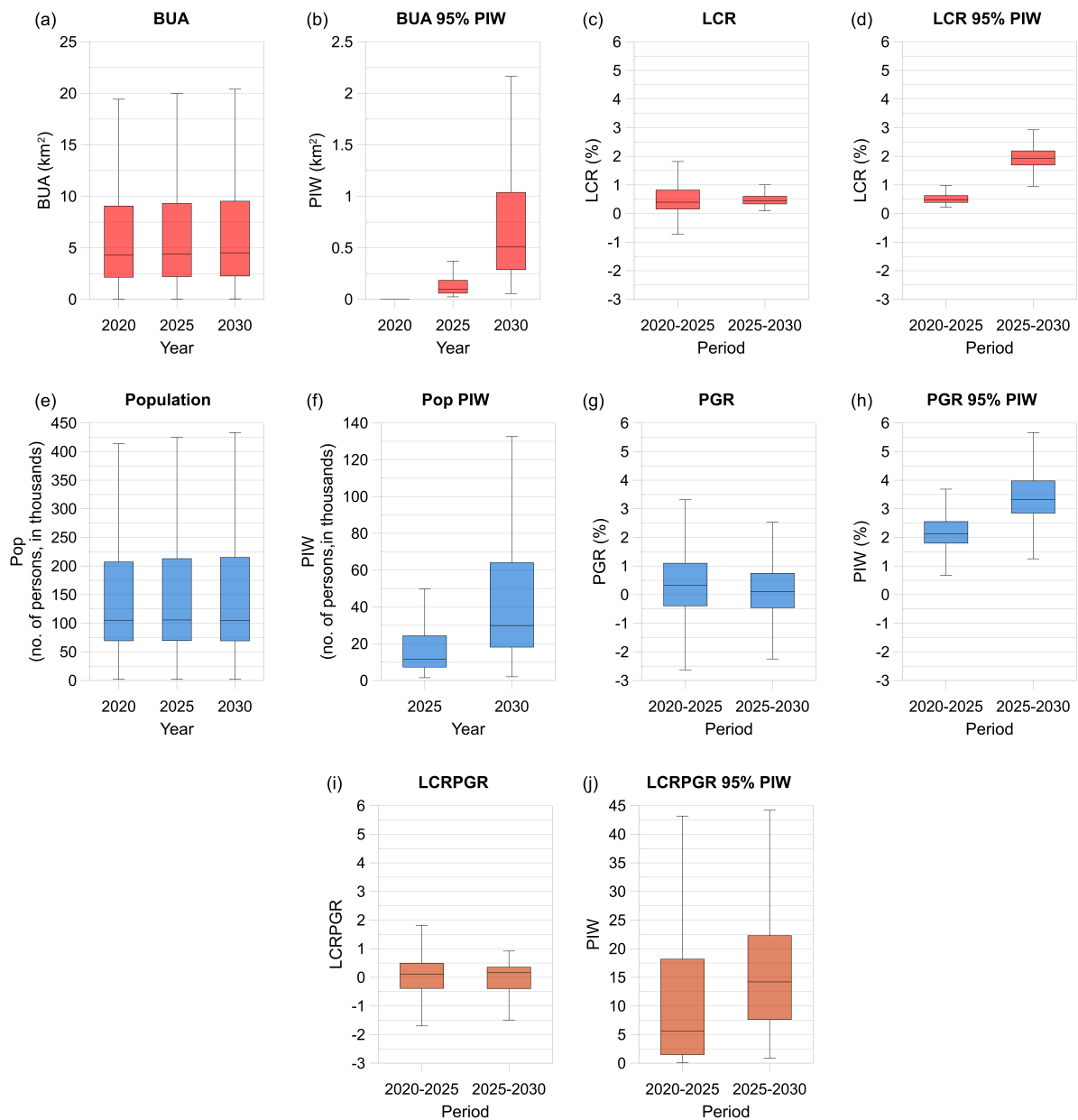


Figure 5.8: Distributions of observed (2020) and forecasted (2025–2030) BUA, Pop, and the derived SDG Indicator 11.3.1 components across the 8,478 urban centers in the study.

Table 5.9: Aggregated BUA and Pop projections and derived SDG Indicator 11.3.1 metrics for 8,478 urban centers (2020–2030). Values for 2020 are observed. Values for 2025 and 2030 are median estimates from 250,000 Monte Carlo simulations of global totals, with 95% prediction intervals [2.5th–97.5th percentiles]. Percentage change columns indicate relative shifts across forecast horizons.

Metric	2020 (Observed)	2025 (Forecast)	2030 (Forecast)	% Change (2020– 2025)	% Change (2025– 2030)	% Change (2020– 2030)
Total BUA ($\times 1000$ km^2)	108.81	111.47 [110.21–113.28]	114.00 [111.34–123.39]	+2.44	+2.27	+4.77
Total Pop (Billions)	2.88	3.03 [2.85–3.21]	3.17 [2.71–3.67]	+5.28	+4.61	+10.13

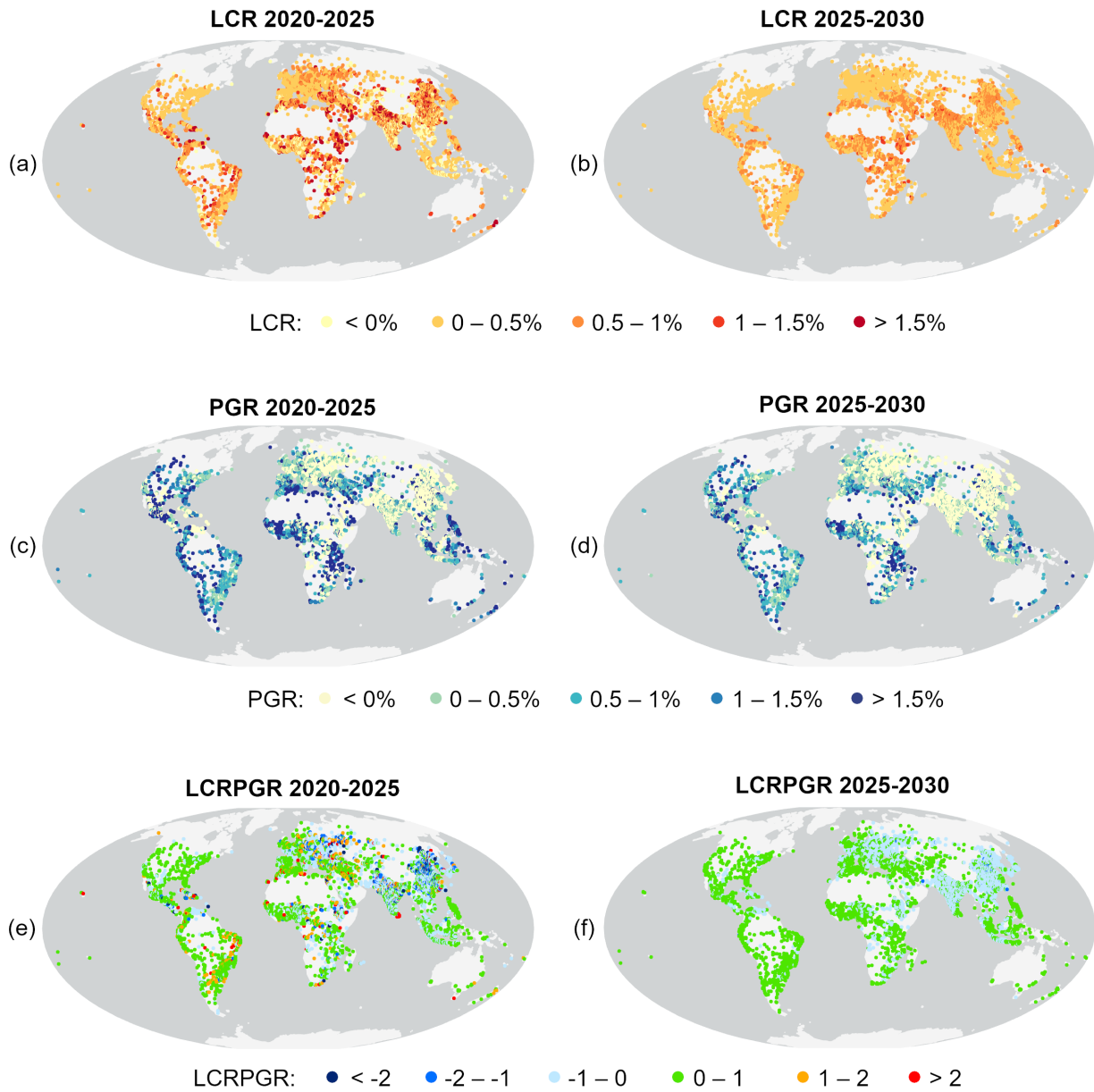


Figure 5.9: Global distribution of projected SDG Indicator 11.3.1 components across 8,478 urban centers for two forecast horizons. Maps are based on median estimates from uncertainty-aware forecasts.

Table 5.10: Aggregated SDG Indicator 11.3.1 projections for 8,478 urban centers (2020–2030). Results are reported with 95% prediction intervals [2.5th–97.5th percentiles].

Period	LCR (% yr ⁻¹)	PGR (% yr ⁻¹)	LCRPGR
2020–2025	0.49 [0.26–0.82]	1.03 [−0.19–2.18]	0.46 [−3.06–2.12]
2025–2030	0.45 [0.16–1.79]	0.91 [−0.99–2.72]	0.41 [−4.15–4.80]

For BUA (Figure 5.8a), the distributions were forecasted to shift upward from 2020 to 2025 and 2030, indicating continued urban expansion across UCs. This result is consistent with Table 5.9, which shows steady global BUA growth, with a cumulative increase of 4.77% from 2020 to 2030. The BUA 95% PIW (Figure 5.8b) remained tightly clustered at low values in 2025, mostly below 0.5 km², but became more dispersed by 2030, consistent with the wider aggregated prediction intervals reported in Table 5.9.

For LCR (Figure 5.8c), both interval distributions were projected to remain centered on positive values, indicating continued land consumption across most UCs. However, the 2025–2030 distribution shifted toward lower values, suggesting moderation in land consumption rates. This pattern is also reflected in Table 5.10, where median global LCR declines slightly from 0.49% in 2020–2025 to 0.45% in 2025–2030. At the same time, LCR uncertainty increased, with wider PIWs in 2025–2030 (Figure 5.8d), consistent with Table 5.10.

Forecasted population distributions (Figure 5.8e) likewise shifted upward from 2020 to 2030, reflecting continued growth across UCs. Table 5.9 shows the global population increasing from about 2.88 billion in 2020 to about 3.03 billion in 2025 and 3.17 billion in 2030, corresponding to cumulative growth of just over 10%. This growth was accompanied by higher uncertainty, as population PIWs (Figure 5.8f) widened markedly from 2025 to 2030.

The PGR distributions (Figure 5.8g) were projected to remain centered near zero in both periods but shifted downward in 2025–2030, with more UCs showing near-zero or negative growth. This suggests slowing demographic momentum, as also reflected in Table 5.10, where median global PGR declines from 1.03% in 2020–2025 to 0.91% in 2025–2030. As with LCR, this moderation was accompanied by wider uncertainty intervals in the later period (Figure 5.8h).

For LCRPGR, both distributions remained centered near zero but shifted toward lower values in 2025–2030, with more UCs concentrated at negative or near-zero ratios (Figure 5.8i). This result reflects the combined effect of moderating land consumption and slowing population growth. Table 5.10 shows a corresponding decline in median global LCRPGR from 0.46 in 2020–2025 to 0.41 in 2025–2030. At the same time, LCRPGR uncertainty increased substantially: the 95% PIW was already wide in 2020–2025 and became even more dispersed in 2025–2030 (Figure 5.8j). Thus, even where LCRPGR was projected to decline, the range of plausible efficiency outcomes widened, calling for cautious interpretation of longer-term projections for SDG Indicator 11.3.1.

5.5.10 Probabilistic LUE Classification and Confidence Patterns

Figures 5.10–5.11 present the global spatial distribution of probabilistic LUE classes and their associated classification confidence for the 2020–2025 and 2025–2030 periods, respectively, while Tables 5.11–5.12 summarize the corresponding distributions across the 8,478 global UCs.

For the 2020–2025 period, where BUA and Pop values for 2020 were observed and uncertainty arose only from the 2025 forecasts, LUE outcomes were dominated by *Efficient Built-up Expansion under Faster Population Growth* (R2) and *Inefficient Built-up Expansion under Population Decline* (R12–R14), accounting for about 39% and 36% of UCs, respectively (Table 5.11). *Inefficient Built-up Expansion under Slower Population Growth* (R16) accounted for about 14%. The projected

spatial patterns in Figure 5.10a showed broad clustering of R2 and R12–R14 across different world regions, with R12–R14 appearing prominently across several areas and R2 also widely distributed..

The remaining projected regimes accounted for much smaller shares. *Efficient Built-up Reduction under Population Growth* (R4–R6), in which the forecasted built-up area in 2025 was lower than the observed built-up area in 2020 while population increased, accounted for about 7% of UCs. *Efficient Built-up Reduction under Population Decline* (R8), in which both LCR and PGR were negative but the decline in built-up area was stronger, was extremely rare, with only 7 UCs, or less than 1% of the total. *Inefficient Built-up Reduction under Population Decline* (R10), in which both rates were negative but population decline was stronger, accounted for about 3% of UCs.

The forecast classification confidence for 2020–2025 was relatively concentrated in the upper categories, from moderate to very high. About 55% of UCs fell in the high or very high confidence classes, while about 25% fell in the low or very low confidence classes (Figure 5.10b; Table 5.11). R2 and the combined R12–R14 contained the largest numbers of high- and very-high-confidence classifications. For R2, about 65% of its UCs fell in the high or very high confidence categories, while for the combined R12–R14, the corresponding share was about 62%. By contrast, R16 was more concentrated in the lower confidence categories, with about 48% in very low or low confidence and only about 24% in high or very high confidence. The combined R4–R6 was centered more in the middle confidence range, with most UCs in the low to high categories and relatively few in the very high class. R10 was also concentrated in the lower to middle confidence levels. R8 was limited to the very low and low confidence classes.

For the 2025–2030 period, in which both temporal endpoints were derived from probabilistic forecasts, the projected class distribution became more concentrated into two regimes (Figure 5.11a; Table 5.12). *Inefficient Built-up Expansion under Population Decline* (R12–R14) accounted for about 55% of UCs, while *Efficient Built-up Expansion under Faster Population Growth* (R2) accounted for about 42%. *Inefficient Built-up Expansion under Slower Population Growth* (R16) declined sharply to about 3%. In contrast to 2020–2025, the combined R4–R6, R8, and R10 no longer appeared as dominant projected classes in Table 5.12. Spatially, Figure 5.11a showed an intensification of the combined R12–R14 patterns across large parts of Europe, East Asia, and other regions previously shown to be affected by projected population decline (Figure 5.9c–d).

The forecast confidence distribution also shifted downward in 2025–2030. About 29% of UCs fell in the high or very high confidence categories, compared with about 55% in 2020–2025, while about 45% fell in the very low or low confidence classes, compared with about 25% in the earlier period (Figure 5.11b; Table 5.12). Moderate-confidence classifications also increased slightly. For R2, the projected classifications were distributed mainly across the very low, low, moderate, and high confidence categories, with very few UCs in the very high category. The combined R12–R14 retained the largest number of high-confidence classifications, but its distribution was also more spread toward lower confidence levels than in 2020–2025. For R16, all projected classifications fell in the very low confidence category in 2025–2030. These results show that the later forecast period was characterized by a narrower set of dominant projected LUE regimes and lower forecast classification confidence overall.

Comparing the two periods shows two clear patterns. While the spatial structure and dominance of LUE classes evolved gradually, the reliability of regime assignment deteriorated sharply as uncertainty propagated from one uncertain endpoint in 2020–2025 to two uncertain endpoints in 2025–2030. Figure 5.11b shows that areas with dominant regimes could nevertheless experience substantial reductions in classification confidence.

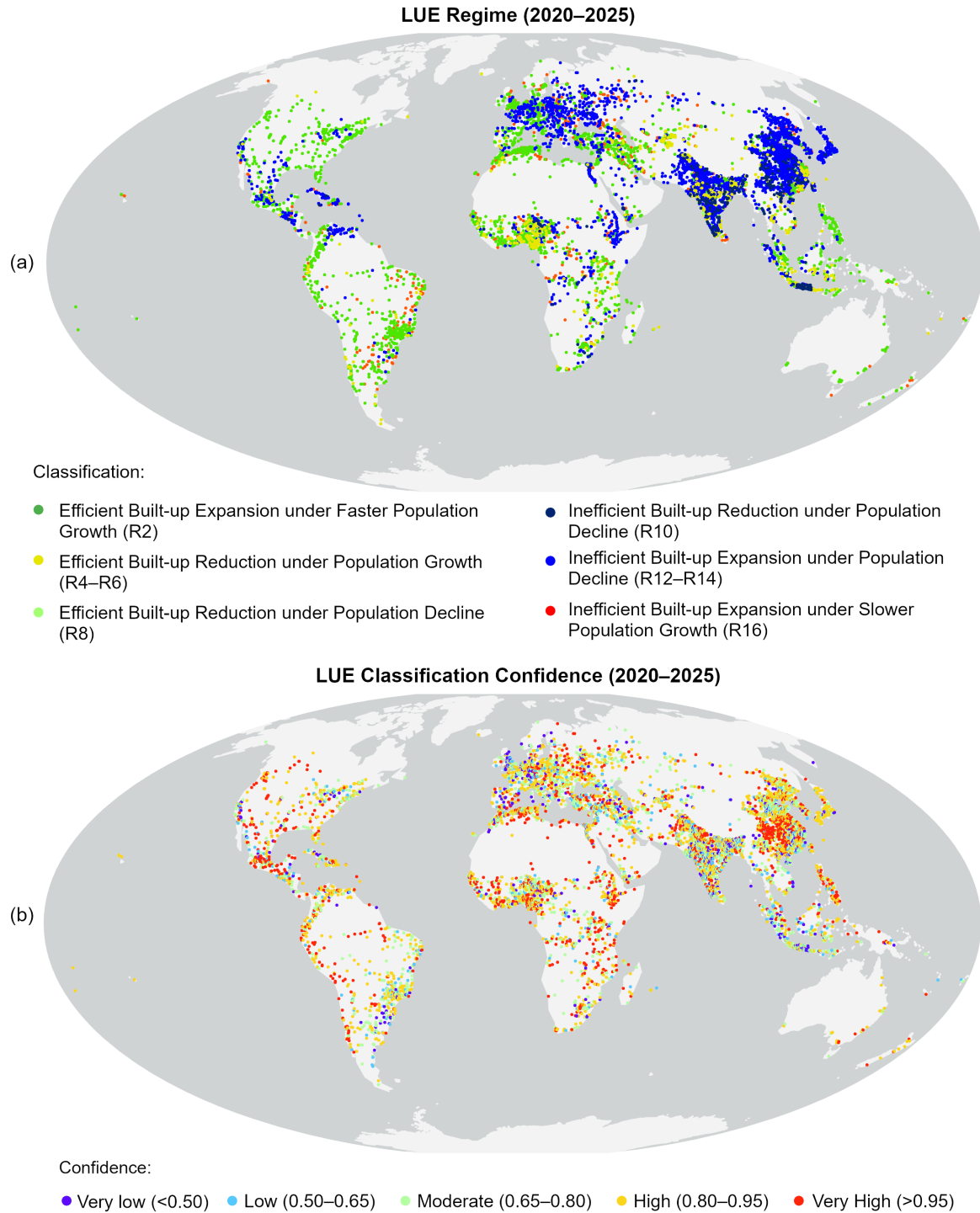


Figure 5.10: Probabilistic LUE regime classification and classification confidence for the 2020–2025 period across 8,478 global urban centers.

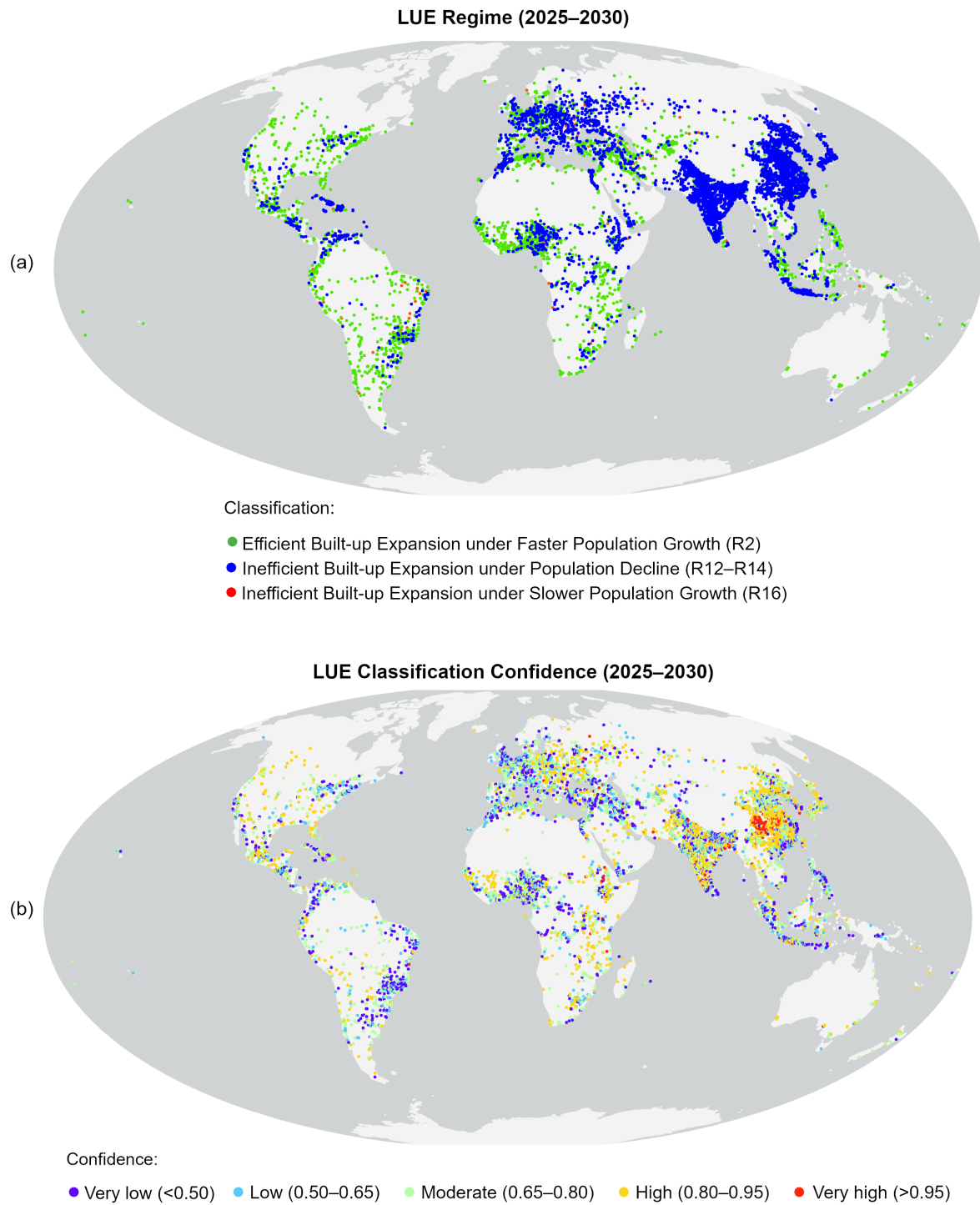


Figure 5.11: Probabilistic LUE regime classification and classification confidence for the 2025–2030 period across 8,478 global urban centers.

Table 5.11: Distribution of urban centers by LUE regime and classification confidence (2020–2025).

LUE Regime	Classification Confidence					Total UCs
	Very low (< 0.50)	Low ($0.50\text{--}0.65$)	Moderate ($0.65\text{--}0.80$)	High ($0.80\text{--}0.95$)	Very High (> 0.95)	
Efficient Built-up Expansion under Faster Population Growth (R2)	203 (2.4%)	427 (5.0%)	536 (6.3%)	1,216 (14.3%)	953 (11.2%)	3,335 (39.3%)
Efficient Built-up Reduction under Population Growth (R4–R6)	74 (0.9%)	123 (1.5%)	173 (2.0%)	172 (2.0%)	50 (0.6%)	592 (7.0%)
Efficient Built-up Reduction under Population Decline (R8)	5 (0.1%)	2 (0.0%)	–	–	–	7 (0.1%)
Inefficient Built-up Reduction under Population Decline (R10)	71 (0.8%)	75 (0.9%)	82 (1.0%)	54 (0.6%)	10 (0.1%)	292 (3.4%)
Inefficient Built-up Expansion under Population Decline (R12–R14)	204 (2.4%)	382 (4.5%)	598 (7.1%)	1,277 (15.1%)	633 (7.5%)	3,094 (36.5%)
Inefficient Built-up Expansion under Slower Population Growth (R16)	184 (2.2%)	375 (4.4%)	322 (3.8%)	255 (3.0%)	22 (0.3%)	1,158 (13.7%)
Total	741 (8.7%)	1,384 (16.3%)	1,711 (20.2%)	2,974 (35.1%)	1,668 (19.7%)	8,478 (100%)

5.6 Discussion

5.6.1 Implications of Formulating Joint BUA–Pop Forecasting as a Global Pooled Learning Task

The results indicate that formulating BUA and Pop forecasting as a pooled, joint-target learning problem across thousands of UCs was effective for capturing broad temporal regularities while still accommodating locality-level heterogeneity. Across 8,478 UCs, the workflow was designed explicitly to learn from a shared collection of historical BUA–Pop trajectories rather than from isolated city-specific models. The test-set results show that this global formulation generated accurate out-of-sample forecasts for both variables across diverse regional and income-group contexts. The representative UC examples further support this interpretation, showing that the global DL models

Table 5.12: Distribution of urban centers by LUE regime and classification confidence (2025–2030).

LUE Regime	Classification Confidence					Total UCs
	Very low (< 0.50)	Low ($0.50\text{--}0.65$)	Moderate ($0.65\text{--}0.80$)	High ($0.80\text{--}0.95$)	Very High (> 0.95)	
Efficient Built-up Expansion under Faster Population Growth (R2)	858 (10.1%)	883 (10.4%)	1,025 (12.1%)	794 (9.4%)	4 (0.1%)	3,564 (42.0%)
Inefficient Built-up Expansion under Population Decline (R12–R14)	903 (10.7%)	942 (11.1%)	1,189 (14.0%)	1,483 (17.5%)	175 (2.1%)	4,692 (55.3%)
Inefficient Built-up Expansion under Slower Population Growth (R16)	222 (2.6%)	–	–	–	–	222 (2.6%)
Total	1,983 (23.4%)	1,825 (21.5%)	2,214 (26.1%)	2,277 (26.9%)	179 (2.1%)	8,478 (100%)

generally preserved the direction and continuation of observed post-2010 trajectories more consistently than the local ES benchmark, especially for BUA. The results also show that Pop is more difficult to forecast than BUA, indicating that pooled learning improves generalization but does not eliminate the intrinsic complexity of demographic dynamics across UCs.

This finding is important for the broader context of this dissertation because the forward-looking SDG Indicator 11.3.1 assessment requires coherent forecasts of both land and population inputs, not just single-variable extrapolations. The present results suggest that shared representation learning across UCs can provide that coherence in a scalable way. In other words, the global formulation appears to offer a practical compromise between purely local models, which may be too weak to learn robustly from very short series, and highly city-specific modeling strategies, which are difficult to scale consistently across the world. This finding, however, requires some qualification. Later residual-dependence analysis (Section 5.5.5) showed only weak cross-variable dependence between BUA and Pop forecast residuals. This does not mean that the forecasting task was not genuinely joint. Rather, it means that after forecasting, the remaining dependence between the two variables was weak in the residual or uncertainty structure examined here. In other words, the model may still have benefited from learning BUA and Pop together even if the final forecast errors did not show strong coupling. The main advantage of the joint formulation, therefore, lies less in capturing strong cross-variable dependence structure and more in enabling simultaneous, scalable forecasting of both SDG Indicator 11.3.1 inputs within a shared global framework. Additionally, it should be made clear that this forecasting exercise is conditional on the observed historical input segment being treated as fixed, so the gains demonstrated here should be interpreted as improvements in forecasting future states given the available history, rather than as a full representation of all uncertainty affecting future LUE assessment.

5.6.2 Comparative Behavior of the Candidate DL Architectures

The architecture comparison shows that the main performance differences among the DL models were modest, with the best candidates separated by only small margins. The Friedman test indicated overall differences among architectures. However, the top mean ranks of GRU and 1D CNN+GRU were nearly identical, and the corresponding pairwise effect sizes were small. This pattern suggests that, under the short-sequence setting used here, the key benefit came less from architectural sophistication per se than from the model’s ability to learn stable nonlinear temporal summaries from the eight-step historical window. GRU emerged as the most balanced final choice because it combined the best overall joint rank with the strongest BUA point performance, competitive Pop performance, and acceptable interval behavior, while avoiding the added structural complexity of the hybrid alternative. The strong performance of GRUs further suggests that recurrent architectures are especially well-suited for handling the nonlinear, intermittent, and often asymmetric growth patterns characteristic of urban expansion. This observation is consistent with findings from sequence modeling research showing that gated recurrent mechanisms perform well when temporal signals are sparse, noisy, or irregular, and when forecasting horizons are short (e.g., Abotaleb & Dutta, 2024; Barua et al., 2024; Che et al., 2018; Li & Xu, 2019; Mateus et al., 2021). The finding that such architectures generalize effectively across thousands of highly diverse urban contexts adds weight to the argument that global, data-driven models can extract transferable temporal representations even when local conditions differ markedly.

The performance comparison also reveals that more architectural complexity did not automatically translate into more reliable forecasting for this application. The Transformer, for example, achieved a very favorable prediction interval, as reflected in the Winkler score, but this came at the cost of sub-nominal coverage for Pop. Conversely, the recurrent and hybrid models tended to be somewhat less sharp but more conservative and better calibrated. The implication is that, for uncertainty-aware SDG forecasting, model selection cannot be based solely on point accuracy or interval sharpness. A model that yields narrow intervals but systematically undercovers is less suitable for downstream uncertainty propagation than a slightly less sharp model whose predictive distributions remain more credible. From that perspective, the final choice of GRU is defensible not because it dominated every metric, but because it offered the most robust trade-off across joint forecasting accuracy, calibration, and parsimony.

5.6.3 Role of Deep Ensembling in Stabilizing Predictive Uncertainty

The deep ensemble results show that ensembling contributed meaningful refinement even though the selected GRU architecture was already fairly stable at the single-model level. Replicate-to-replicate variability across the 30 independently trained GRU models was limited, indicating that the fitted forecasting behavior was not highly sensitive to random initialization or optimization path. Even so, aggregating across replicates still improved the final forecast, especially for BUA, as evidenced by reduced median MAE, narrower prediction interval widths, and increased empirical coverage. For Pop, the gains were smaller but still relevant. These results suggest that the main contribution of

deep ensembling in this study was not to compensate for an unstable model, but to suppress residual stochastic variation and produce a more reproducible predictive distribution for downstream use.

These findings are especially relevant in the present chapter because the forecasts are not the final analytical product; they are intermediate, uncertainty-bearing inputs to the propagation of SDG Indicator 11.3.1. In that setting, even modest improvements in stability matter because small changes in BUA and Pop uncertainty can be magnified once propagated into LCR, PGR, and especially LCRPGR. The ensemble therefore strengthens the methodological credibility of the workflow by ensuring that the propagated distributions reflect the systematic behavior of the selected architecture rather than the stochastic variation from a single training run. The later use of denser ensemble quantiles to reconstruct empirical predictive CDFs also makes the uncertainty representation sufficiently smooth and informative for Monte Carlo propagation.

5.6.4 Added Value of the Global DL Workflow Relative to the ES Baseline

One of the objectives of the chapter was to determine whether the proposed global DL workflow offers real value beyond a classical local benchmark, and the results clearly show that it does. As shown in Tables 5.2 and 5.3, all DL models substantially outperformed the ES baseline in terms of point-forecast accuracy and prediction-interval quality for both BUA and Pop. The gap was especially pronounced for BUA, where ES showed much larger MAE and MAPE values. This difference should be interpreted in light of the distinct role of ES in the experimental design. ES was included as a classical local, univariate baseline, as described in Section 5.4.5, to represent what can be achieved using only each UC’s own historical trajectory, without cross-UC learning, joint BUA–Pop modeling, or deep feature representation. It is therefore not expected to match the representational capacity of the global DL models; rather, it serves as a simple local trend-extrapolation benchmark against which the added value of cross-UC learning, joint BUA–Pop modeling, and nonlinear feature representation can be assessed.

The qualitative examples show that the ES forecasts frequently became too flat and often failed to preserve the continuation of observed upward trajectories. In contrast, the DL models more often maintained the observed temporal direction and magnitude of change. This difference is important because the forecasting problem here is not simply short-term extrapolation of an isolated series, but joint projection across many UCs with diverse growth paths. The stronger performance of the DL models in Tables 5.2 and 5.3 therefore suggests that the forecasting task benefits from pooled learning across many UCs, especially under short five-year interval time series where purely local smoothing methods have limited information from which to estimate future change. The results therefore support the claim that cross-UC learning carries substantial added value over purely local, univariate smoothing in this application.

Moreover, the baseline comparison is useful because it clarifies the source of the added value. The advantage of the DL workflow appears to lie in its ability to exploit pooled information and nonlinear structure across related series, rather than merely in using a more modern modeling approach. ES remained transparent and computationally simple, but its poorer performance indicates that

locality-specific univariate extrapolation is generally insufficient for the coupled BUA–Pop forecasting task required by forward-looking LUE assessment. Thus, the large difference between ES and the DL models does not undermine the suitability of ES as a baseline; rather, it confirms its function as a simple classical reference against which the added value of global joint-target DL forecasting can be evaluated. This finding reinforces the chapter’s broader methodological argument. If the goal is not just to forecast single trajectories, but to obtain coherent probabilistic inputs for derived SDG indicators across thousands of UCs, then a global joint-target framework is more appropriate than a collection of independent local models.

5.6.5 Implications for Forecast-Uncertainty Propagation and Probabilistic LUE Assessment

The uncertainty-propagation results show that forecast uncertainty is not transmitted uniformly across the components of SDG Indicator 11.3.1. LCR remained the most stable derived quantity, with comparatively narrow propagated intervals and only mild under-coverage. PGR showed broader, more heterogeneous uncertainty, reflecting the larger variability already present in population forecasts. LCRPGR, however, displayed the strongest amplification; its median 95% interval width was far larger than those of LCR and PGR, and its distribution was strongly right-skewed across the test-set UCs.

The results further show that this amplification cannot be explained solely by the size of the input intervals. The fitted dependence model indicated strong temporal dependence within BUA and within Pop forecast residuals, but only weak cross-variable dependence. Moreover, the correlation analysis showed that the LCRPGR interval width was nearly uncorrelated with the widths of the BUA and Pop forecast intervals. Instead, LCRPGR uncertainty widened sharply as the absolute projected median PGR approached zero. This finding closely aligns with the literature’s emphasis on denominator sensitivity as a central source of anomalous LCRPGR behavior. Studies in the deterministic setting have shown that small population fluctuations can induce large magnitude changes, sign reversals, and oscillatory behavior in the ratio (Angel et al., 2024; Zhong et al., 2025b). The forecasting results extend that insight by showing that even reasonably calibrated forecasts of BUA and Pop may still produce highly unstable LCRPGR projections when projected demographic growth is weak. Thus, the main source of instability is not merely forecast imprecision in the inputs, but the interaction between forecast uncertainty and the mathematical form of the indicator itself.

The classification results further clarify the decision relevance of uncertainty-aware forecasting. The projected probabilistic LUE classification achieved an overall accuracy of 63.6% against observed classes, with substantially stronger producer’s accuracy for R2 and combined R12–R14 than for R16. Moreover, correct classifications were associated with higher dominant-regime probabilities than incorrect ones, showing that the posterior dominance measure is informative. It follows that the workflow does not only output a future class label; it also communicates how strongly the propagated uncertainty supports that class. That is a practical advantage over deterministic forecasting, because it allows regime prevalence and regime reliability to be distinguished. At the

same time, the weak performance of the sprawl-driven inefficient class indicates that some regimes are inherently less separable under forecast uncertainty.

5.6.6 Interpretation of the 2020–2025 and 2025–2030 Projections

The 2025–2030 projections extend the discussion from evaluation to prospective insight. At the aggregate level, the forecasts suggest continued growth in both BUA and Pop across the global UC system up to 2030. These forecasting results can be interpreted in relation to earlier global and regional urban growth projections and SDG Indicator 11.3.1 monitoring efforts. In terms of magnitude, the projected growth rates of BUA and Pop are also consistent with broader global urban growth projections reported in the literature. Li et al. (2019) estimate that global urban land area will expand by approximately 40–67% between 2013 and 2050, corresponding to annual growth rates of about 1.1–1.8%.

In contrast, Gao & O'Neill (2020) project much larger increases—ranging from 80% to 490% by 2100, approximately 0.8–4.9% per year—depending on underlying socioeconomic trajectories and policy assumptions. For the 8,478 UCs analyzed in this study, which represent a substantial but incomplete subset of the 11,422 UCs in the GHS-UCDB 2025 and exclude smaller urban settlements and other urban areas, BUA is forecast to grow by approximately 0.48% annually between 2020 and 2030 (Table 5.9). Although this projected expansion rate is lower than most long-term global estimates, it remains consistent with the lower end of reported projections when differences in spatial coverage, urban definitions, and forecast horizons are taken into account. Population is projected to increase by about 10% cumulatively over the same decade, closely aligning with the United Nations' medium-variant projection of approximately 9% for the same period (UN Population Division, 2025). These comparisons suggest that the forecast trajectories produced by the workflow are not only internally consistent but also externally plausible within the context of large UCs.

Moreover, the indicator projections suggest moderation rather than acceleration in built-up land consumption. Median LCR declines slightly from 0.49 in 2020–2025 to 0.45 in 2025–2030; median PGR declines from 1.031 to 0.907; and median LCRPGR also declines from 0.46 to 0.403, indicating improved efficiency. Thus, the likely future signaled by the workflow is not one of urban contraction, but one in which continued horizontal and demographic expansion occurs under slowing rates, particularly on the population side. This projected pattern is consistent with retrospective findings showing that more than half of global functional urban areas improved their efficiency between 2000 and 2015 compared to 1990–2000 (Schiavina et al., 2022b). The present projections extend this historical assessment by demonstrating how similar efficiency tendencies may persist into future decades, while also revealing the degree to which such interpretations depend on forecast uncertainty and joint BUA–Pop dynamics.

However, it is important to note that uncertainty grows materially as the projection horizon extends. This increase in uncertainty is visible not only in the widening PIWs of BUA, Pop, LCR, PGR, and especially LCRPGR, but also in the probabilistic LUE outputs. For 2020–2025, when only one endpoint is forecast, about 55% of UCs fall in the high or very high confidence categories; by 2025–2030, when both endpoints are forecast, that share drops to about 29%, while about 45%

fall into the very low or low confidence categories. At the same time, the dominant regime shifts from *Efficient Built-up Expansion under Faster Population Growth* (R2) toward *Inefficient Built-up Expansion under Population Decline* (R12–R14), which rises from 39% to 55% of UCs. *Inefficient Built-up Expansion under Slower Population Growth* (R16) nearly disappears as a dominant class. The key implication is that projected regime prevalence should not be interpreted independently of confidence: a regime may become more common in median terms while becoming less certain at the same time. That distinction is central for policy interpretation, because it cautions against reading longer-horizon LUE maps as deterministic forecasts of future status.

A particular point of concern is the appearance of built-up reduction-associated LUE regimes (R4–R6, R8, and R10) in the 2020–2025 probabilistic classification results, where LCR is negative. This is difficult to reconcile with the historical record used for forecasting, in which all LCR values from 1975 to 2020 at 5-year intervals were positive. By contrast, negative PGR values were already present in the historical input data for some UCs.

One possible explanation could be the interaction between BUA and Pop in the forecasting process, since the two variables were modeled jointly. However, this explanation appears less convincing in the present case because the residual dependence analysis on the validation set showed only weak cross-variable dependence, which was subsequently simplified to independence in the uncertainty modeling. A more plausible explanation is that the forecasting model was not constrained to preserve non-decreasing BUA trajectories. As a result, it could yield lower 2025 BUA values for some UCs, particularly where recent historical patterns were weak, flat, or irregular. A further possible reason is that the 2025 and 2030 projections were generated from later input windows than those used for model training and evaluation. The model was trained on sequences ending in 2010 and evaluated through forecasts of 2015 and 2020, whereas the 2025 forecasts were based on more recent time steps. If the temporal structure of BUA changed in these later periods, the model may have extrapolated less reliably, contributing to projected BUA decline in some UCs. The projected occurrence of R4–R6, R8, and R10, therefore, appears unusual and deserving of validation, especially because these classes, particularly R4–R6, also appear in moderate to very high confidence categories.

5.6.7 Synthesis in Relation to RQ4

The results of this chapter provide a qualified but generally positive answer to RQ4. Across the six chapter objectives, it was shown that a global, joint-target, uncertainty-aware DL workflow can generate useful forecasts of BUA and Pop across diverse UCs and translate them into a forward-looking probabilistic assessment of SDG Indicator 11.3.1. The comparative evaluation identified GRU as the most balanced architecture among the tested models, while deep ensembling further improved the stability and calibration of the predictive distributions. Relative to the ES baseline, the global DL models consistently provided better point forecasts and more informative uncertainty estimates, indicating clear added value from pooled learning across localities and from modeling BUA and Pop jointly rather than through separate local univariate forecasts. These results support the view that uncertainty-aware forecasting is a feasible and meaningful basis for prospective LUE assessment at the global UC scale.

The chapter also showed, however, that the workflow’s usefulness is not uniform across all outputs or forecast settings. Forecast reliability was generally stronger for BUA than for Pop, and uncertainty propagation affected the derived components of SDG Indicator 11.3.1 unevenly. LCR remained comparatively stable, whereas LCRPGR was much more sensitive, particularly when projected population growth was weak and denominator effects became pronounced. Even under these challenges, the framework produced probabilistic LUE classifications that retained meaningful agreement with observed classes while also expressing regime confidence through dominant posterior probabilities. This practical outcome is important because it shows that the workflow does not merely generate future LUE class labels but also indicates how strongly those labels are supported under forecast uncertainty.

5.7 Limitations and Further Work

Several limitations should be considered when interpreting the findings of this chapter. First, the forecasting design relied on a fixed 8–2 input–output sequence, using eight historical observations to predict the next two time steps. While this provided a consistent experimental setup across models, it may not be optimal for all UCs or for both target variables. Different localities may benefit from shorter or longer memory lengths, and the fixed window may have limited the models’ ability to capture slower structural shifts or, conversely, shorter-term changes not well represented in the five-year series. As noted in Section 5.6.6, this issue is also relevant to the unusual appearance of built-up reduction-associated LUE regimes in the 2020–2025 projections. One possible limitation is that the forecasting stage relied on later input windows than those used in model training and evaluation, so the temporal structure encountered in forecasting may not have been fully represented in the original learning setup. This design choice likely contributed to the widening uncertainty observed at the later forecast horizon, particularly for Pop, and to the amplification of that uncertainty as it was propagated into PGR and LCRPGR.

Second, the forecasting workflow was intentionally restricted to BUA and Pop only, ensuring close alignment with the SDG Indicator 11.3.1 formulation but limiting the range of explanatory information available to the models. In the broader forecasting literature, global probabilistic models such as DeepAR and the Temporal Fusion Transformer are explicitly designed to incorporate static and time-varying covariates alongside the target series (Lim et al., 2021; Salinas et al., 2020). In urban growth and land-use studies, commonly used explanatory drivers include variables such as population density, distance to roads and city centers, slope, land prices, and zoning or planning constraints, all of which can help explain heterogeneity in growth trajectories and improve predictive performance or interpretability (e.g., Hu & Lo, 2007; Mahmoudzadeh et al., 2022). Their omission here means that the models captured shared temporal structure mainly through the historical BUA–Pop sequences themselves, without explicitly representing broader spatial, socioeconomic, infrastructural, or environmental factors that shape urban change.

Third, although the models were trained globally across many UCs, the forecasting framework did not explicitly account for geographic location or for spatial relationships and connectivity between UCs. In reality, urban change is not fully independent across localities; nearby or functionally

connected UCs may exhibit correlated development patterns through regional spillovers, shared infrastructure, migration systems, or broader economic linkages (Clark, 1998; Wang et al., 2022b). Because such spatial structures were ignored during the modeling process, the model has limited ability to represent inter-UC dependence beyond what is implicitly learned from pooled temporal data alone. Explicitly accounting for these factors is particularly relevant in a global UC system, where growth trajectories may be shaped not only by local history but also by a UC's position within wider spatial and networked urban systems.

Fourth, the workflow treated historical BUA and Pop as fixed inputs and assumed that the dependence structure inferred from historical forecast residuals remained reasonably stable over the forecast horizon. These assumptions made the uncertainty propagation feasible, but they also mean that the reported intervals primarily reflect forecast uncertainty. They do not fully capture the uncertainty already present in the EO-derived inputs or possible future changes in the dependence structure between BUA and Pop.

Fifth, a further limitation concerns the spatial form of the forecasts. The workflow developed in this chapter predicts UC-level BUA and Pop trajectories for forward-looking SDG Indicator 11.3.1 assessment, but it does not model the within-UC spatial configuration of future built-up change. This means that the approach can indicate how much BUA and Pop are expected to change, together with the resulting implications for projected LCR, PGR, LCRPGR, and probabilistic LUE classification, but not where built-up expansion or contraction is likely to occur within the urban extent. This delimitation is consistent with the objective of the chapter, which is to support large-scale, uncertainty-aware monitoring across thousands of UCs rather than spatial simulation of urban form. The historical BUA series used here are derived from EO-based 2D built-up maps, meaning that the underlying spatial information exists outside the present forecasting framework and could be analyzed separately in studies focused on spatial pattern or urban growth simulation.

Further work could address these limitations by testing alternative or adaptive input–output window configurations, rather than relying on a single 8–2 setup. In relation to the projection behavior discussed in Section 5.6.6, future work should also examine whether forecasts based on later rolling windows lead to systematic changes in BUA trajectory behavior, and whether structural constraints or post hoc plausibility checks are needed to reduce implausible projected declines in BUA.

Another line of future work is the incorporation of relevant exogenous covariates into the forecasting framework. Relevant static contextual factors may include geographic location, region, income class, density, or terrain, whereas dynamic covariates may include nighttime lights, accessibility, economic activity, policy or zoning information, and environmental constraints. Adding these covariates would allow the models to move beyond reliance on historical BUA–Pop trajectories alone and to better represent the broader drivers of urban change. This could improve predictive performance, strengthen interpretability, and help explain why forecast accuracy, interval width, and LUE regime stability vary across UCs.

Future extensions could also explicitly model spatial dependence and inter-UC connectivity, for example, through graph-based representations linking UCs by geographic proximity, regional connectivity, or functional interaction. In this regard, graph neural networks offer a promising direction

because they can combine temporal dynamics with spatial or network structure and may therefore improve both forecast skill and the representation of geographically structured uncertainty (Jin et al., 2024a; Luo & Chen, 2024).

Further work should also move beyond the assumption that historical BUA and Pop inputs are fixed and error-free. One possible extension is to integrate input uncertainty from EO-derived BUA and population datasets directly into the forecasting stage, for example, by drawing historical input trajectories from their uncertainty distributions or by training models that explicitly account for measurement uncertainty in the target and predictor series. If exogenous covariates are added to the forecasting framework, uncertainty in these covariates should also be considered, but as a secondary extension to the more fundamental issue of uncertainty in the core BUA and Pop inputs. In this broader formulation, the forecasting workflow would propagate uncertainty from the historical BUA–Pop series, from any auxiliary covariates used to support prediction, and from the forecasting model itself. This would allow forecast intervals to reflect both model uncertainty and uncertainty already present in the historical observations and auxiliary inputs. In addition, the dependence structure between BUA and Pop could be modeled dynamically rather than assumed to remain stable over the forecast horizon. Future studies could compare alternative dependence assumptions, update copula or residual-dependence models across rolling windows, or test whether dependence parameters vary by region, income group, urban size, or growth regime. These extensions would provide a more complete representation of uncertainty propagation from historical EO-derived inputs and supporting covariates to future BUA, Pop, LCR, PGR, LCRPGR, and probabilistic LUE classifications.

Future work could also link the present forecasting framework with spatially explicit urban growth or land-change simulation models to project not only the magnitude of future BUA change, but also its 2D spatial manifestation and its implications for spatially differentiated LUE assessment.

5.8 Chapter Conclusion

This chapter showed that a global, joint-target, uncertainty-aware DL workflow can generate useful probabilistic forecasts of BUA and Pop across diverse UCs and propagate them into forward-looking SDG Indicator 11.3.1 assessment. The main conclusion of this chapter is not that future LUE can be predicted with certainty, but that it can be assessed probabilistically in a way that is more transparent, more informative, and more consistent with the structural sensitivities of SDG Indicator 11.3.1 than deterministic forecasting alone. This conclusion should, however, be interpreted within the scope of the chapter, which treated the historical BUA–Pop sequences as fixed inputs and relied on a dependence structure inferred from out-of-sample residuals for subsequent copula-based propagation. Within those bounds, the chapter shows that data-driven, uncertainty-aware forecasting can support forward-looking, decision-relevant LUE assessment, while also making clear where predictive uncertainty remains substantial and where further methodological refinement is needed.

6 Enhancing LUE Assessment Through Built-Up Area–Built-Up Volume Trajectories²⁰

6.1 Chapter Overview

This chapter introduces the built-up area–built-up volume (BUA–BUV) trajectory framework to explicitly integrate horizontal and vertical dimensions of urban growth into LUE assessment. This chapter addresses the final research question: *RQ5. To what extent does integrating BUA and BUV within a trajectory-based framework enhance the interpretability and robustness of SDG Indicator 11.3.1–based LUE assessment?* The framework produces a trajectory-based typology of urban growth, enabling systematic classification of centers according to their prevailing built form and growth direction. The framework is motivated by limitations observed in existing approaches to incorporating vertical urban growth into efficiency assessments (Section 2.2.3.5). Unlike previous approaches that relied on proxy measures of verticality, were resource-intensive, or remained conceptually disconnected from SDG Indicator 11.3.1 monitoring, the proposed framework leverages harmonized global datasets on built-up area and volume. This framework enables analysis of the co-evolution of extent and form, and the distinction between horizontal expansion and vertical intensification. The integration of these trajectories with LCRPGR aims to improve interpretability without collapsing multiple growth processes into a single composite or higher-dimensional index whose interpretation is ambiguous.

6.2 Related Work

Beyond the SDG Indicator 11.3.1 framework, volumetric indicators have been proposed to more fully characterize urbanization processes by incorporating the vertical dimension of growth. For instance, Zięba-Kulawik et al. (2020) introduced the Building 3D Density Index (B3DI) to quantify changes in the volume of buildings and how it relates to urban expansion; Zambon et al. (2019)

²⁰This chapter is based on the author’s published article: Santillan et al. (2025b). *Enhancing land use efficiency assessment through built-up area–built-up volume trajectories: Integrating vertical urban growth into SDG 11.3.1 monitoring*, *ISPRS International Journal of Geo-Information*, 14(10), 404. <https://doi.org/10.3390/ijgi14100404>. In that article, the author’s contributions included conceptualization, methodology, data curation, formal analysis, visualization, writing of the original draft, project administration, and funding acquisition; writing–review and editing were undertaken jointly with the coauthors, and supervision was provided by Christian Heipke. Material from the published article is reused here in revised form. Portions of the text, figures, concepts, and methodological descriptions have been adapted for consistency with the structure, notation, and argumentation of this dissertation. The source article is an open-access publication distributed under the Creative Commons Attribution 4.0 International (CC BY 4.0) license; this chapter includes adapted material from that source.

proposed the Vertical-to-Horizontal Growth (VHG) ratio to discriminate vertical from horizontal urban expansion; and Kim & Kim (2024) developed a 3D Land Use Index (3D LUI) to quantify vertical land use. Other studies have examined the coupling between built-up land intensity and efficiency using volumetric or hybrid indicators (e.g., Ruan et al., 2022). Collectively, these works demonstrate that excluding volumetric information risks obscuring key structural and functional dimensions of urban growth.

Despite their conceptual value, existing volumetric indicators remain limited in their suitability for large-scale and standardized LUE monitoring. Many formulations collapse horizontal and vertical processes into aggregate ratios, thereby obscuring the distinction between expansion and intensification. Others depend on detailed building-level or height information that is not consistently available across space and time, or are not explicitly aligned with the operational logic of SDG Indicator 11.3.1. As a result, while volumetric indicators enrich urban analysis, they have yet to provide a scalable, interoperable solution for integrating vertical growth into LUE assessments.

Parallel to the development of volumetric indicators, trajectory-based and typological frameworks have been widely used to classify urban growth dynamics. Classical two-dimensional approaches track planimetric expansion patterns such as infill, edge expansion, and leapfrog development (Forman, 1995; Camagni et al., 2002; Jabareen, 2006; Schneider & Woodcock, 2008; Shi et al., 2012). Related work has also examined the co-evolution of built-up area and population over time (Sun et al., 2024), with some studies linking these trajectories to LUE assessments within the SDG Indicator 11.3.1 framework (e.g., Estoque et al., 2021; Santillan & Heipke, 2023, 2024a; Schiavina et al., 2019). While these approaches improve the interpretability of horizontal growth processes, they remain confined to planimetric change and do not capture the contribution of vertical development to efficiency outcomes.

More recent studies have attempted to integrate vertical dimensions in urban growth analysis. Mahtta et al. (2019) and Frolking et al. (2024) combined built-up extent, population, and microwave backscatter as a proxy for height to identify global urban growth modes. Although these approaches highlight the importance of vertical processes, their reliance on proxy measures limits temporal consistency, spatial transferability, and sensitivity to urban morphology. In addition, cluster-based typology classifications introduce methodological subjectivity, reducing reproducibility across contexts and weakening alignment with SDG Indicator 11.3.1 monitoring requirements.

Other studies rely on explicit three-dimensional data. Liu et al. (2024) proposed a global morphology typology based on planar density and average height, while Zhao et al. (2020) developed a LiDAR-based framework to identify processes such as intensification, efficient expansion, and sprawl. These approaches demonstrate the analytical value of explicit 3D information but remain constrained by static representations, heavy data requirements, or limited scalability. Importantly, none of these frameworks were designed to integrate directly with SDG Indicator 11.3.1, thereby reducing their applicability for standardized, globally comparable LUE assessment.

In summary, existing volumetric indicators and trajectory-based typologies highlight the importance of vertical urban growth but remain unsuitable for LUE monitoring. Proxy-based methods lack semantic and temporal consistency, LiDAR-based approaches are data-intensive and locally

constrained, and existing typologies are not interoperable with the SDG Indicator 11.3.1 framework. These limitations motivate the development of a reproducible trajectory-based approach that explicitly distinguishes horizontal expansion from vertical intensification, traces their co-evolution over time, avoids reliance on proxy or data-intensive inputs, and aligns directly with the operational logic of SDG Indicator 11.3.1. The BUA–BUV trajectory framework is proposed to address these requirements.

The framework was selected because it combines three essential properties absent in earlier methods: (i) it captures both horizontal and vertical processes without collapsing them into aggregate density ratios, (ii) it provides a transparent and reproducible trajectory-based classification that avoids the subjectivity of clustering or typology schemes, and (iii) it aligns with the scalability and comparability requirements of SDG Indicator 11.3.1 monitoring. These properties make the BUA–BUV framework a suitable and globally transferable basis for embedding vertical growth into LUE assessments.

6.3 The BUA–BUV Trajectory Framework

The framework conceptualizes urban growth as a trajectory in a two-dimensional space, with the x -axis representing BUA and the y -axis representing BUV. For each urban spatial unit u , a trajectory is constructed as an ordered sequence of observations across n temporal snapshots $t_1, t_2, \dots, t_n$, where each t_k ($k = 1, 2, \dots, n$) denotes a specific year or epoch. The raw trajectory is defined as

$$T_u = \{(BUA_{t_1}, BUV_{t_1}), (BUA_{t_2}, BUV_{t_2}), \dots, (BUA_{t_n}, BUV_{t_n})\}. \quad (6.1)$$

As illustrated in Figure 6.1a, this trajectory represents the temporal co-evolution of urban spatial extent and built volume as a sequence of connected line segments in BUA–BUV space. Each segment, constructed by linearly linking consecutive observations, conveys the direction and magnitude of growth between time steps. While this captures the scale and cumulative magnitude of change, raw trajectories are insufficient for systematic comparison across UCs. They do not allow for the analytical determination of the prevailing built form or growth modality, which are essential for linking the structural characteristics of urban development with LCRPGR outcomes.

Two issues explain this limitation. First, BUA and BUV are incommensurable, as they represent different physical dimensions and operate on different numerical ranges, rendering the direct interpretation of trajectory slopes ambiguous. Second, interpretation requires a reference trajectory against which relative growth modes can be assessed. Without such a benchmark, raw plots offer only subjective visual impressions, unsuitable for consistent typological classification. As a result, reliance on visual inspection prevents reproducible identification of growth modalities, whether predominantly horizontal, vertical, or balanced, across UCs and temporal intervals.

To overcome these limitations, the framework normalizes BUA and BUV to a common, dimensionless scale. Specifically, a min–max transformation is performed across all urban units and time steps, rescaling values to the $[0, 1]$ range. This normalization ensures that each trajectory is evaluated relative to the global distribution of observed BUA and BUV values. A normalized value of

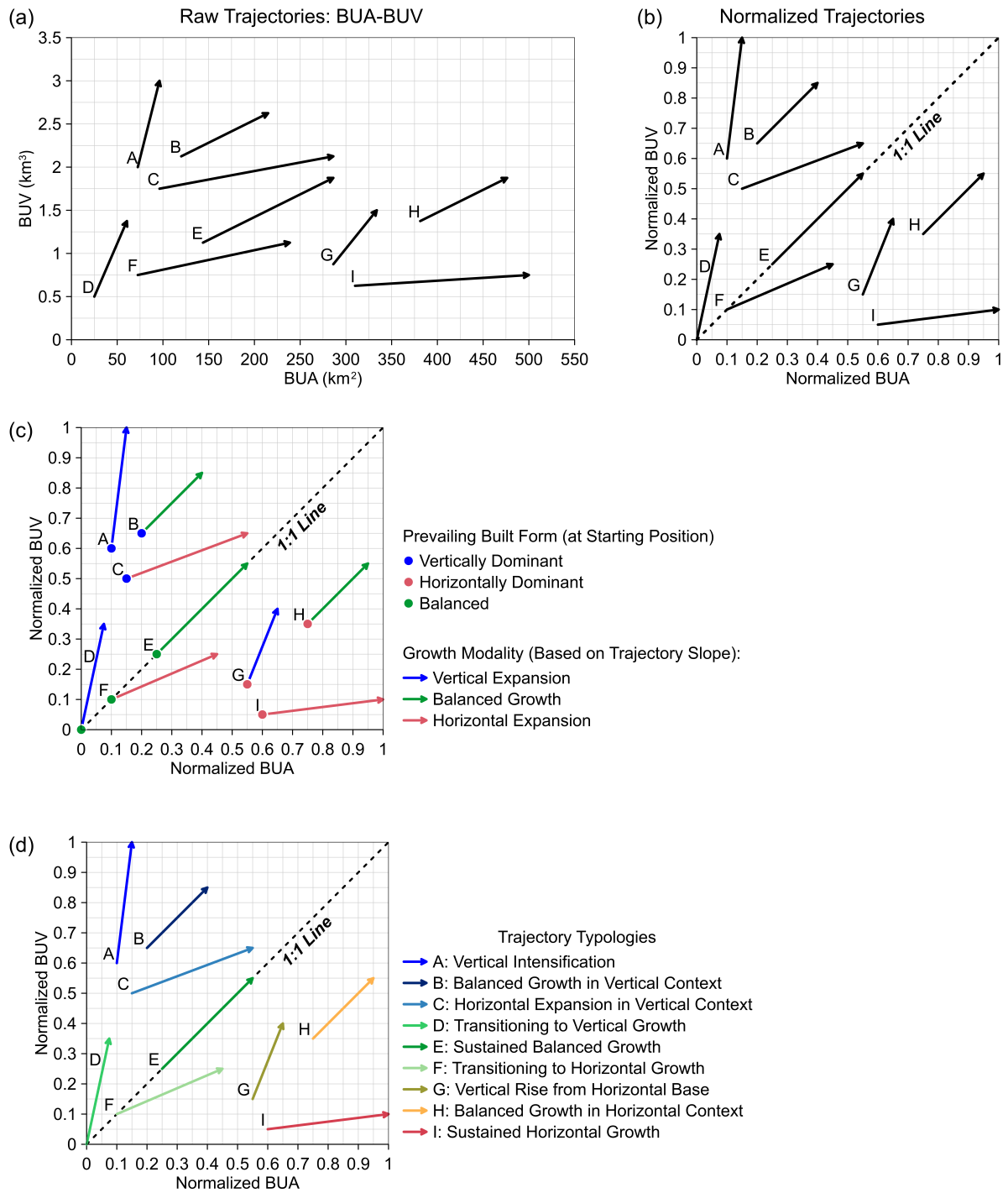


Figure 6.1: Hypothetical example illustrating the BUA–BUV trajectory framework for classifying urban growth typologies under the assumption of monotonic increases in BUA and BUV over time. (a) Raw BUA–BUV trajectories derived from plotted built-up area (BUA) and built-up volume (BUV) values at two epochs for nine hypothetical urban units (City A to City I), with connecting lines illustrating the direction and magnitude of growth between time steps. (b) Normalized trajectories, where BUA and BUV values are rescaled to the $[0, 1]$ range relative to the maximum across all cities and epochs, enabling comparability of growth directions independent of absolute size. (c) Classification of prevailing built form at the starting position, combined with growth modality based on trajectory slope. (d) Resulting trajectory typologies.

0 corresponds to the lowest observed expansion or volume, while a value of 1 corresponds to the highest, with intermediate values reflecting relative positions within the global range. This transformation creates a standardized basis for comparison, enabling consistent classification of growth modes independent of absolute size.

In this normalized trajectory $T_u^* \subseteq [0, 1]^2$ (Figure 6.1b), the values for each time step t_k are calculated as

$$\text{BUA}_{t_k}^* = \frac{\text{BUA}_{t_k} - \min(\text{BUA})}{\max(\text{BUA}) - \min(\text{BUA})}, \quad (6.2)$$

$$\text{BUV}_{t_k}^* = \frac{\text{BUV}_{t_k} - \min(\text{BUV})}{\max(\text{BUV}) - \min(\text{BUV})}, \quad (6.3)$$

where $\min(\text{BUA})$, $\max(\text{BUA})$, $\min(\text{BUV})$, and $\max(\text{BUV})$ are calculated from the full spatiotemporal dataset of all urban units and all time points.

The central premise of the BUA–BUV framework is the 1:1 line, which represents a state of proportional equilibrium between horizontal expansion and vertical intensification. This benchmark is meaningful only when both axes are normalized to the same $[0, 1]$ range. Under this condition, the line $y = x$ indicates that an urban unit’s relative position in BUA growth is identical to its relative position in BUV growth.

In the normalized BUA–BUV space, the position of trajectory points relative to the 1:1 line reflects their prevailing built form compared to the global sample. Because the 1:1 line captures equal relative change with respect to each variable’s observed range, it functions as a comparative reference. Trajectory points above it reflect proportionally larger upward expansion, those below it indicate dominant horizontal expansion, and those on or near it represent balanced growth, where horizontal expansion and vertical development proceed in comparable proportions (Figure 6.1c). Because normalization is performed across the whole dataset, these positions highlight relative morphological tendencies rather than absolute magnitudes.

Trajectory direction between two epochs indicates the prevailing growth modality. It is quantified as the slope of the trajectory in normalized space, calculated from the counterclockwise angle between the x -axis and the line segment connecting the earliest and latest time points. Shallow slopes indicate predominantly area-driven expansion, steeper slopes reflect volume-led intensification, and intermediate slopes suggest more balanced growth. Combined with a trajectory’s starting position relative to the 1:1 line, these directional measures enable consistent identification of urban growth typologies (Figure 6.1d). For example, a city whose initial trajectory point lies above the 1:1 line and follows a steep upward slope indicates pronounced vertical intensification, whereas one that starts below the line and follows a shallow slope signifies sustained horizontal growth. Transitional cases, such as centers on or near the 1:1 line, signal balanced growth trajectories that can shift toward vertical or horizontal dominance over time.

6.4 Operationalizing the BUA–BUV Framework for Typology Classification

To operationalize the classification, a “zone of effective balance” is first defined to capture trajectory points that fall sufficiently close to proportional equilibrium between horizontal expansion and vertical intensification. Because perfect balance is only a theoretical ideal, this zone is set as the 5th percentile of absolute orthogonal distances to the 1:1 line, calculated from the pooled global sample of all centers and years. Year-specific percentiles would introduce a moving decision boundary and confound temporal comparisons; the pooled threshold yields a fixed, globally defined tolerance that is consistent with global normalization and the trajectory interpretation. The prevailing built form of an urban unit at each time step is then determined by the position of its normalized trajectory vertices, or the trajectory’s endpoints in the case of only two time steps, relative to this band. Urban units within this band are classified as balanced, while those outside are assigned to vertical or horizontal dominance depending on the sign of deviation.

Second, growth modality is determined by slope categories. Slopes between 0° and 30° represent horizontal expansion, 30° to 60° correspond to balanced growth, and 60° to 90° indicate vertical intensification. While trajectory directions can, in principle, extend from -180° to 180° if either BUA or BUV decreases over time, such contractions are rare in practice, given the cumulative nature of EO-derived built-up area and volume (e.g., European Commission, 2023; Liu et al., 2024). For interpretive clarity, the present study restricts classification to the 0° – 90° domain, assuming monotonic increases in BUA and BUV. This assumption can be relaxed in future applications of the framework to explicitly accommodate contraction scenarios such as those resulting from natural disasters, armed conflict, or deliberate downsizing. By combining positional classification relative to the 1:1 line with slope-based growth modality, the framework establishes a systematic and reproducible scheme for deriving urban growth typologies that explicitly incorporate the vertical dimension (Figure 6.1d; Table 6.1).

Both conceptual and geometric considerations guide the choice of the 5th percentile for defining the balance band and the selection of the slope thresholds. The 5th percentile is adopted here as a conservative and data-driven threshold. Stricter values, such as the 1st percentile, would produce a narrower balance band and may exclude some urban units that are close to proportional equilibrium, particularly in the presence of data uncertainty. Broader values, such as the 10th percentile, would widen the balance band and may increase the likelihood that urban units with clear horizontal or vertical dominance are classified as balanced. In this sense, the 5th percentile provides a compromise by restricting the balanced class to urban units with relatively small deviations from the 1:1 line.

For the slope thresholds, dividing the 0° – 90° continuum evenly into three 30° intervals provides a neutral and symmetric partition that avoids privileging one growth mode over another. This scheme also aligns with the established triad of urban growth processes: horizontal expansion, vertical densification, and mixed forms (Frolking et al., 2024; Liu et al., 2024; Soltani et al., 2025). Moreover, slope angles are directly related to the ratio of normalized BUV change to normalized BUA change, since $\theta = \arctan(\Delta\text{BUV}^*/\Delta\text{BUA}^*)$. This makes the approach conceptually comparable to ratio-based indices such as the B3DI (Zięba-Kulawik et al., 2020) and VHG (Zambon et al., 2019), while

Table 6.1: Typology of urban growth patterns represented as a 3×3 matrix, based on the trajectory of normalized built-up area (BUA) and built-up volume (BUV). Rows correspond to the prevailing built form at the starting position, with trajectory starting position shown in parentheses; columns represent growth modes, with slope description and ranges shown in parentheses; and cells contain the typology names with corresponding interpretations. The classification assumes simultaneous increases in BUA and BUV, limiting trajectory slopes to $0\text{--}90^\circ$. If the decline of BUA or BUV is also considered, the slope domain extends to $-180^\circ\text{--}180^\circ$, yielding additional typologies not covered in this table. “On or near the 1:1 line” denotes balanced cases, defined through a zone of effective balance as described in Section 6.4.

	Vertical Expansion (Steeper, $60\text{--}90^\circ$)	Balanced Growth (Moderate, $30\text{--}60^\circ$)	Horizontal Expansion (Shallow, $0\text{--}30^\circ$)
Vertically Dominant (Above the 1:1 line)	A: Vertical Intensification (Growth is driven primarily by built-up volume; vertically dominant cities become more compact.)	B: Balanced Growth in Vertical Context (BUA and BUV increase at relatively balanced rates while retaining vertical dominance.)	C: Horizontal Expansion in Vertical Context (BUA increases more rapidly than BUV, but vertical dominance remains.)
Balanced (On or near the 1:1 line)	D: Transitioning to Vertical Growth (Urban form shifts from balance toward increasing vertical development.)	E: Sustained Balanced Growth (BUA and BUV grow proportionally; urban structure remains balanced.)	F: Transitioning to Horizontal Growth (Urban form shifts from balance toward increasing horizontal expansion.)
Horizontally Dominant (Below the 1:1 line)	G: Vertical Rise from Horizontal Base (Cities previously dominated by BUA show increased BUV; a shift toward vertical development.)	H: Balanced Growth in Horizontal Context (BUA and BUV increase at relatively balanced rates while retaining horizontal dominance.)	I: Sustained Horizontal Growth (Horizontal expansion continues to dominate; vertical growth remains limited.)

retaining an explicitly geometric interpretation. However, unlike B3DI, which summarizes vertical and horizontal contributions through a single density measure, and unlike VHG, which requires detailed building-level data or loses precision when applied at the aggregated level, the slope-based classification embeds both dimensions within a normalized trajectory framework and applies evenly defined regimes. In doing so, it avoids some of the interpretive limitations of aggregated ratio measures. Finally, the use of equal angular partitions ensures transparency and reproducibility. The classification is straightforward, interpretable across contexts, and independent of dataset-specific distributions. Unlike empirically derived thresholds, such as quartiles of observed slopes, which vary across datasets and reduce comparability, evenly spaced angular partitions provide a consistent, generalizable basis for classification.

6.5 Comparison with Other Trajectory Typology Classifications

Unlike clustering methods used in earlier studies for urban growth trajectory classification (e.g., Frohking et al., 2024; Mahtta et al., 2019), which are exploratory and often sensitive to algorithmic choices, the urban growth typology classification approach developed in this study is explicitly

theory-driven. It begins with a predefined conceptual benchmark—the 1:1 line of proportional equilibrium—and classifies urban units according to their prevailing built form (vertically dominant, balanced, or horizontally dominant) and their growth mode (horizontal expansion, balanced growth, or vertical intensification). This two-dimensional scheme yields interpretable categories (Table 6.1) such as “Vertical Intensification,” “Sustained Balanced Growth,” or “Sustained Horizontal Growth,” each directly linked to structural processes of urban change. The resulting typologies are transparent, reproducible, and designed to assess how urban units deviate from or approach structural balance. Thus, this trajectory classification scheme provides a more robust basis for interpreting LUE outcomes than unsupervised clustering approaches.

It is also possible to condense trajectories into a single scalar, such as the angle between the overall trajectory vector and the 1:1 line, but doing so removes critical contextual information. For instance, an urban unit that begins in a vertically dominant position and subsequently undergoes moderate horizontal expansion may yield an overall trajectory angle that suggests horizontal growth, even though its built form remains predominantly vertical. This loss of temporal and structural information limits the interpretability of scalar metrics. By retaining the full trajectory while classifying it into interpretable categories, the BUA–BUV framework balances methodological rigor with explanatory clarity, providing a stronger basis for assessing structural drivers of LUE outcomes.

6.6 Interpretation in the Context of SDG Indicator 11.3.1

The BUA–BUV trajectory-based typologies, when used in conjunction with LCRPGR, provide a more comprehensive basis for evaluating LUE and its alignment with sustainable urban development under SDG Indicator 11.3.1.

Urban units with similar LCRPGR values may belong to different trajectory typologies, and vice versa. Referring to the hypothetical example (Figure 6.1d), City A and City I might both fall within the “efficient” range, yet their growth typologies diverge significantly. City A’s steep trajectory above the 1:1 line indicates vertical intensification and compact urban form. At the same time, City I follows a shallow path below the line, suggesting continued horizontal expansion with minimal vertical development. Although both may appear numerically equivalent under the SDG Indicator 11.3.1 standard, their long-term sustainability implications differ: City A’s compact form supports resource efficiency and reduced land consumption, whereas City I may incur higher infrastructure costs and contribute to urban sprawl (Bibri et al., 2020).

Conversely, cities with similar spatial trajectories may exhibit varying LCRPGR values depending on their population growth dynamics. For instance, City G shows a steep upward trajectory from a low-rise base—suggesting a vertical correction or densification strategy—yet may still have an LCRPGR larger than one if its population growth remains stagnant. In such a case, the indicator might not yet reflect this progress despite the improving spatial form.

These distinctions underscore the value of interpreting LCRPGR together with BUA–BUV typologies. While LCRPGR provides a numerical measure of efficiency, the trajectories act as a spatial diagnostic tool that clarifies the structural pathways through which efficiency or inefficiency

arises and whether these pathways are likely to be sustainable in the long term. Importantly, the BUA–BUV framework is diagnostic by identifying whether growth is realized through horizontal expansion, vertical intensification, or balanced forms. However, it does not explain the socioeconomic, institutional, or cultural drivers behind these patterns. As such, the framework complements rather than replaces causal analyses of urbanization, providing a reproducible and scalable means to embed structural context into SDG Indicator 11.3.1 monitoring.

The framework’s application relies on BUA and BUV data that are already harmonized across time, with consistent boundaries and definitions. It further assumes that the inputs are free of major errors or extreme outliers. The procedures for detecting and removing such anomalies are described in the following section, where the framework is applied to assess LUE and global urban growth trajectories.

6.7 Application to Global Urban Centers for Enhanced LUE Assessment

6.7.1 Dataset Used, Preparation, and Temporal Scope of Analysis

The BUA–BUV trajectory framework was applied to global urban centers (UCs) using the GHS-UCDB 2025 dataset described in Section 5.4.1. As described there, the BUA and Pop attributes for each UC were obtained from the GHSL GHS-BUILT-S and GHS-POP layers, respectively. To incorporate the vertical dimension of urban form, this chapter used the BUV attribute provided in GHS-UCDB 2025. This UC-level attribute was derived from the GHS-BUILT-V product, which provides multitemporal built-up volume estimates based on GHS-BUILT-S and GHS-BUILT-H (European Commission et al., 2024). In turn, GHS-BUILT-H provides average building height estimates derived from AW3D30, SRTM30, and Sentinel-2 composites (European Commission, 2023; Pesaresi & Politis, 2023).

The analysis covered the period 1980–2020 and considered two 20-year intervals (1980–2000 and 2000–2020) as well as the cumulative period (1980–2020). These longer intervals were selected to capture major phases of global urban development while avoiding the lower reliability and limited structural change often associated with the native 5-year steps. This 20-year framing provided a more robust and interpretable basis for global trajectory analysis.

LCR, PGR, and LCRPGR were calculated for each UC for 1980–2000, 2000–2020, and the cumulative period 1980–2020, following Eqs. 2.1–2.3. Preliminary checks showed that the LCR and PGR distributions were highly skewed and contained implausible extreme values. To ensure consistency, the same data-cleaning and filtering procedure applied in Chapter 5 was used here: the modified IQR method of Barbato et al. (2011). This procedure was applied separately to LCR and PGR for each analysis period. A UC was excluded if any LCR or PGR value in any period fell outside the corresponding acceptance range. After this quality-control step, 10,856 UCs were retained for analysis (Figure 6.2). This retained sample is larger than the 8,478 UCs used in Chapter 5 because the exclusion rule here was less restrictive. Chapter 5 required complete temporal consistency across

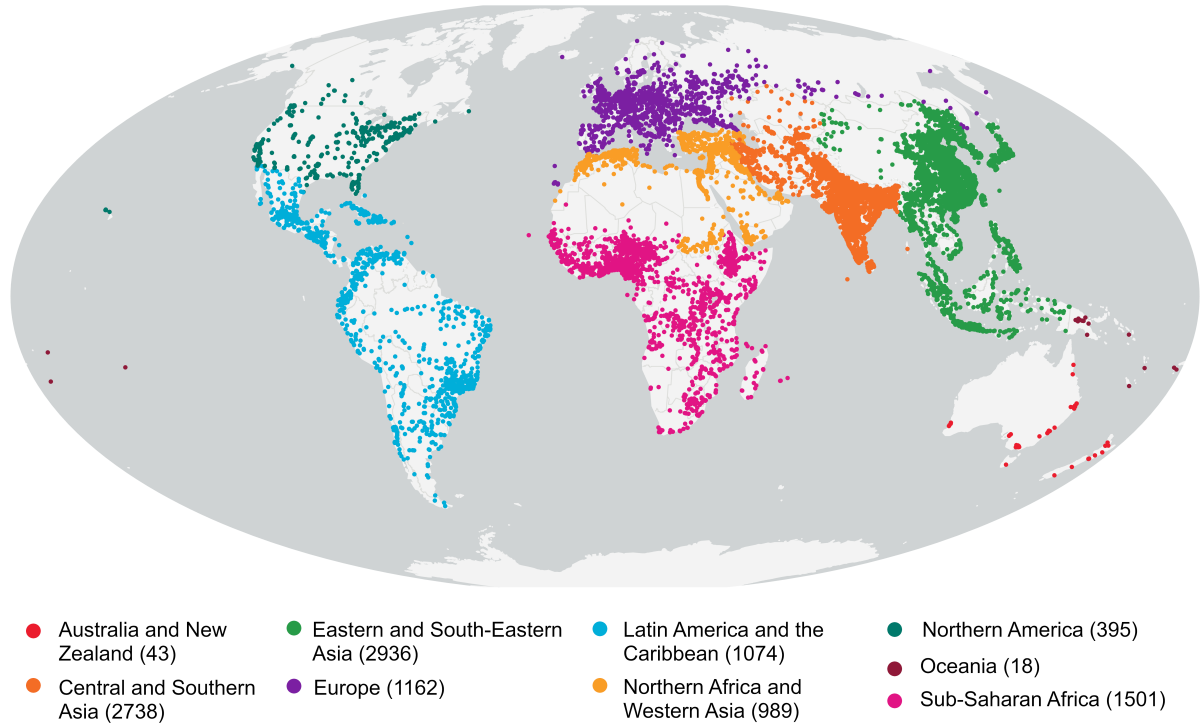


Figure 6.2: Geographic distribution of the 10,856 urban centers (UCs) included in this study, categorized by UN SDG region. Each point corresponds to an individual UC from the Global Human Settlement Layer Urban Centre Database (GHS-UCDB 2025). The number of centers per region is shown in parentheses following the region name.

all 5-year BUA and Pop transitions from 1975 to 2020 for forecasting, whereas the present chapter filtered only the LCR and PGR values derived for the selected 20-year and cumulative periods.

6.7.2 Conventional LUE Assessment

The LUE of each UC retained after quality control was assessed using the conventional SDG Indicator 11.3.1 calculation and interpretation framework described in Chapter 2. For each analysis period, frequency distributions of LCR, PGR, and LCRPGR were generated to describe the overall efficiency of global UCs. To provide a robust representation of central tendencies that is not influenced by distributional asymmetry, medians were used as the primary measure.

Using the LUE classification scheme (Table 2.1), each UC was assigned to an LUE regime for the three analysis periods: 1980–2000, 2000–2020, and 1980–2020. The number of UCs falling into each regime was then tabulated for each period. As in Chapters 4 and 5, UCs assigned to R4–R6 were grouped as *Efficient Built-up Reduction under Population Growth*, and those assigned to R12–R14 were grouped as *Inefficient Built-up Expansion under Population Decline*; all other regimes were retained individually.

6.7.3 Trajectory Typology Assignment and Efficiency Evaluation

BUA and BUV values for 1980, 2000, and 2020 were normalized using global minimum and maximum values (Eqs. 6.2 and 6.3). Each UC was then classified by prevailing built form (horizontal, balanced, vertical) and growth direction (horizontal expansion, balanced growth, vertical intensification) using the scheme outlined in Section 6.4. The combination of these two dimensions yielded nine possible trajectory typologies, as summarized in Table 6.1. Each UC was assigned to one of these typologies for the three analysis periods: 1980–2000, 2000–2020, and 1980–2020.

The LUE classifications were cross-tabulated with the trajectory typologies to evaluate how growth form and efficiency were interlinked. For each analysis period, the number and proportion of UCs in each LUE class were tabulated within every trajectory category. This tabulation enabled the identification of which trajectory pathways were most frequently associated with efficient or inefficient outcomes, thereby providing an integrated perspective on how horizontal and vertical growth modalities shaped LUE across time.

6.7.4 Statistical Assessment of Distributional Differences

To assess whether categorical distributions differed significantly, the chi-square (χ^2) test of independence, complemented by Cramer’s V , was applied to indicate the strength of association between the variables (McHugh, 2013). The analysis first tested whether the distribution of LUE classes differed significantly between the two multi-decadal periods. The second test examined differences in the prevalence of horizontally dominant, balanced, and vertically dominant built-form classes across years. The third test assessed whether the distribution of growth typologies exhibited statistically significant shifts between the two multi-decadal periods, irrespective of efficiency outcomes. The fourth test evaluated whether typology distributions changed significantly across periods when efficiency classes were considered separately. The final test examined whether efficiency classes were disproportionately associated with specific growth typologies within each period. For each analysis, the chi-square statistic, degrees of freedom (dof), p -value, and strength of association, Cramer’s V , were reported, and standardized residuals were inspected to identify the categories contributing most to significant departures from independence.

6.7.5 Sensitivity Analysis

The sensitivity of the BUA–BUV trajectory framework was examined by varying the parameters used to classify prevailing built form and growth modality. For built form, the zone of effective balance (Section 6.4) was modified by adjusting the threshold defining proximity to the 1:1 line. The baseline threshold was derived from the 5th percentile of absolute orthogonal distances across all UCs and years and was compared with stricter (1st percentile) and looser (10th percentile) alternatives. For growth modality, slope partitions were modified from the baseline $0^\circ\text{--}30^\circ/30^\circ\text{--}60^\circ/60^\circ\text{--}90^\circ$ scheme to alternative configurations of $0^\circ\text{--}25^\circ/25^\circ\text{--}55^\circ/55^\circ\text{--}90^\circ$ and $0^\circ\text{--}20^\circ/20^\circ\text{--}50^\circ/50^\circ\text{--}90^\circ$. Built-form and growth-typology assignments were recalculated for the three analysis periods, 1980–2000,

2000–2020, and 1980–2020, and compared with the baseline classifications using percentage agreement and Cohen’s kappa index (Cohen, 1960; Congalton, 1991).

Percentage agreement was evaluated using two complementary measures analogous to the well-established producer’s and user’s accuracies in remote sensing accuracy assessment (Congalton, 1991). Class retention rate, analogous to producer’s accuracy, quantified the share of UCs assigned to a given class in the baseline that remained in that class under an alternative specification. In contrast, assignment purity, analogous to user’s accuracy, measured the share of UCs assigned to a class in the alternative specification that also belonged to the same class in the baseline. Cohen’s kappa index was additionally reported to adjust for agreement expected by chance, thereby providing a more robust indicator of classification stability. Kappa (κ) values were interpreted following Landis and Koch’s (1977) scale: < 0.00 , poor; $0.00–0.20$, slight; $0.21–0.40$, fair; $0.41–0.60$, moderate; $0.61–0.80$, substantial; and $0.81–1.00$, almost perfect. These tests were exploratory rather than exhaustive and were intended to illustrate the framework’s sensitivity to reasonable variations in classification parameters.

6.8 Results

6.8.1 Temporal Trends in LCR, PGR, and LCRPGR

The temporal evolution of the SDG Indicator 11.3.1 component metrics was summarized in Table 6.2 for the 10,856 UCs covered in the analysis. Between 1980–2000 and 2000–2020, the global median LCR decreased from 3.47% to 1.72%, indicating a notable deceleration in horizontal land expansion. The median PGR also declined, from 1.97% to 0.95%, though at a slower rate. Consequently, the median LCRPGR²¹ dropped from 1.67 to 1.05, suggesting improved alignment between land consumption and population growth. When evaluated over the 1980–2020 period, the global median LCR remained high at 3.44%, while the median LCRPGR reached 1.94.

Table 6.2: Median LCR, PGR, and LCRPGR of global urban centers (UCs) for the periods 1980–2000, 2000–2020, and 1980–2020. Values are calculated at the UC level for each period and then summarized using the median across UCs. The reported LCRPGR values are therefore the medians of UC-level LCRPGR values and are not calculated as the ratio of the median LCR and median PGR shown in the table.

SDG Indicator 11.3.1 Components	1980–2000	2000–2020	1980–2020
LCR (% yr ⁻¹)	3.47	1.72	3.44
PGR (% yr ⁻¹)	1.97	0.95	1.52
LCRPGR	1.67	1.05	1.94

²¹The reported median LCRPGR was obtained by first calculating LCRPGR separately for each UC and then taking the median of the resulting UC-level LCRPGR values. Therefore, it should not be expected to equal the median LCR divided by the median PGR, because the medians of LCR, PGR, and LCRPGR are calculated independently from their respective UC-level distributions.

6.8.2 Spatiotemporal Patterns of LUE

Figure 6.3 shows the spatial distribution of UCs by LUE class. *Inefficient Built-up Expansion under Slower Population Growth* (R16) was the most common outcome, affecting 7442 UCs (69%) in 1980–2000 and 5610 UCs (52%) in 2000–2020. The share of the UCs classified as *Inefficient Built-up Expansion under Population Decline* (R12–R14) rose from 1,009 UCs (9%) to 2,121 UCs (20%), while the share of efficient centers (R2) increased modestly, from 2,405 UCs (22%) to 3,125 UCs (29%).

Statistical testing confirmed that these distributions differed significantly across the two periods ($\chi^2 = 745.9$, $\text{dof} = 2$, $p < 0.001$). The effect size was small to moderate, with Cramer’s $V = 0.19$, indicating that while the association between period and LUE outcome was not strong, the shift was systematic and consistent across the global sample. Inefficient expansion (R16) cases were overrepresented in 1980–2000, whereas efficiency (R2) and population-decline-associated inefficient expansion (R12–R14) cases were disproportionately common in 2000–2020. These results provide statistical evidence of a structural shift in efficiency outcomes, from widespread land-consumption-driven inefficiency in the earlier period toward a more mixed profile in the later period, marked by both efficiency gains and increasing inefficiency driven by population decline.

6.8.3 BUA–BUV Trajectories of UCs

The co-evolution of BUA and BUV from 1980 to 2020 is illustrated in Figure 6.4, which includes all UCs (panels a–b) and the top 10 most populous UCs (panels c–d). In the raw BUA–BUV space (panels a and c), cities such as Tokyo, Guangzhou, and Shanghai exhibit long trajectories extending toward the upper-right quadrant, indicating substantial absolute increases in both BUA and BUV. However, this representation primarily captures the scale and cumulative magnitude of growth, providing limited insight into each UC’s prevailing built form and directional characteristics relative to others. This plot alone is insufficient for reliably classifying UCs by urban growth trajectory in a systematic or replicable manner.

Normalized trajectories (Figure 6.4b–d), which rescale growth to a standard 0–1 range, enable more meaningful comparisons across UCs of varying sizes. Aided by the 1:1 line, these plots reveal heterogeneous developmental pathways. Some UCs show steep upward trajectories aligned with vertical intensification, while others follow flatter paths consistent with horizontal expansion. For example, among the top 10 most populous UCs, Tokyo, Guangzhou, and Seoul exhibit vertically oriented trajectories. Jakarta, Dhaka, and Manila maintain trajectories indicative of sustained horizontal growth. These contrasting patterns highlight the interpretive value of the trajectory framework in distinguishing spatial development strategies that are not readily discernible using absolute values alone.

6.8.4 Prevailing Built Forms of UCs

Figure 6.5 illustrates the evolving spatial distribution of prevailing built form types, which is one of the outputs of the BUA–BUV trajectory framework. In 1980, most UCs (9,306 UCs or 86%),

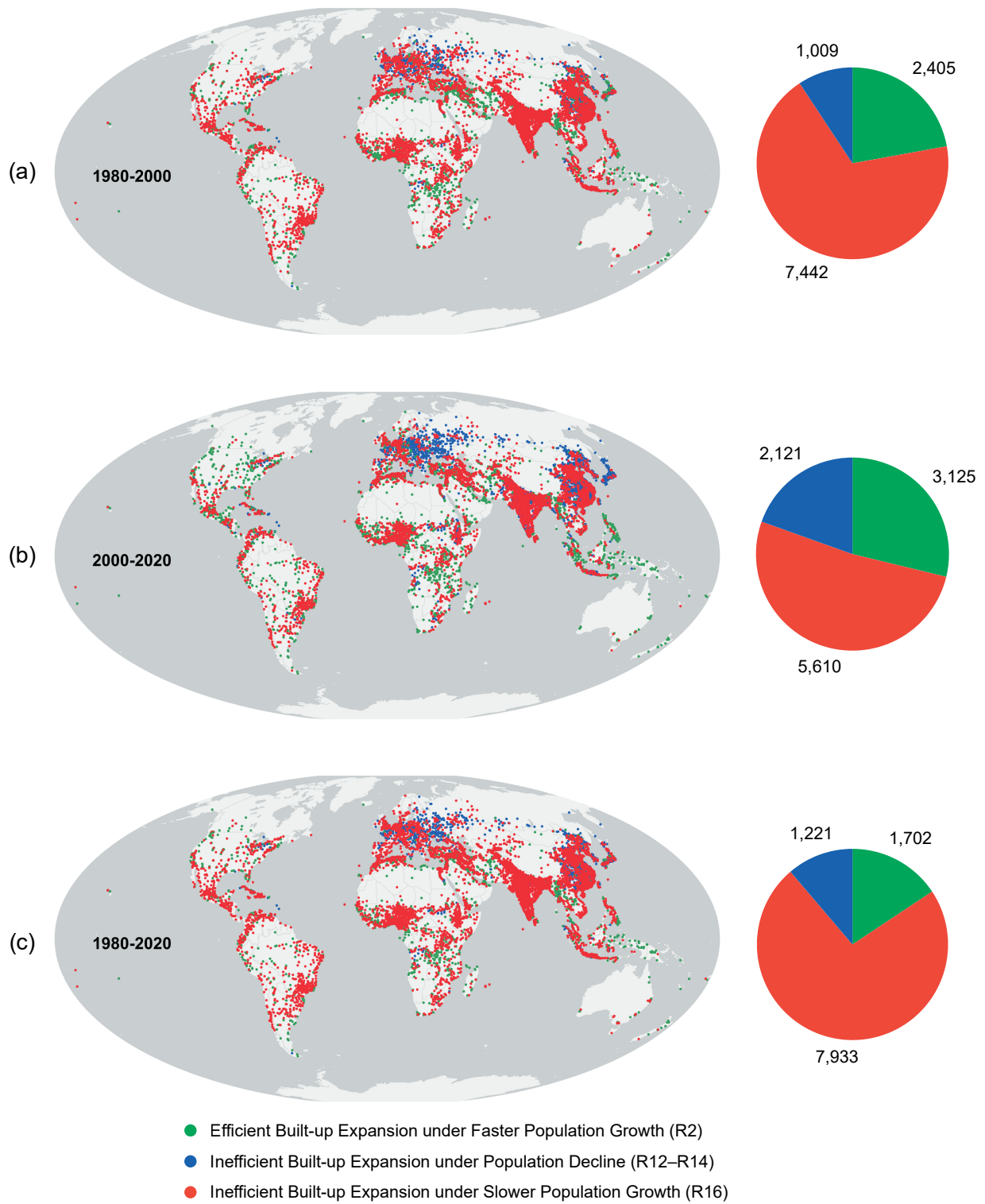


Figure 6.3: Distribution of LUE classes across global UCs for the periods (a) 1980–2000, (b) 2000–2020, and (c) 1980–2020. The pie charts on the right show the number of UCs assigned to each LUE class for the corresponding period.

were classified as horizontally dominant, while 1,172 UCs or 11% exhibited a balanced built form, and 378 UCs or 3% were vertically dominant. Horizontally dominant UCs were geographically widespread, occurring across various SDG regions. In contrast, UCs with balanced built forms

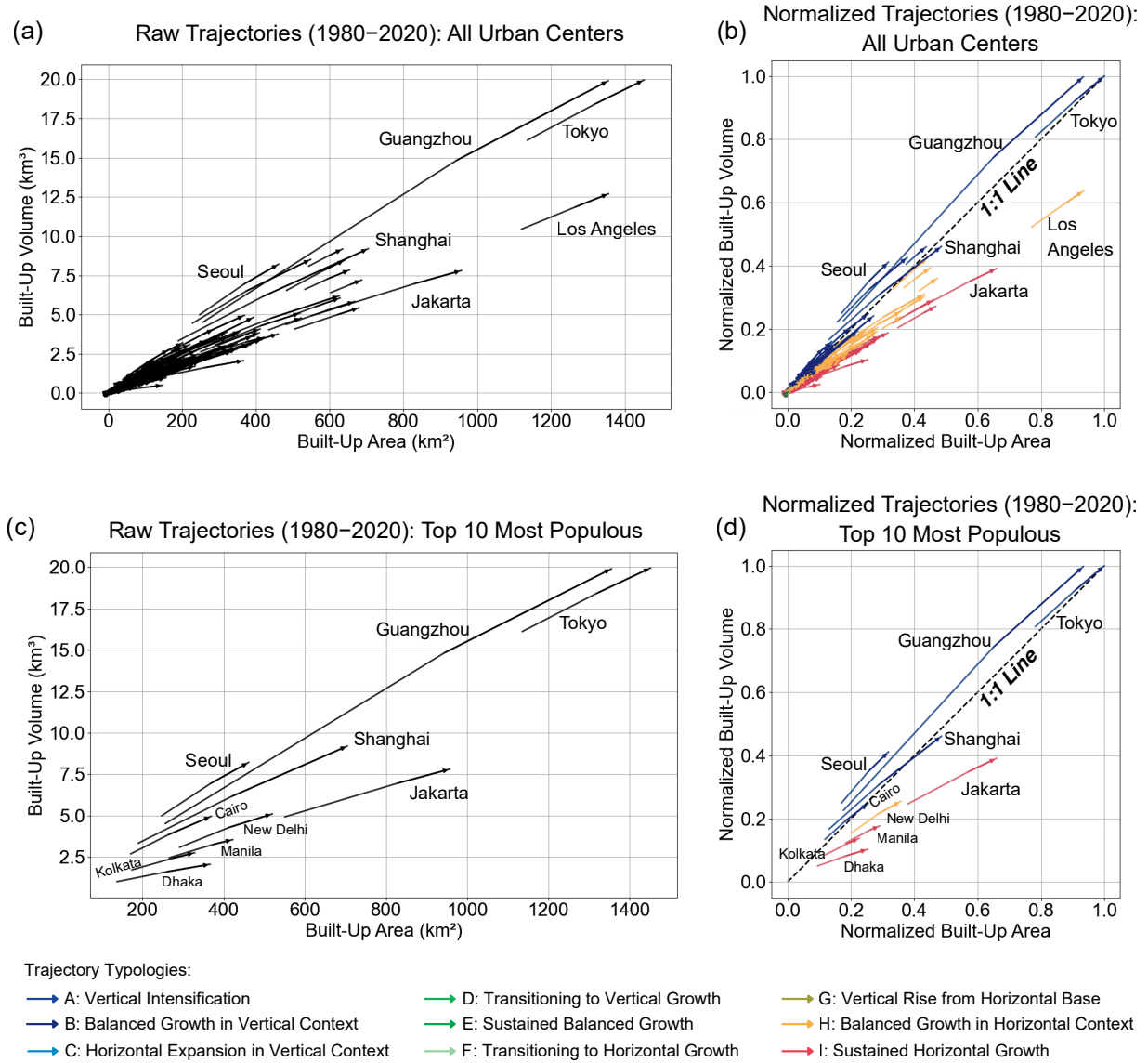


Figure 6.4: Trajectories of urban centers (UCs) in raw and normalized BUA–BUV space from 1980 to 2020. Panels (a, b) show raw and normalized trajectories for all UCs in the dataset. Panels (c, d) present the corresponding trajectories for the top 10 most populous UCs in 2020 based on the Global Human Settlement Layer Urban Centre Database (GHS-UCDB 2025).

were disproportionately concentrated in Asia, accounting for 76% of all balanced UCs globally. Notably, Asia also hosted the largest number of vertically dominant UCs, with 323 out of the 378 located primarily in countries such as China, Japan, and South Korea. By 2020, the proportion of horizontally dominant UCs had increased to 97%, or 10,528 UCs, while the shares of balanced and vertically dominant centers declined to 1%, 143 UCs, and 2%, 185 UCs, respectively, both categories remaining primarily concentrated in Asia.

Statistical testing confirmed that these shifts in the global distribution of built forms are statistically significant. However, the strength of association is modest ($\chi^2 = 1266.9$, $\text{dof} = 4$, $p < 0.001$, Cramer's $V = 0.14$). Balanced forms were disproportionately frequent in 1980 but declined sharply thereafter, while horizontal forms became increasingly more prevalent than expected if built form and time period were independent by 2020. Vertical forms, which were slightly overrepresented in

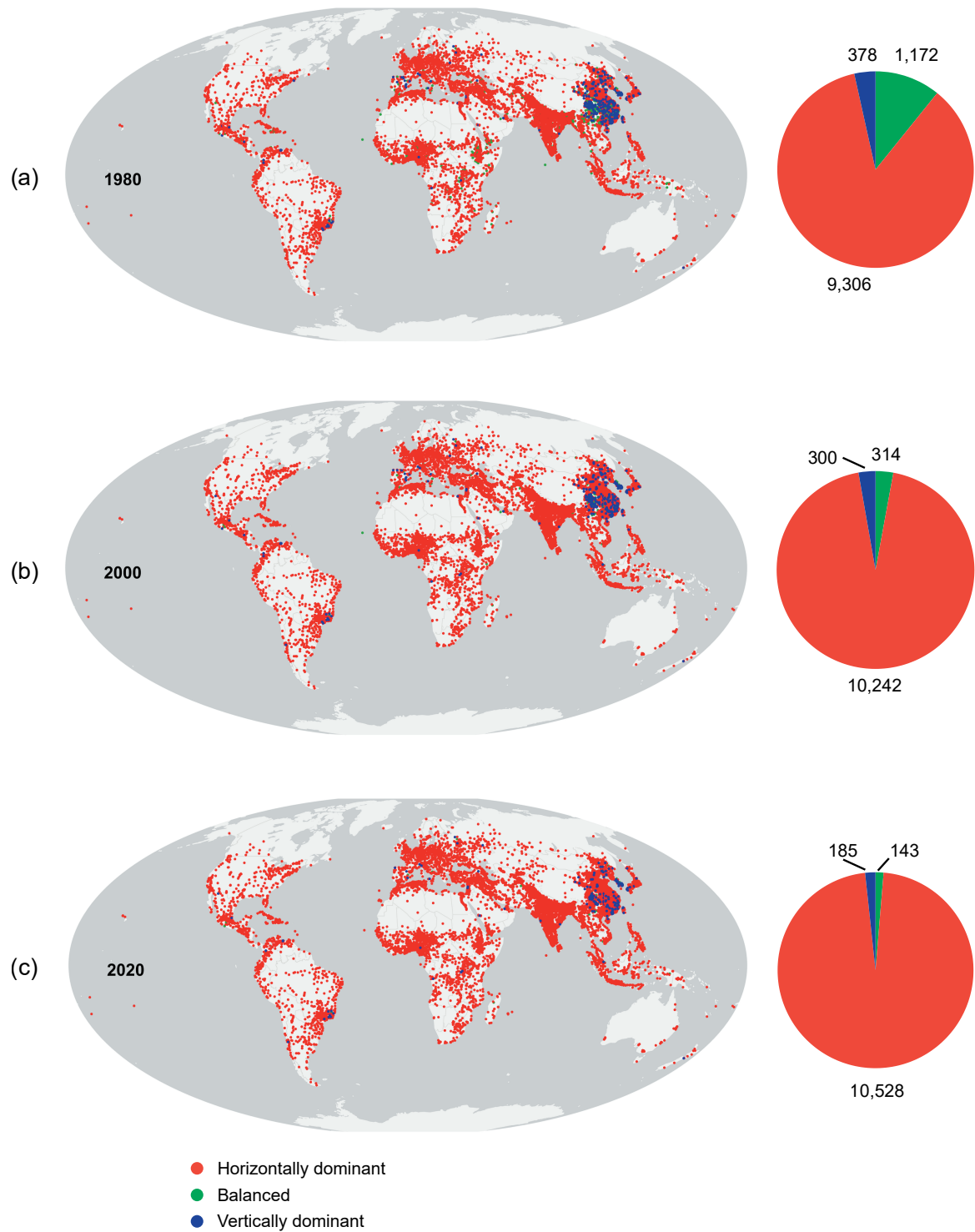


Figure 6.5: Distribution of prevailing built form dominance among urban centers (UCs) in (a) 1980, (b) 2000, and (c) 2020 based on normalized BUA–BUV values. The pie charts on the right show the number of UCs assigned to each built-form class for the corresponding period.

1980, fell below expectations under independence by 2020. These results indicate a global shift toward horizontal dominance accompanied by a reduction in balanced and vertical forms over the four-decade period.

6.8.5 Trajectory Typologies of UCs

Trajectory typologies derived from the normalized BUA–BUV space (Figure 6.4b) and their corresponding spatial distributions (Figure 6.6) reveal the dominant structural evolution pathways of UCs. Over the 1980–2000 and 2000–2020 intervals, *Sustained Horizontal Growth* (Typology I) was the most common growth pattern. This typology accounted for 6,307 UCs, or 58%, in the earlier period, and rose to 74% in the latter, reinforcing horizontal expansion as the prevailing global trend. In both periods, UCs in *Central and Southern Asia*, *Eastern and South-Eastern Asia*, and *Sub-Saharan Africa* collectively accounted for over 60% of all cities exhibiting this growth pattern.

Balanced Growth in Horizontal Context (Typology H) was the second most common growth pattern, accounting for 28% of UCs in 1980–2000 but declining to 20% in 2000–2020. During the first period, this typology was concentrated mainly in *Eastern and South-Eastern Asia* and *Europe*, accounting for approximately 53% of all UCs in this category. By the end of the second period, *Eastern and South-Eastern Asia* significantly expanded its share, representing about 48% of UCs under this typology. At the same time, most other regions experienced a decline in their numbers.

Typologies indicative of more structurally balanced or transitional growth, such as *Sustained Balanced Growth* (Typology E) and *Transitioning to Horizontal Growth* (Typology F), were relatively uncommon. Typology E, mostly observed in *Eastern and South-Eastern Asia*, declined from 6% to 2% between the two periods. Typology F similarly decreased from 5%, primarily concentrated in *Central and Southern Asia* and *Sub-Saharan Africa*, to just 0.7%.

Typologies associated with substantial vertical intensification, namely *Vertical Intensification* (Typology A), *Transitioning to Vertical Growth* (Typology D), and *Vertical Rise from Horizontal Base* (Typology G), were absent in both periods. Typology B, *Balanced Growth in Vertical Context*, was observed in approximately 3% of all UCs in both periods. This typology was heavily concentrated in *Eastern and South-Eastern Asia*, which accounted for over 80% of all UCs in this class. Typology C, *Horizontal Expansion in Vertical Context*, was observed in only 6 UCs, or 0.06%, in 2000–2020, all located in *Eastern and South-Eastern Asia* and *Europe*.

The chi-square analysis further demonstrated that trajectory typology distributions varied significantly across the three multi-decadal periods ($\chi^2 = 888.2$, $\text{dof} = 5$, $p < 0.001$; Cramer's $V = 0.20$). The association strength was moderate, similar to the previous analysis of prevailing built forms, and the standardized residuals revealed systematic shifts in growth pathways. In 1980–2000, balanced and transitional types, namely Typologies E, F, and H, occurred more frequently than expected if typology and period were independent. By 2000–2020, these categories fell to levels well below those expected under independence, while *Sustained Horizontal Growth* (Typology I) was strongly overrepresented. These results point to the consolidation of horizontal expansion as the dominant long-term trajectory, coupled with the retreat of balanced and transitional growth forms.

6.8.6 Linking Urban Growth Typologies to Efficiency Outcomes

Figure 6.7 presents the percentage distribution of BUA–BUV trajectory typologies across the three LUE classes for the periods 1980–2000 and 2000–2020. During 1980–2000, the *Efficient Built-*

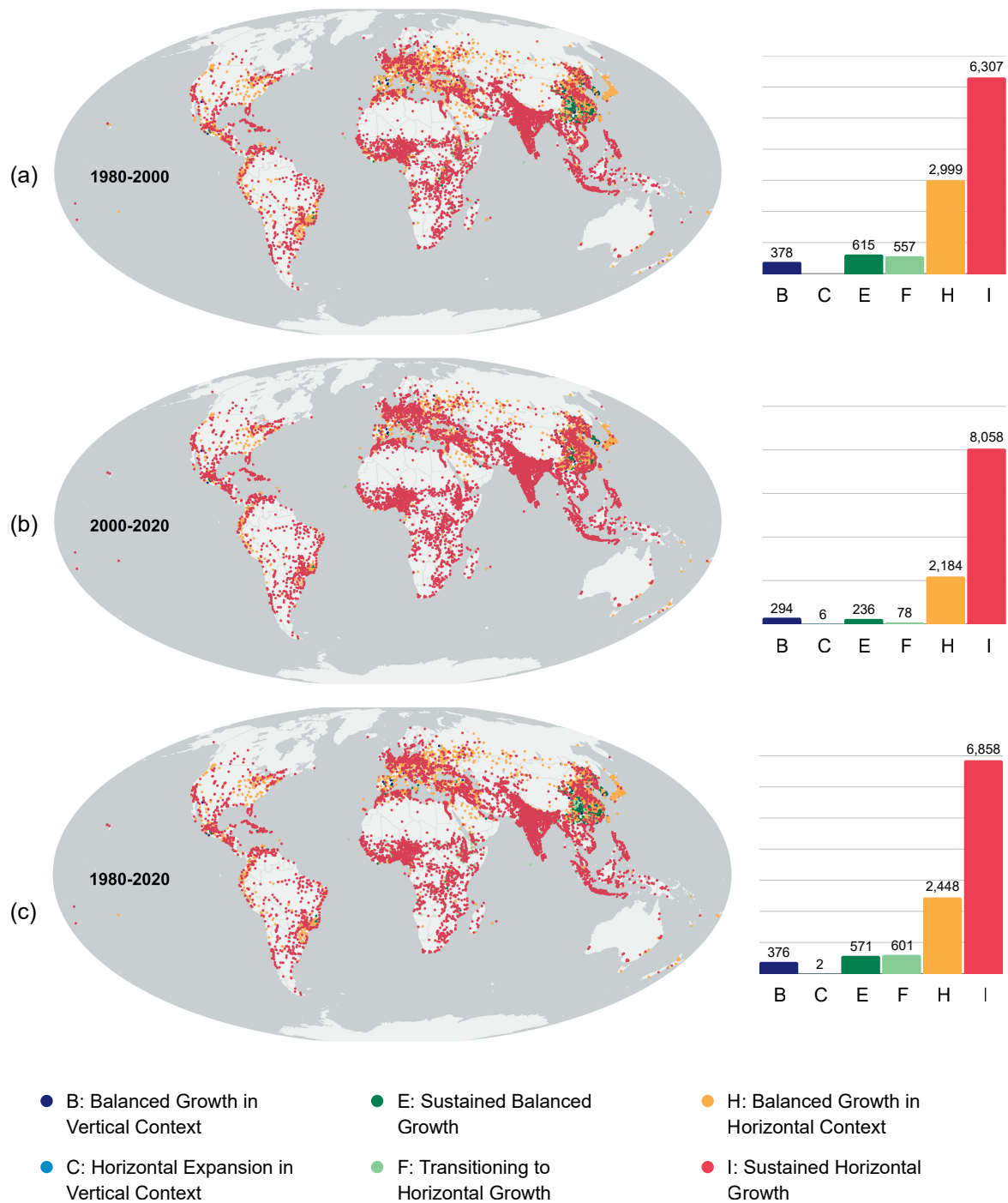


Figure 6.6: Global distribution of urban centers (UCs) by BUA-BUV trajectory typology for (a) 1980-2000, (b) 2000-2020, and (c) 1980-2020. Each point on the map represents one UC classified into one of nine typologies based on prevailing built form and growth direction. Only six of the nine typologies were observed among the UCs for the three analyzed periods. The bar charts on the right show the number of UCs assigned to each trajectory typology for the corresponding period.

up Expansion under Faster Population Growth (R2) class was strongly dominated by Typology I, with 1,480 of 2,405 UCs (62%) exhibiting sustained horizontal growth. The class characterized by *Inefficient Built-up Expansion under Population Decline* (R12-R14) was also primarily associated with Typology I, with 349 of 1,009 centers (35%), alongside a substantial contribution from Typology

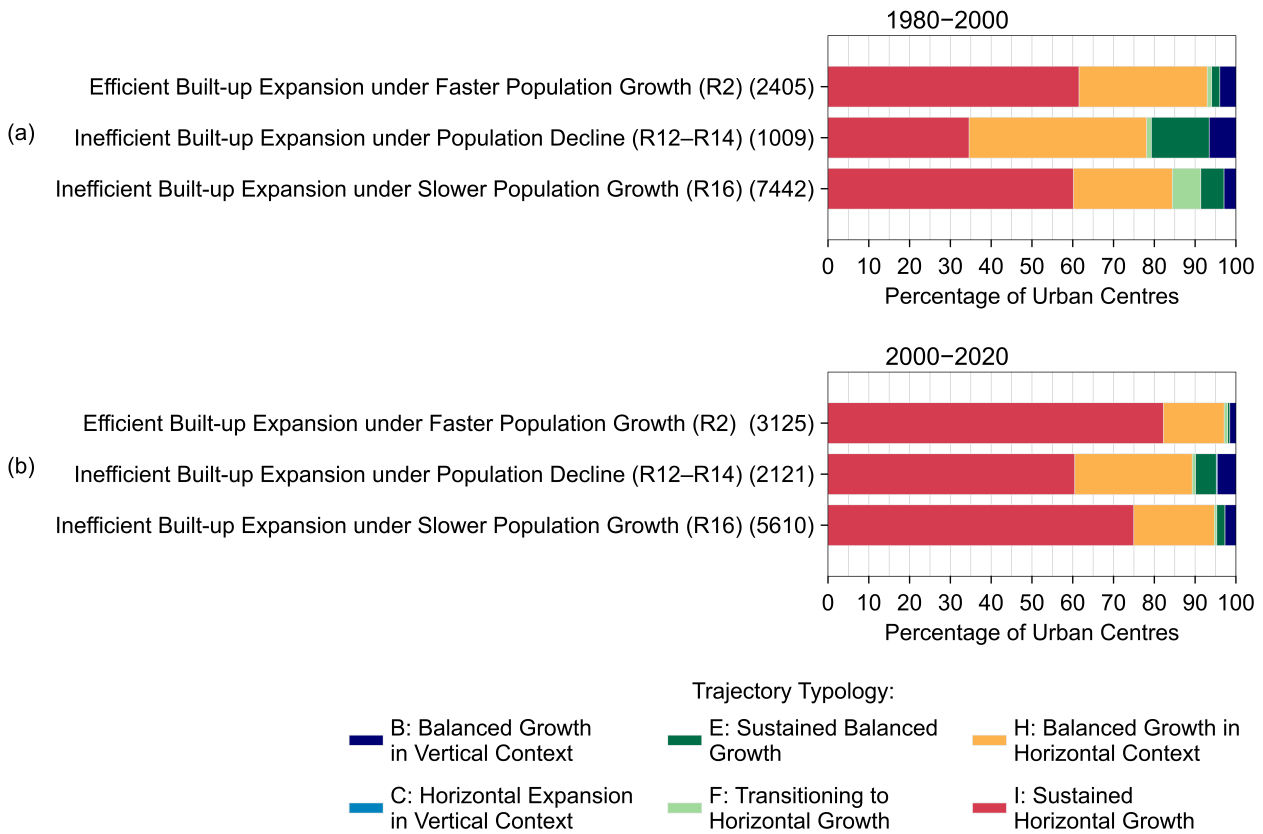


Figure 6.7: Percentage distribution of urban centers (UCs) across BUA–BUV trajectory typologies for three LUE classes during (a) 1980–2000 and (b) 2000–2020. The total number of UCs per LUE class is indicated in parentheses.

H, *Balanced Growth in Horizontal Context*, with 449 centers (44%). In the *Inefficient Built-up Expansion under Slower Population Growth* (R16) class, Typology I again prevailed, accounting for 4,487 of 7,442 centers (60%), further underscoring the strong association between sustained horizontal expansion and spatially inefficient land use.

The dominance of Typology I persisted into the 2000–2020 period. Around 60% of UCs categorized as *Inefficient Built-up Expansion under Population Decline*, or 1,283 of 2,121 centers, followed a sustained horizontal growth trajectory, as did 4,205 of the 5,610 centers within the *Inefficient Built-up Expansion under Slower Population Growth* class. Even among efficient centers, horizontal expansion was predominant, with 2,570 of 3,125 centers (82%) classified under Typology I.

Across all LUE classes and both periods, vertical and balanced typologies, such as Typologies B and E, remained relatively rare. Typology B, *Balanced Growth in Vertical Context*, reflects efficiency achieved within a vertical-dominant regime, while Typology E, *Sustained Balanced Growth*, indicates proportional growth in both horizontal and vertical dimensions. Both can be regarded as structurally sustainable trajectories, yet their combined presence was minimal. In the 2000–2020 period, only 47 UCs followed Typology B and 16 followed Typology E while also achieving efficient land use.

Chi-square tests confirmed that these typology distributions differed significantly across periods within each efficiency class ($p < 0.001$ in all cases). The strength of association was consistently moderate, with Cramer's V values ranging from 0.21 to 0.26 across LUE classes, indicating sys-

tematic but not strong shifts in composition over time. Additional tests across efficiency classes within each period also showed significant associations ($p < 0.001$ in all cases; Cramer's $V = 0.17$ for 1980–2000 and 0.13 for 2000–2020). These effect sizes point to weaker but still significant links between efficiency outcomes and growth typologies, with standardized residuals highlighting the persistent overrepresentation of sustained horizontal growth and the underrepresentation of structurally sustainable forms.

6.8.7 Sensitivity of the BUA–BUV Trajectory Framework

6.8.7.1 Effect of Thresholds on Prevailing Built Form

Figure 6.8 and Table 6.3 summarize the effects of adjusting the threshold used to define the zone of effective balance on the stability of prevailing built form classifications, using 2000 and 2020 as case examples.

The most pronounced changes were observed in the balanced category (Figure 6.8). Under the stricter 1st percentile threshold, only about one-quarter of the UCs classified as balanced in the baseline scheme were retained, 25% in both years. By contrast, the horizontal class showed complete retention and near-perfect purity, while the vertical class showed complete retention but lower purity. Assignment purity was 100% for the balanced class, but only about 80% in 2000 and 87% in 2020 for the vertical class, indicating that some centers assigned as vertical under the stricter threshold had belonged to other classes in the baseline classification. The horizontal class remained highly consistent, with purity of about 98–99%. Overall agreement with the baseline was still very high, 98% in 2000 and 99% in 2020, but Cohen's kappa was lower, 0.77–0.81, reflecting reduced agreement after accounting for chance (Table 6.3).

When the balance zone was widened to the 10th percentile, the balanced class composition changed substantially (Figure 6.8). In this case, balanced cases achieved full retention, 100%, but lower assignment purity, about 50% in 2000 and 57% in 2020, reflecting the inclusion of many UCs that had been classified as horizontal or vertical in the baseline scheme. For this wider threshold, the horizontal and vertical classes showed the opposite pattern from the 1st percentile case: both had perfect purity, but their retention was slightly lower. Horizontal and vertical classes retained high purity, 100%, but somewhat lower retention, about 97–99% for horizontal and 85–91% for vertical. Overall agreement remained high, 97% in 2000 and 99% in 2020, and kappa values increased slightly, 0.78–0.85, relative to the 1st percentile threshold, indicating substantial to almost perfect agreement (Table 6.3).

The collapsed two-class analysis, which contrasted balanced cases with all others, confirmed that the balanced class was the most threshold-sensitive group (Table 6.3). For the 1st percentile threshold, kappa dropped to 0.40 in both 2000 and 2020, corresponding to only fair agreement. By contrast, the 10th percentile threshold yielded stronger performance and substantial agreement with the baseline, with kappa improving to 0.65 in 2000 and 0.72 in 2020.

These results demonstrate that the horizontal class is the most robust to threshold variation, whereas the balanced class is strongly parameter-dependent. The vertical class is also relatively

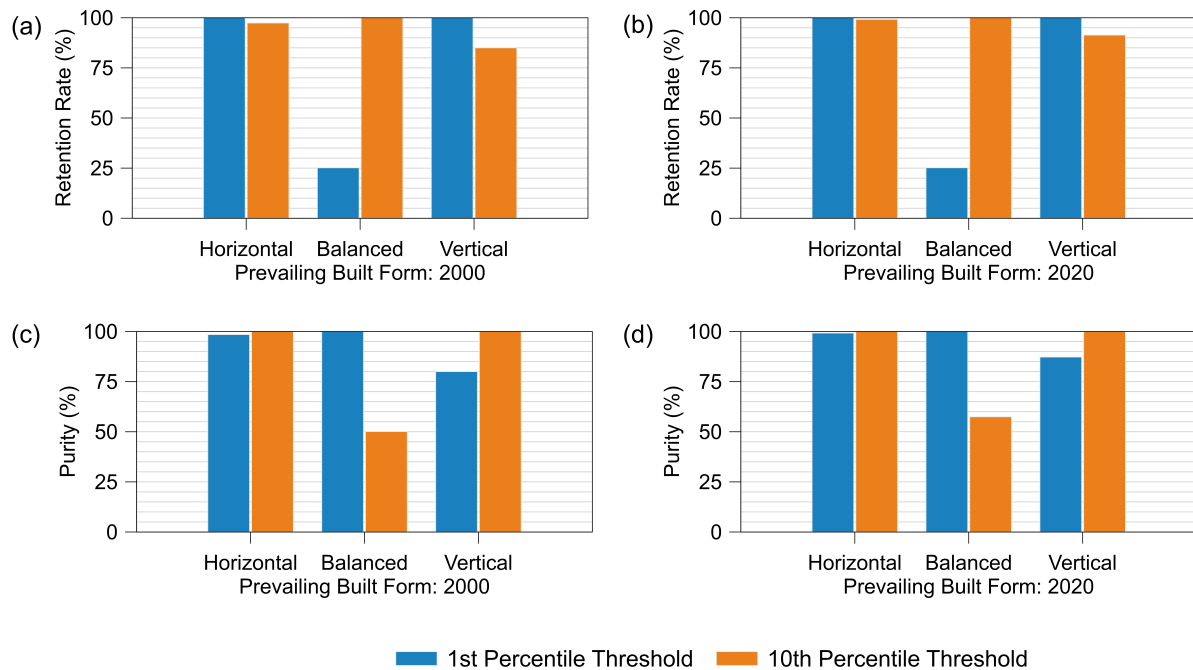


Figure 6.8: Sensitivity of prevailing built form classification of urban centers to the zone of effective balance thresholds, comparing the 1st and 10th percentile scenarios relative to the baseline 5th percentile classification. Panels (a, b) show class retention rates for 2000 and 2020, respectively, while panels (c, d) show assignment purity for 2000 and 2020, respectively.

Table 6.3: Overall agreement and Cohen's kappa index with the baseline (5th percentile) prevailing built form classification for 2000 and 2020 under alternative effective zone of balance thresholds (1st and 10th percentiles).

Metric	Year	1st Percentile	10th Percentile
Overall Agreement	2000	98%	97%
Overall Agreement	2020	99%	99%
Cohen's kappa (Three-class case: Horizontal, Balanced, Vertical)	2000	0.77	0.78
Cohen's kappa (Three-class case: Horizontal, Balanced, Vertical)	2020	0.81	0.85
Cohen's kappa (Two-class case: Balanced vs. All Other Built Forms)	2000	0.40	0.65
Cohen's kappa (Two-class case: Balanced vs. All Other Built Forms)	2020	0.40	0.72

stable, but less so than the horizontal class, because its retention and purity shift depending on whether the threshold is tightened or widened. The balanced category either contracts sharply under stricter definitions or expands markedly under looser ones, leading to systematic variations in retention, assignment purity, overall agreement, and kappa.

6.8.7.2 Effect of Slope Parameters on Trajectory Typology

Tables 6.4 and 6.5 summarize how the stability of trajectory typology classifications varies relative to the baseline scheme depending on the threshold and slope partition applied. Adjusting only the percentile threshold, 1st, 5th, and 10th percentiles, while keeping the baseline slope partition resulted in minimal change, with overall agreement remaining very high, 97–98%, and $\kappa = 0.93$ –0.94, indicating almost perfect agreement. In contrast, altering the slope partitions substantially reduced stability: the 0° – 25° / 25° – 55° / 55° – 90° partition produced agreement of 73–76% with $\kappa = 0.51$ –0.55, corresponding to moderate agreement, while the 0° – 20° / 20° – 50° / 50° – 90° partition yielded only 49–51% agreement with $\kappa = 0.24$ –0.27, corresponding to fair agreement. These results show that overall classification agreement is far more sensitive to slope partitioning than to percentile threshold choice.

Beyond overall agreement, the detailed retention and purity analysis highlights how parameter combinations differentially affect individual trajectory categories (Figure 6.9). Balance-related typologies, namely B, E, and H, show high retention stability across most parameterizations, but purity is sensitive to slope partitioning, with H particularly prone to misclassification. Horizontal-leaning categories, C, F, and I, exhibit the most substantial volatility, with retention rates often collapsing under alternative settings. Stricter balance thresholds, such as the 1st percentile, fragment transitional classes, while looser thresholds, such as the 10th percentile, inflate balanced categories, E and H, at the expense of purity. This finding indicates that typology distributions are not invariant and that parameterization choices significantly affect the allocation of transitional and horizontal categories, even when dominant balance groups remain stable.

Table 6.4: Overall agreement with the baseline trajectory typology classification (5th percentile balance threshold, 0° – 30° / 30° – 60° / 60° – 90° slope partition) for 2000–2020 under alternative percentile thresholds and slope partitions.

Slope Partition	5th Percentile	1st Percentile	10th Percentile
0° – 30° / 30° – 60° / 60° – 90°	(Baseline)	98%	97%
0° – 25° / 25° – 55° / 55° – 90°	76%	74%	73%
0° – 20° / 20° – 50° / 50° – 90°	51%	50%	49%

Table 6.5: Cohen’s kappa index with the baseline trajectory typology classification (5th percentile balance threshold, 0° – 30° / 30° – 60° / 60° – 90° slope partition) for 2000–2020 under alternative percentile thresholds and slope partitions.

Slope Partition	5th Percentile	1st Percentile	10th Percentile
0° – 30° / 30° – 60° / 60° – 90°	(Baseline)	0.94	0.93
0° – 25° / 25° – 55° / 55° – 90°	0.55	0.51	0.51
0° – 20° / 20° – 50° / 50° – 90°	0.27	0.24	0.25

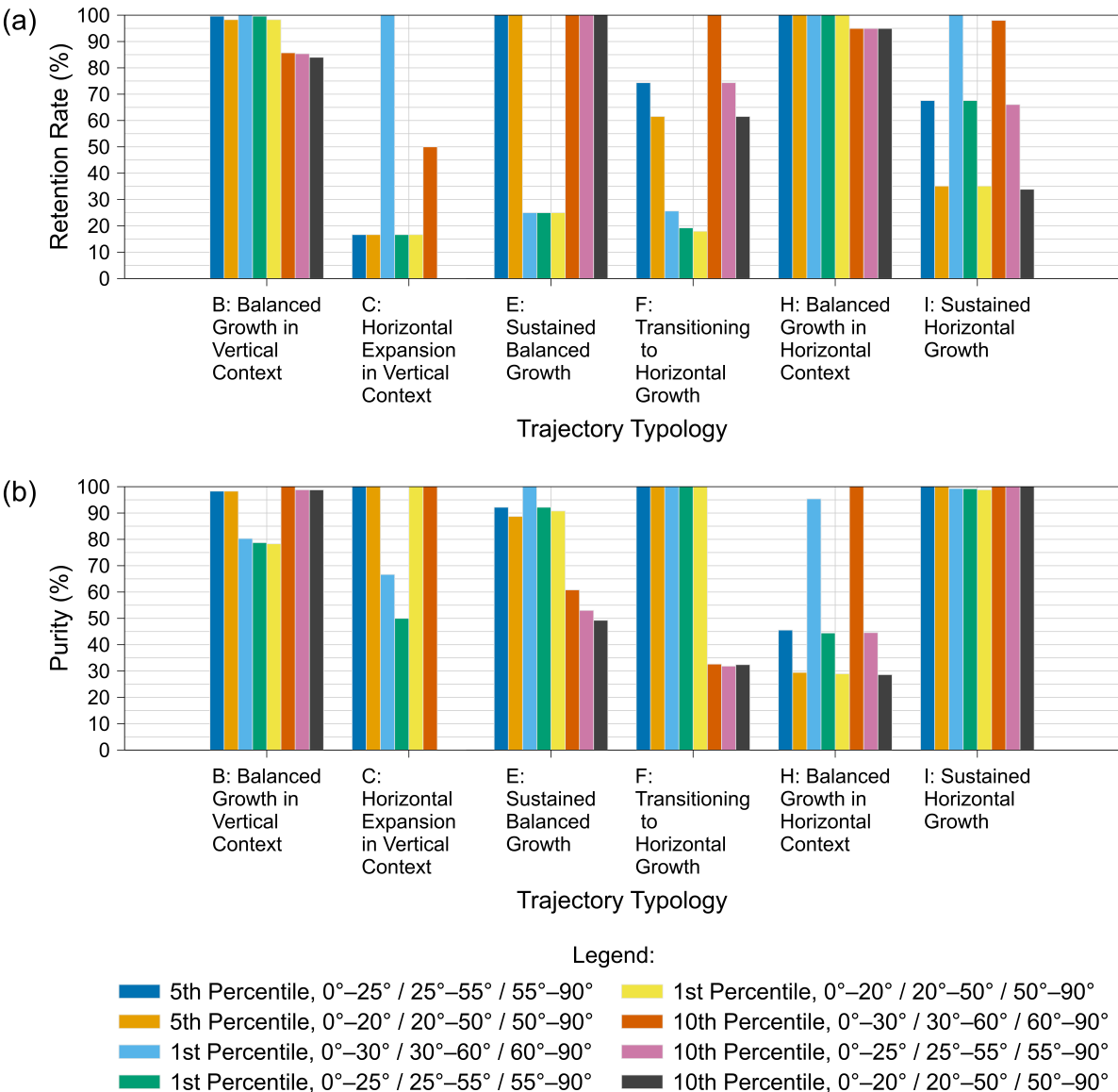


Figure 6.9: Sensitivity of trajectory typology classifications to variations in the percentile threshold for the zone of effective balance (1st, 5th, and 10th percentiles) and the slope partitioning schemes: retention (a) and assignment purity (b) for the period 2000–2020, relative to the baseline configuration (5th percentile, 0°–30°/30°–60°/60°–90°).

6.9 Discussion

6.9.1 Enhanced Interpretation of Efficiency Outcomes Through BUV Integration

The BUA–BUV trajectory framework enhances the interpretation of efficiency outcomes under SDG Indicator 11.3.1 by revealing not only whether urban growth is efficient, but also how that efficiency is associated with structural form. The conventional LCRPGR indicator, even when considered alongside its LCR and PGR components, can only show whether efficiency stems from reduced land consumption or from demographic dynamics. Because LCRPGR’s calculation relies solely on built-up area, its two-dimensional formulation captures outward expansion but cannot account for vertical

growth. The findings confirm this limitation: UCs with fundamentally different trajectories, such as those following sustained horizontal growth and those exhibiting balanced growth in a vertical context, were both classified as “efficient,” despite reflecting structurally divergent pathways.

Integrating the vertical dimension through BUV addresses this limitation. By considering BUV alongside BUA within a single trajectory framework, it becomes possible to distinguish efficiency arising from genuine densification from that arising from slower sprawl or demographic change. The added vertical dimension allows growth patterns to be traced and directly linked to efficiency outcomes, clarifying whether compactness arises from vertical expansion, constrained outward growth, or population dynamics without structural improvement. As demonstrated in the analysis, pairing the framework with LCRPGR transforms SDG Indicator 11.3.1 monitoring from a purely outcome-based assessment into a diagnostic tool that can reveal the structural sustainability, or fragility, of observed efficiency gains. Applying this framework at a global scale reveals distinct and persistent patterns in how UCs have structurally evolved over the past four decades. This explicit structural integration is a key innovation absent in earlier global SDG Indicator 11.3.1 monitoring.

6.9.2 Global Patterns of Built Form and Growth Modality

The analysis of BUA–BUV trajectories from 1980 to 2020 reveals that although most UCs experienced significant increases in both built-up extent and volume, their normalized trajectories underscore strongly contrasting structural modes of growth. A dominant pattern is the persistence of horizontally oriented built form, accompanied by a decline in balanced and vertically dominant forms over time. Typology I, *Sustained Horizontal Growth*, remained the most prevalent across both periods, underscoring the global predominance of outward expansion as the prevailing urban growth modality. In contrast, typologies associated with marked vertical expansion, A, D, and G, were rare. At the same time, transitional categories, C and F, appeared only intermittently, suggesting that shifts toward more balanced or vertical configurations remain exceptions rather than the rule.

These findings are consistent with earlier global assessments, which have consistently documented horizontal expansion as the principal mode of urban growth (e.g., Mahtta et al., 2019; Taubenböck et al., 2025; Zhou et al., 2022). The concentration of vertically dominant trajectories in Asian UCs corroborates prior studies by Liu et al. (2024) and Frolicking et al. (2024), which highlighted the regional prominence of high-rise intensification in large Asian metropolises. However, in line with those studies, the current study’s results emphasize that vertical intensification remains geographically constrained and secondary at the global scale. While compact vertical urbanism is often highlighted in sustainability debates as a desirable model for reducing land consumption and enhancing density (Bibri et al., 2020; Lin, 2018; Munn & Dragičević, 2021; Pelczynski & Tomkowicz, 2019), the empirical evidence presented here suggests that the dominant global trajectory of urban development continues to favor outward sprawl rather than upward consolidation.

6.9.3 Intersection of Structural Dynamics and Efficiency Metrics

Temporal trends in LCR, PGR, and LCRPGR from the conventional LUE assessment suggest a movement toward improved efficiency from 2000 onwards, echoing results from earlier GHSL-based

global assessments (Estoque et al., 2021; Schiavina et al., 2019, 2022b). However, the BUA–BUV trajectory framework clarifies that these improvements are not necessarily linked to vertical intensification. Although the number of sprawl-driven, inefficient UCs declined over time, horizontal expansion remained dominant across all efficiency classes. Even efficient centers were primarily associated with sustained horizontal growth rather than compact forms. Conversely, vertical and balanced typologies remained marginal, even where centers exhibited favorable LCRPGR values.

This divergence highlights that numerical efficiency does not always align with structural compactness. Improvements in LUE may stem from slower expansion or demographic stabilization, rather than genuine densification. Similarly, vertical intensification may initially appear inefficient if not matched by population growth, producing lagged efficiency gains. These findings emphasize that efficiency outcomes cannot be interpreted reliably without structural context, as LCRPGR alone risks obscuring the mechanisms driving observed values.

This concern is consistent with Taubenböck et al. (2025), who showed that global urban expansion from 1985 to 2015 was associated with structurally less-dense urban growth, as newly developed areas were dominated by lower-density morphologies and lower population densities than older urban cores. Their scenario analysis further indicated substantial unrealized potential for more compact land use, supporting the dissertation’s argument that numerically improved efficiency outcomes may still mask structurally non-compact development. These results underscore the importance of complementing efficiency metrics with structural diagnostics to avoid misinterpretations of urban sustainability.

6.9.4 Structural Sustainability and Policy Implications

The alignment or misalignment between efficiency metrics and structural growth has critical implications for urban policy. The results show that many UCs classified as “efficient” under SDG Indicator 11.3.1 continue to follow predominantly horizontal trajectories. While such centers may achieve numerical efficiency, their form entails higher long-term infrastructure costs, fragmented service delivery, and larger carbon and ecological footprints (OECD, 2018).

These insights indicate that SDG Indicator 11.3.1 monitoring should move beyond the conventional assessment and incorporate structural diagnostics such as the BUA–BUV trajectory framework. For policy, this means that urban efficiency assessments can no longer be limited to ratios of land consumption and population growth but should also evaluate how land is consumed. Embedding trajectory-based diagnostics into monitoring systems would allow policymakers to achieve the following:

- Identify unsustainable pathways early, flagging UCs where apparent efficiency masks sprawl-driven growth that may lock in high resource use (Hasse & Lathrop, 2003).
- Promote densification through measures such as transit-oriented development (Kim & Li, 2021; Yang et al., 2024) and mixed-use zoning (Mualam et al., 2019), focusing on areas where vertical growth is both practical and necessary.

- Differentiate regional strategies by promoting vertical or balanced growth in land-scarce contexts (Grace Wong, 2004), while focusing on managed expansion in regions with lower demographic pressure.
- Support transitional pathways by recognizing centers currently in mixed or evolving forms and ensuring these shifts are reinforced through infrastructure and governance interventions (Li et al., 2024; Swilling & Hajer, 2017).

These implications highlight the importance of reframing SDG Indicator 11.3.1 monitoring as both an evaluative and diagnostic tool. The BUA–BUV trajectory framework can help cities and regions align urban growth management with broader goals of resilience, inclusivity, and low-carbon development by revealing whether sustainable structural forms underpin efficiency outcomes.

6.10 Limitations and Further Work

The findings of this chapter should be interpreted in light of several limitations, particularly in relation to uncertainty considerations developed in earlier chapters. First, the use of globally normalized BUA and BUV values constrains the interpretation of the results. The global min–max transformation places all UCs on a common scale, which is necessary for consistent typological classification and global comparison. Because the normalization is defined by the smallest and largest BUA and BUV values in the dataset, the relative positions of many small- and mid-sized UCs can be compressed, especially when a few megacities set a very high upper bound. As a result, differences among these UCs may appear less pronounced than they are in absolute terms. Users of the results should therefore recognize normalization as both a methodological necessity and a trade-off, balancing comparability with the potential loss of local detail.

Additionally, classifications of prevailing built form and growth modality are contingent on operational parameter choices, including the definition of the zone of effective balance and the slope thresholds separating horizontal, balanced, and vertical growth. The sensitivity analysis demonstrated that alternative parameterizations could result in moderate to substantial reassignment of typologies. Trajectory assignments are therefore subject to structural uncertainty analogous to the data and parameter sensitivities discussed in Chapters 3 and 4. They should be interpreted as comparative rather than absolute descriptors of urban structure.

Moreover, the framework is based on BUA and BUV values aggregated at the UC level. This formulation assumes homogeneity of built form and growth dynamics within each UC, meaning that intra-urban variability, such as districts simultaneously undergoing horizontal expansion and vertical densification, cannot be represented. While this assumption is necessary for global-scale comparability, it may mask important sub-city dynamics that influence sustainability outcomes.

A related limitation concerns the input data. The analysis depends on GHSL-derived built-up area and volume data from GHS-UCDB 2025. Although these datasets represent state-of-the-art global EO products, uncertainties related to classification, height estimation, and temporal harmonization were not explicitly quantified or propagated in this chapter. Consequently, while the results capture robust global patterns, local-scale deviations from ground reality cannot be excluded. Unlike the

probabilistic LUE assessments in Chapters 4 and 5, BUA–BUV trajectories should be interpreted as deterministic structural summaries rather than uncertainty-aware representations.

Additionally, the analysis relied on a filtered version of the GHS-UCDB 2025, where UCs with extreme LCRs and PGRs were excluded. While this step was necessary to ensure methodological consistency, it also means that the assessment is incomplete and that the reported distributions do not cover the complete global set of UCs. Moreover, because normalization statistics for BUA and BUV were derived from the reduced sample, the resulting balance thresholds may not fully reflect the entire spectrum of built-up characteristics. Consequently, the classification of prevailing forms and trajectories may be biased.

Finally, the framework was applied using three benchmark years, 1980, 2000, and 2020, and BUA data extracted from 100 m GHSL layers. While appropriate for global analysis, this spatial and temporal resolution limits the detection of short-term fluctuations and fine-scale structural changes in urban growth.

Further work should extend the BUA–BUV trajectory framework by explicitly incorporating uncertainty in built-up area and volume estimates, for example, by propagating input data uncertainty into trajectory positions and typology assignments. Exploring alternative normalization schemes and parameterizations would help assess the robustness of structural classifications. Applying the framework at finer spatial and temporal scales and to additional global EO datasets would enable evaluation of its transferability and sensitivity to data sources. Structural trajectories could also be linked at the UC level to additional socioeconomic, environmental, and governance variables through multivariate, panel, or typology-based comparative analysis. For example, trajectory classes or continuous trajectory attributes such as position, angle, or distance from the 1:1 line could be modeled against variables related to income, accessibility, land regulation, infrastructure provision, environmental conditions, or institutional capacity. This would help explain why particular urban growth pathways emerge and how they relate to broader sustainability outcomes.

6.11 Chapter Conclusion

In this chapter, the BUA–BUV trajectory framework was developed to extend SDG Indicator 11.3.1-based LUE assessment beyond planar land consumption. The findings show that integrating BUA and BUV within a trajectory-based framework enhances the interpretability and robustness of the LUE assessment based on SDG Indicator 11.3.1, primarily by reducing interpretive ambiguity in LCRPGR. The results of the framework’s application to global UCs show that improvements in conventional LCRPGR patterns over time were not generally accompanied by a widespread shift toward vertical intensification. Sustained horizontal growth dominated across periods and across all LUE classes, including many centers classified as “efficient” using LCRPGR alone. The trajectory typologies, therefore, reveal structurally divergent pathways that are otherwise conflated under similar LCRPGR values, distinguishing efficiency associated with horizontal expansion, vertical densification, or more balanced forms. In this way, the framework transforms SDG Indicator 11.3.1 monitoring from a purely outcome-based assessment into a diagnostic interpretation that

clarifies how efficiency outcomes emerge and whether they are underpinned by structurally compact development or by demographic and planimetric dynamics. Overall, the framework provides a scalable way to describe structural pathways of urbanization, supporting more differentiated and policy-relevant interpretation of SDG Indicator 11.3.1.

7 Synthesis, Summary of Implications, and Future Research Directions

7.1 Synthesis of Key Findings Across Research Questions

This dissertation set out to examine whether prevailing deterministic, retrospective, and planar applications of SDG Indicator 11.3.1 are sufficient for a reliable and policy-relevant assessment of urban LUE. Taken together, the results show that they are not. Across the empirical and methodological chapters, a consistent pattern emerged: LUE outcomes are conditional on how BUA and Pop are represented, highly sensitive to uncertainty once propagated through the indicator transformations, increasingly difficult to interpret when extended into the future without explicit predictive uncertainty, and only partially explained when assessed through LCRPGR alone. Therefore, the overall conclusion is that SDG Indicator 11.3.1 remains useful, but only when treated as an uncertainty-sensitive, forward-looking, and structurally incomplete indicator that benefits from complementary analytical frameworks rather than as a precise, standalone scalar.

In relation to RQ1, the dissertation first showed that, even before formal uncertainty propagation is introduced, SDG Indicator 11.3.1 already exhibits substantial deterministic sensitivity to routine input data choices. When alternative EO-derived BUA and Pop datasets were substituted under otherwise fixed analytical conditions, LCR, PGR, and especially LCRPGR shifted systematically, with BUA representation exerting the stronger influence. In the Philippine LGU analysis, threshold crossings between efficient and inefficient classes were not rare, and a substantial share of units changed class purely because the underlying data representation changed. These findings establish an important baseline conclusion: many reported LUE outcomes depend on dataset definitions and measurement logic, rather than representing a single unambiguous state of LUE.

That baseline leads directly to the findings for RQ2 and RQ3. Once BUA and Pop are treated as uncertain rather than fixed, SDG Indicator 11.3.1 is more appropriately interpreted as a distribution than as a point value. The uncertainty propagation framework showed that LCR and PGR remain reasonably interpretable under uncertainty, but LCRPGR is structurally much more unstable due to its nonlinear ratio form and sensitivity to denominator behavior. The dissertation therefore moves the interpretation of LUE from deterministic labels to probabilistic statements. Rather than asking only which regime a city belongs to, the framework asks how strongly that regime is supported and what alternative regimes remain plausible. This change in interpretation approach directly improves transparency in borderline cases and provides a more defensible basis for comparison across cities and periods.

In relation to RQ4, the dissertation showed that forward-looking, uncertainty-aware LUE assessment is feasible at the global UC scale. The global joint-target DL workflow identified GRU as the most balanced architecture for BUA and Pop forecasting; deep ensembling improved stability and calibration; and the proposed approach outperformed the ES baseline, demonstrating clear value from pooled learning across UCs. At the same time, forecast uncertainty did not propagate uniformly: LCR remained comparatively stable, PGR was more heterogeneous, and LCRPGR again emerged as the most fragile quantity, especially where projected population growth approached zero. The projected probabilistic LUE classifications showed moderate agreement with the observed classes and, importantly, the dominant regime probability provided useful confidence information. Thus, the main conclusion is not that future LUE can be predicted precisely, but that it can be assessed probabilistically, providing more informative and decision-relevant information than deterministic projection alone.

In relation to RQ5, the BUA–BUV trajectory framework showed that area-based efficiency outcomes do not adequately explain how urban growth occurs. Many UCs that appear efficient in LCRPGR terms remain structurally dominated by horizontal growth, while similar ratio values can arise from materially different combinations of horizontal expansion and vertical intensification. By integrating BUA and BUV in a common trajectory space, the dissertation shows that SDG Indicator 11.3.1 captures an outcome but not the structural pathway that produced it. Although uncertainty was not explicitly incorporated into the BUA–BUV framework, it nevertheless adds an essential dimension that is absent from conventional SDG Indicator 11.3.1 applications. Accordingly, the most complete interpretation of LUE comes from combining probabilistic SDG Indicator 11.3.1 assessment with a trajectory-based structural reading of urban growth. In this sense, the dissertation’s overall contribution is not a replacement for SDG Indicator 11.3.1, but rather its repositioning: from a deterministic, planar indicator to a probabilistic, forecast-enabled, and multidimensional diagnostic framework for urban LUE assessment.

7.2 Summary of Implications for SDG Indicator 11.3.1 Monitoring and Broader Urban Sustainability Assessment

Building on the preceding synthesis, this section summarizes the main implications of the dissertation for the interpretation, reporting, and practical use of SDG Indicator 11.3.1. It also considers the broader relevance of the findings for urban sustainability assessment, especially where EO-derived indicators are used to inform planning and policy under uncertainty.

A first implication is methodological. SDG Indicator 11.3.1 should not be reported or interpreted as though BUA and Pop were exact quantities. The results of this dissertation show that indicator values and regime assignments can shift materially under alternative BUA and Pop representations and under plausible uncertainty ranges. For routine monitoring, this means that dataset definitions, built-up semantics, and data provenance should be documented explicitly. It also means that a single deterministic label should be treated as conditional rather than absolute. At a minimum, LCR and PGR should continue to be reported alongside LCRPGR, in line with SDG Indicator 11.3.1 methodology, and where feasible, uncertainty intervals and regime probabilities should be

reported alongside central estimates. In practical terms, this implies a shift from reporting a single “final” LUE class toward reporting the most likely regime together with the degree of confidence attached to that classification. Such reporting would better distinguish between cities with clearly supported classifications and those whose apparent status is sensitive to plausible variation in the inputs. For policymakers and decision-makers, this means that reported uncertainties should be used to judge how strongly a result can support action, comparison, or prioritization. Results with narrow intervals and clearly dominant regime probabilities may support firmer conclusions, whereas wide intervals or less concentrated regime probabilities should prompt more cautious interpretation, additional validation, or adaptive responses.

A second implication is practical. For EO-based SDG monitoring to become genuinely uncertainty-aware, validation information must be provided in a form that supports inference rather than only descriptive accuracy reporting. It should include sampling design information, stratum definitions, sample sizes, and full error or confusion matrices. Without these, uncertainty in built-up class areas cannot be propagated transparently to SDG Indicator 11.3.1 metrics. The dissertation therefore suggests that an extended reporting template for SDG Indicator 11.3.1 should combine point summaries, interval estimates, and probabilistic LUE regime membership. This extension preserves compatibility with the official regime taxonomy while avoiding overconfident interpretation. For national statistical offices, international agencies, and other reporting bodies, this suggestion would support more transparent communication of indicator reliability and make cross-city or cross-country comparisons more defensible. It would also encourage producers of EO-derived built-up datasets to publish validation outputs in forms that are analytically usable for uncertainty propagation. In this context, uncertainty information should be treated as part of the indicator evidence itself rather than as optional supplementary material, since the credibility of downstream interpretation depends directly on how well that uncertainty can be quantified and communicated.

A third implication concerns planning and prospective use. If SDG Indicator 11.3.1 is to support decision-making rather than only retrospective description, then future LUE assessment must be based on predictive distributions, not only on point forecasts. The results show that forecast uncertainty is itself informative, since wider uncertainty and weaker dominant regime probabilities identify UCs where future LUE outcomes are less stable and where planning decisions should be made more cautiously. In this sense, uncertainty-aware forecasting enables SDG Indicator 11.3.1 to serve not only as a monitoring tool but also as a risk-aware planning support framework for the period leading up to 2030 and beyond. From a planning perspective, this means that multi-class LUE regime probabilities can be interpreted as a decision-support signal. For example, if a city has a high probability of remaining in a population-intensive or relatively balanced growth regime, planners may treat existing densification-oriented strategies as broadly consistent with likely future development. By contrast, if probability mass is distributed across multiple LUE regimes, especially between more and less land-intensive outcomes, this signals greater uncertainty in the future urban trajectory and may justify more cautious or adaptive planning responses. Similarly, where projected LCRPGR outcomes appear unfavorable but with low confidence, planners should avoid treating those results as definitive warnings and instead interpret them as prompts for closer investigation, scenario testing, or targeted data improvement. In this way, predictive uncertainty is not merely a statement about model imperfection; it is also a practical indicator of planning risk.

As such, decision-makers are guided in distinguishing between futures that appear robust enough to inform near-term policy and those for which flexibility, contingency measures, or incremental implementation would be more appropriate.

Finally, the dissertation implies that SDG Indicator 11.3.1 should be complemented structurally when used to interpret urban LUE. Because the indicator captures land consumption only in two-dimensional planar terms, it cannot, by itself, distinguish whether apparent efficiency is associated with compact vertical intensification, extensive horizontal growth, or other mixed pathways. The BUA–BUV trajectory framework should therefore be applied alongside SDG Indicator 11.3.1 as a complementary framework for interpreting the structural form and direction of urban growth. For policy interpretation, this means that similar LCRPGR values should not automatically be treated as equivalent, especially where structural compactness and vertical development are important planning concerns. In practical use, the trajectory framework can help policymakers distinguish whether a seemingly similar LUE outcome arises from compact intensification, outward expansion, or a mixed development pathway. This distinction matters because the policy implications of these pathways differ substantially for infrastructure provision, transport demand, service accessibility, environmental pressure, and land preservation. A city with an apparently favorable LCRPGR but a strongly horizontal trajectory may still warrant concern regarding sprawl, infrastructure extension, or loss of open land, whereas a city with a comparable LCRPGR but a more vertical trajectory may reflect a more compact form of urban accommodation. Thus, the BUA–BUV framework adds diagnostic depth by clarifying how urban growth is being spatially and physically realized, not merely whether the aggregate ratio appears efficient. Accordingly, the framework can support more targeted policy interpretation by helping decision-makers align intervention strategies with the underlying growth pathway rather than with the aggregate indicator value alone.

Altogether, the uncertainty-aware SDG Indicator 11.3.1 framework and the BUA–BUV trajectory framework support more informed interpretation of urban LUE for policy and planning. The former helps decision-makers assess how confident they should be in a reported or projected LUE regime, while the latter helps explain the structural pathway underlying that outcome. Used jointly, they enable a shift from single-number reporting toward a richer evidence base in which cities can be evaluated not only by their most likely LUE class, but also by the uncertainty surrounding that class and by the horizontal–vertical growth pattern through which it is being produced. This is particularly valuable in contexts where planning decisions must balance land conservation, housing demand, infrastructure efficiency, and long-term urban resilience.

Beyond its direct contribution to SDG Indicator 11.3.1 monitoring, this dissertation also has broader implications for urban sustainability assessment and for the wider use of EO-based SDG indicators. It demonstrates the value of uncertainty-aware, forward-looking analysis for indicators derived from modeled geospatial inputs, particularly when policy conclusions may be sensitive to data quality, predictive uncertainty, and nonlinear transformations. In this respect, the conceptual and methodological contributions developed here are not limited to LUE assessment. They are also relevant to other SDG indicators that depend on EO-derived land, environmental, or exposure data, such as indicators related to public transport and access to green spaces (e.g., SDG 11.2 and 11.7), ecosystem extent and change (e.g., SDG 6.6 and 15.1), and hazard or exposure-sensitive

urban assessments linked to resilience and risk reduction (e.g., SDG 11.5) (United Nations Statistics Division, 2026).

7.3 Future Research Directions

Chapter-specific limitations and future work have already been discussed in the relevant chapters. Rather than repeating those points in detail, this section highlights broader research directions that emerge across the dissertation as a whole.

A first direction is the development of a more fully integrated uncertainty-aware retrospective and prospective LUE assessment framework. This dissertation has shown the value of treating BUA, Pop, and their derived indicators probabilistically, both for monitoring past LUE conditions and for interpreting future LUE trajectories based on forecasts. However, these retrospective and prospective components were implemented largely in isolation, with uncertainty-aware monitoring applied to historical assessment and forecast-based LUE interpretation developed as a separate analytical component. Future work should therefore aim to connect these components more consistently across the full analytical chain. In such a framework, historical EO-derived BUA and Pop data would first be used to establish and characterize input uncertainties, which would then be propagated to LCR, PGR, LCRPGR, and regime classification for retrospective monitoring. The same historical data and associated uncertainty information would also support forward-looking forecasting of BUA and Pop, allowing future LUE outcomes to be assessed probabilistically rather than deterministically. Such integration points toward a unified framework in which uncertainty is not treated as an isolated add-on at particular stages, but as a central property of SDG Indicator 11.3.1 monitoring across both past assessment and future projection.

A second direction is the further development of forecasting approaches for BUA and Pop. The results of this dissertation support the usefulness of uncertainty-aware deep learning for forward-looking LUE assessment, while also indicating that additional methodological advances are possible. Future research may further develop forecasting by allowing greater flexibility in model structure, incorporating richer contextual information, and explicitly representing spatial interdependence. This broader analytical scope would move beyond joint BUA–Pop trajectories at the UC level by accounting more explicitly for spatial context, linkages with other UCs, and other factors that influence urban change.

A third direction is to extend the multidimensional perspective on urban LUE introduced through the BUA–BUV trajectory framework. The dissertation has shown that incorporating the vertical dimension can reveal aspects of urban development that remain hidden in conventional area-based interpretations of SDG Indicator 11.3.1. Future work should build on this perspective by developing trajectory-based approaches that are more uncertainty-aware, spatially detailed, and explanatory. These enhancements would support a more comprehensive understanding of how different horizontal and vertical growth pathways influence urban LUE and its interpretation.

Bibliography

- Aas, K., Czado, C., Frigessi, A., & Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2), 182–198. <https://doi.org/10.1016/j.insmatheco.2007.02.001>
- Abotaleb, M., & Dutta, P. K. (2024). Optimizing gated recurrent unit networks for univariate time series forecasting: A comprehensive guide. In *Hybrid Information Systems: Non-Linear Optimization Strategies with Artificial Intelligence* (pp. 473–490). De Gruyter. <https://doi.org/10.1515/9783111331133>
- Adams, D. S., Zimmer, A., Tuccillo, J., Moehl, J., Stewart, R., Walters, S., Bhaduri, B., Li, P., & Urban, M. (2025). LandScan mosaic enables high-resolution gridded population estimates with explicit uncertainty. *Scientific Reports*, 15(1), 44493. <https://doi.org/10.1038/s41598-025-28125-z>
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- 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. <https://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. <https://doi.org/10.1007/s44196-024-00413-y>
- Al-shalabi, M., Billa, L., Pradhan, B., Mansor, S., & Al-Sharif, A. A. A. (2013). Modelling urban growth evolution and land-use changes using GIS based cellular automata and SLEUTH models: The case of Sana’a metropolitan city, Yemen. *Environmental Earth Sciences*, 70(1), 425–437. <https://doi.org/10.1007/s12665-012-2137-6>
- Angel, S., Mackres, E., & Guzder-Williams, B. (2024). Measuring change in urban land consumption: A global analysis. *Land*, 13(9), 1491. <https://doi.org/10.3390/land13091491>
- Angelopoulos, A. N., & Bates, S. (2021). A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *arXiv Preprint*. <https://doi.org/10.1561/2200000101>

- Aquilino, M., Adamo, M., Blonda, P., Barbanente, A., & Tarantino, C. (2021). Improvement of a dasymetric method for implementing sustainable development goal 11 indicators at an intra-urban scale. *Remote Sensing*, 13(14). <https://doi.org/10.3390/rs13142835>
- Arumugam, V., & Natarajan, V. (2023). Time series modeling and forecasting using autoregressive integrated moving average and seasonal autoregressive integrated moving average models. *Instrumentation Mesure Métrologie*, 22(4), 161–168. <https://doi.org/10.18280/i2m.220404>
- Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.1803.01271>
- Bai, X., McPhearson, T., Cleugh, H., Nagendra, H., Tong, X., Zhu, T., & Zhu, Y.-G. (2017). Linking urbanization and the environment: Conceptual and empirical advances. *Annual Review of Environment and Resources*, 42(1), 215–240. <https://doi.org/10.1146/annurev-environ-102016-061128>
- Baloloy, A., Sta. Ana, R. R., Cruz, J. A., Blanco, A. C., Lubrica, N. V., Valdez, C. J., & Cajucom, E. P. (2019). Spatiotemporal multi-satellite biophysical data analysis of the effect of urbanization on land surface and air temperature in Baguio City, Philippines. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W19, 47–54. <https://doi.org/10.5194/isprs-archives-XLII-4-W19-47-2019>
- 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. <https://doi.org/10.1080/02664763.2010.545119>
- Barua, M., Kumar, T., Raj, K., & Roy, A. M. (2024). Comparative analysis of deep learning models for stock price prediction in the Indian market. *FinTech*, 3(4), 551–568. <https://doi.org/10.3390/fintech3040029>
- Bedford, T., & Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32(1–4), 245–268. <https://doi.org/10.1023/A:1016725902970>
- Bedford, T., & Cooke, R. M. (2002). Vines—A new graphical model for dependent random variables. *The Annals of Statistics*, 30(4). <https://doi.org/10.1214/aos/1031689016>
- 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. <https://doi.org/10.1145/3533382>
- Berse, K. (2024). Regional: Advancing Inclusive and Resilient Urban Development Targeted at the Urban Poor Philippines: Urbanization Trends, Hazard Exposure and Sectoral Realities: A Macro View of Philippine Cities (Supplementary Report). In *ADB Technical Assistance Consultant's*

- Report, Project Number: 51325-001*. Asian Development Bank. Retrieved March 17, 2024, from https://www.adb.org/sites/default/files/project-documents/51325/51325-001-tacr-en_5.pdf
- Bhandari, R., Xue, W., Virdis, S. G. P., Winijkul, E., Nguyen, T. P. L., & Joshi, S. (2023). Monitoring and assessing urbanization progress in Thailand between 2000 and 2020 using SDG indicator 11.3.1. *Sustainability (Switzerland)*, 15(12). <https://doi.org/10.3390/su15129794>
- Bibri, S. E., Krogstie, J., & Kärrholm, M. (2020). Compact city planning and development: Emerging practices and strategies for achieving the goals of sustainability. *Developments in the Built Environment*, 4, 100021. <https://doi.org/10.1016/j.dibe.2020.100021>
- Bounoua, L., Bachir, N., Souidi, H., Bahi, H., Lagmiri, S., Khebiza, M. Y., Nigro, J., & Thome, K. (2023). Sustainable development in Algeria's urban areas: Population growth and land consumption. *Urban Science*, 7(1). <https://doi.org/10.3390/urbansci7010029>
- Bruyns, G. J., Higgins, C. D., & Nel, D. H. (2021). Urban volumetrics: From vertical to volumetric urbanisation and its extensions to empirical morphological analysis. *Urban Studies*, 58(5), 922–940. <https://doi.org/10.1177/0042098020936970>
- Budsaba, K., Tipwong, N., Manomat, S., & Phaphan, W. (2025). Total population forecasting in Bangkok: A comparative study of time series and machine learning models. *Lobachevskii Journal of Mathematics*, 46(4), 1552–1566. <https://doi.org/10.1134/S199508022560640X>
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2), 261–304. <https://doi.org/10.1177/0049124104268644>
- Cacciari, I., & Ranfagni, A. (2024). Hands-on fundamentals of 1D convolutional neural networks—A tutorial for beginner users. *Applied Sciences*, 14(18), 8500. <https://doi.org/10.3390/app14188500>
- 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). <https://doi.org/10.1371/journal.pone.0244318>
- Calka, B., Orych, A., Bielecka, E., & Mozuriunaite, S. (2022). The ratio of the land consumption rate to the population growth rate: A framework for the achievement of the spatiotemporal pattern in Poland and Lithuania. *Remote Sensing*, 14(5). <https://doi.org/10.3390/rs14051074>
- Camagni, R., Gibelli, M. C., & Rigamonti, P. (2002). Urban mobility and urban form: The social and environmental costs of different patterns of urban expansion. *Ecological Economics*, 40(2), 199–216. [https://doi.org/10.1016/S0921-8009\(01\)00254-3](https://doi.org/10.1016/S0921-8009(01)00254-3)
- Cardenas-Ritzert, O. S. E., Vogeler, J. C., Shah Heydari, S., Fekety, P. A., Laituri, M., & McHale, M. (2024a). 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). <https://doi.org/10.3390/ijgi13070226>

- Cardenas-Ritzert, O. S. E., Vogeler, J. C., Shah Heydari, S., Fekety, P. A., Laituri, M., & McHale, M. R. (2024b). Effects of land use data spatial resolution on SDG indicator 11.3.1 (urban expansion) assessments: A case study across Ethiopia. *Sustainability*, 16(22), 9698. <https://doi.org/10.3390/su16229698>
- CEOS. (2023). *SDG Indicator 11.3.1 – EO Support Sheet*. Committee on Earth Observation Satellites (CEOS). Retrieved January 21, 2026, from https://ceos.org/document_management/Ad_Hoc_Teams/SDG-AHT/Support%20Sheets/SDG_11.3.1_EO_Support_Sheet_v2023_1.pdf
- 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. <https://doi.org/10.1038/s41467-024-52241-5>
- Che, Z., Purushotham, S., Cho, K., Sontag, D., & Liu, Y. (2018). Recurrent neural networks for multivariate time series with missing values. *Scientific Reports*, 8(1), 6085. <https://doi.org/10.1038/s41598-018-24271-9>
- 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. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1724–1734. <https://doi.org/10.3115/v1/D14-1179>
- Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.1412.3555>
- Cimini, A., De Fioravante, P., Riitano, N., Dichicco, P., Calò, A., Scarascia Mugnozza, G., Marchetti, M., & Munafò, M. (2023). Land consumption dynamics and urban–rural continuum mapping in Italy for SDG 11.3.1 indicator assessment. *Land*, 12(1). <https://doi.org/10.3390/land12010155>
- Clark, D. (1998). Interdependent urbanization in an urban worlds: An historical overview. *The Geographical Journal*, 164(1), 85–95. <https://doi.org/10.2307/3060547>
- Cochran, W. G. (1977). *Sampling Techniques*. John Wiley & Sons.
- Cohen, J. (1960). A Coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1), 35–46. [https://doi.org/10.1016/0034-4257\(91\)90048-B](https://doi.org/10.1016/0034-4257(91)90048-B)
- Dalagnol, R., Wagner, F. H., Galvão, L. S., Braga, D., Osborn, F., Sagang, L. B., Da Conceição Bispo, P., Payne, M., Silva Junior, C., Favrichon, S., Silgueiro, V., Anderson, L. O., Aragão, L. E. O. E. C. D., Fensholt, R., Brandt, M., Ciais, P., & Saatchi, S. (2023). Mapping tropical

- forest degradation with deep learning and Planet NICFI data. *Remote Sensing of Environment*, 298, 113798. <https://doi.org/10.1016/j.rse.2023.113798>
- De Keersmaeker, L., Roggemans, P., Poelmans, L., Priem, F., Taillir, S., Petermans, T., & Van Valckenborgh, J. (2026). Homogenization of Northern Belgian landscapes through centuries of reclamation, agricultural transition, and urbanization. *Nature Communications*, 17(1), 1906. <https://doi.org/10.1038/s41467-026-68594-y>
- Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research*, 7, 1–30.
- Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, 14(2), 241–258. <https://doi.org/10.1007/s11704-019-8208-z>
- Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1(1), 54–75.
- Estoque, R. C., & Murayama, Y. (2011). Spatio-temporal urban land use/cover change analysis in a hill station: The case of Baguio City, Philippines. *Procedia - Social and Behavioral Sciences*, 21, 326–335. <https://doi.org/10.1016/j.sbspro.2011.07.016>
- Estoque, R. C., & Murayama, Y. (2012). Examining the potential impact of land use/cover changes on the ecosystem services of Baguio City, the Philippines: A scenario-based analysis. *Applied Geography*, 35(1–2), 316–326. <https://doi.org/10.1016/j.apgeog.2012.08.006>
- Estoque, R. C., & Murayama, Y. (2013). City profile: Baguio. *Cities*, 30, 240–251. <https://doi.org/10.1016/j.cities.2011.05.002>
- 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. <https://doi.org/10.1016/j.habitatint.2021.102403>
- European Commission. (2023). *GHSL Data Package 2023*. Publications Office of the European Union. <https://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*. Publications Office of the European Union. Retrieved April 20, 2025, from <https://data.europa.eu/doi/10.2760/3046391>
- Fabian Jr., V. I. (1991). Land Use and Land Classification of the Philippines. *Infomapper*, 1(2), 10–11. Retrieved May 24, 2023, from https://www.namria.gov.ph/jdownloads/Info_Mapper/00a_im_dec911.pdf

- FAO. (2016). *Map Accuracy Assessment and Area Estimation: A Practical Guide*. Food and Agriculture Organization of the United Nations. Retrieved February 26, 2026, from <https://openknowledge.fao.org/items/df0886a8-7aa0-4f46-a7c3-0c87591ad519>
- Faye, B., Du, G., & Zhang, R. (2022). Efficiency analysis of land use and the degree of coupling link between population growth and global built-up area in the subregion of West Africa. *Land*, 11(6). <https://doi.org/10.3390/land11060847>
- Ferbar Tratar, L., Mojškerc, B., & Toman, A. (2016). Demand forecasting with four-parameter exponential smoothing. *International Journal of Production Economics*, 181, 162–173. <https://doi.org/10.1016/j.ijpe.2016.08.004>
- Forman, R. T. T. (1995). *Land mosaics: The ecology of landscapes and regions*. Cambridge University Press.
- Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32(200), 675–701.
- Frolking, S., Mahtta, R., Milliman, T., Esch, T., & Seto, K. C. (2024). Global urban structural growth shows a profound shift from spreading out to building up. *Nature Cities*, 1(9), 555–566. <https://doi.org/10.1038/s44284-024-00100-1>
- Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. *Proceedings of The 33rd International Conference on Machine Learning, PMLR*, 48, 1050–1059.
- Gallaun, H., Steinegger, M., Wack, R., Schardt, M., Kornberger, B., & Schmitt, U. (2015). Remote sensing based two-stage sampling for accuracy assessment and area estimation of land cover changes. *Remote Sensing*, 7(9), 11992–12008. <https://doi.org/10.3390/rs70911992>
- Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, 115, 105151. <https://doi.org/10.1016/j.engappai.2022.105151>
- Gao, J., & O'Neill, B. C. (2020). Mapping global urban land for the 21st century with data-driven simulations and shared socioeconomic pathways. *Nature Communications*, 11(1), 2302. <https://doi.org/10.1038/s41467-020-15788-7>
- Gao, K., Yang, X., Wang, Z., Zhang, H., Huang, C., & Zeng, X. (2022). Spatial sustainable development assessment using fusing multisource data from the perspective of production-living-ecological space division: A case of Greater Bay Area, China. *Remote Sensing*, 14(12). <https://doi.org/10.3390/rs14122772>
- Gardner, E. S. (2006). Exponential smoothing: The state of the art—Part II. *International Journal of Forecasting*, 22(4), 637–666. <https://doi.org/10.1016/j.ijforecast.2006.03.005>

- 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. <https://doi.org/10.1109/JURSE57346.2023.10144145>
- Genest, C., Ghoudi, K., & Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3), 543–552. <https://doi.org/10.1093/biomet/82.3.543>
- Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation*, 12(10), 2451–2471. <https://doi.org/10.1162/089976600300015015>
- Ghazaryan, G., Rienow, A., Oldenburg, C., Thonfeld, F., Trampnau, B., Stickse, 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). <https://doi.org/10.3390/rs13091694>
- Gómez, J. A., Patiño, J. E., Duque, J. C., & Passos, S. (2019). Spatiotemporal modeling of urban growth using machine learning. *Remote Sensing*, 12(1), 109. <https://doi.org/10.3390/rs12010109>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Grace Wong, K. M. (2004). Vertical cities as a solution for land scarcity: The tallest public housing development in Singapore. *URBAN DESIGN International*, 9(1), 17–30. <https://doi.org/10.1057/palgrave.udi.9000108>
- Grossman, I., Wilson, T., & Temple, J. (2023). Forecasting small area populations with long short-term memory networks. *Socio-Economic Planning Sciences*, 88, 101658. <https://doi.org/10.1016/j.seps.2023.101658>
- Hall, T., & Rasheed, K. (2025). A survey of machine learning methods for time series prediction. *Applied Sciences*, 15(11), 5957. <https://doi.org/10.3390/app15115957>
- Han, L., Lu, L., Lu, J., Liu, X., Zhang, S., Luo, K., He, D., Wang, P., Guo, H., & Li, Q. (2022). Assessing spatiotemporal changes of SDG indicators at the neighborhood level in Guilin, China: A geospatial big data approach. *Remote Sensing*, 14(19). <https://doi.org/10.3390/rs14194985>
- Hansen, M. C., Potapov, P. V., Pickens, A. H., Tyukavina, A., Hernandez-Serna, A., Zalles, V., Turubanova, S., Kommareddy, I., Stehman, S. V., Song, X.-P., & Kommareddy, A. (2022). Global land use extent and dispersion within natural land cover using Landsat data. *Environmental Research Letters*, 17(3), 034050. <https://doi.org/10.1088/1748-9326/ac46ec>
- Hasse, J. E., & Lathrop, R. G. (2003). Land resource impact indicators of urban sprawl. *Applied Geography*, 23(2–3), 159–175. <https://doi.org/10.1016/j.apgeog.2003.08.002>

- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York. <https://doi.org/10.1007/978-0-387-84858-7>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Holobacă, I.-H., Benedek, J., Ursu, C.-D., Alexe, M., & Temerdek-Ivan, K. (2022). Ratio of land consumption rate to population growth rate in the major metropolitan areas of Romania. *Remote Sensing*, 14(23). <https://doi.org/10.3390/rs14236016>
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65-70.
- Hu, Z., & Lo, C. P. (2007). Modeling urban growth in Atlanta using logistic regression. *Computers, Environment and Urban Systems*, 31(6), 667–688. <https://doi.org/10.1016/j.compenvurbsys.2006.11.001>
- Huang, C., Shen, Y., Chen, Y., & Chen, H. (2021). A novel hybrid deep neural network model for short-term electricity price forecasting. *International Journal of Energy Research*, 45(2), 2511–2532. <https://doi.org/10.1002/er.5945>
- Huang, F., Xie, G., & Xiao, R. (2009). Research on ensemble learning. *2009 International Conference on Artificial Intelligence and Computational Intelligence*, 249–252. <https://doi.org/10.1109/AICI.2009.235>
- 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. <https://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. <https://doi.org/10.1016/j.jag.2022.102787>
- Huang, Y., Chen, S., Chen, Z., Huang, Q., & Hu, W. (2020). Probabilistic load flow computation considering dependence of wind powers and using quasi-Monte Carlo method with truncated regular vine copula. *International Transactions on Electrical Energy Systems*, 30(12). <https://doi.org/10.1002/2050-7038.12646>
- Huber, P. J., & Ronchetti, E. M. (2009). *Robust Statistics* (Second Edition). John Wiley & Sons.
- Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2), 297–307. <https://doi.org/10.1093/biomet/76.2.297>
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (3rd Edition). OTexts.

- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688. <https://doi.org/10.1016/j.ijforecast.2006.03.001>
- Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2005). Prediction intervals for exponential smoothing using two new classes of state space models. *Journal of Forecasting*, 24(1), 17–37. <https://doi.org/10.1002/for.938>
- Jabareen, Y. R. (2006). Sustainable urban forms. *Journal of Planning Education and Research*, 26(1), 38–52. <https://doi.org/10.1177/0739456X05285119>
- Jaber, S. M. (2025). Land use efficiency and governance disparities: Unveiling the nexus in the Arab world. *Environmental Development*, 55, 101169. <https://doi.org/10.1016/j.envdev.2025.101169>
- Jalilov, S. M., Chen, Y., Quang, N. H., Nguyen, M. N., Leighton, B., Paget, M., & Lazarow, N. (2021). Estimation of urban land-use efficiency for sustainable development by integrating over 30-year Landsat imagery with population data: A case study of Ha Long, Vietnam. *Sustainability (Switzerland)*, 13(16). <https://doi.org/10.3390/su13168848>
- Janssen, H. (2013). Monte-Carlo based uncertainty analysis: Sampling efficiency and sampling convergence. *Reliability Engineering & System Safety*, 109, 123–132. <https://doi.org/10.1016/j.ress.2012.08.003>
- JCGM. (2008). *Evaluation of measurement data—Supplement 1 to the “Guide to the expression of uncertainty in measurement”—Propagation of distributions using a Monte Carlo method* (No. JCGM 101:2008). Joint Committee for Guides in Metrology (JCGM). https://www.bipm.org/documents/20126/2071204/JCGM_101_2008_E.pdf
- Jensen, V., Bianchi, F. M., & Anfinssen, S. N. (2024). Ensemble conformalized quantile regression for probabilistic time series forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, 35(7), 9014–9025. <https://doi.org/10.1109/TNNLS.2022.3217694>
- Jiang, C., Zhang, W., Han, X., Ni, B. Y., & Song, L. J. (2015). A vine-copula-based reliability analysis method for structures with multidimensional correlation. *Journal of Mechanical Design*, 137(6), 061405. <https://doi.org/10.1115/1.4030179>
- Jiang, H., Guo, H., Sun, Z., Xing, Q., Zhang, H., Ma, Y., & Li, S. (2022). Projections of urban built-up area expansion and urbanization sustainability in China’s cities through 2030. *Journal of Cleaner Production*, 367. <https://doi.org/10.1016/j.jclepro.2022.133086>
- 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. <https://doi.org/10.1038/s42949-021-00032-y>
- Jiang, X., Na, J., Lu, W., & Zhang, Y. (2017). Coupled Monte Carlo simulation and copula theory for uncertainty analysis of multiphase flow simulation models. *Environmental Science and Pollution Research*, 24(31), 24284–24296. <https://doi.org/10.1007/s11356-017-0030-2>

- Jin, G., Liang, Y., Fang, Y., Shao, Z., Huang, J., Zhang, J., & Zheng, Y. (2024a). Spatio-temporal graph neural networks for predictive learning in urban computing: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 36(10), 5388–5408. <https://doi.org/10.1109/TKDE.2023.3333824>
- Jin, M., Koh, H. Y., Wen, Q., Zambon, D., Alippi, C., Webb, G. I., King, I., & Pan, S. (2024b). 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. <https://doi.org/10.1109/TPAMI.2024.3443141>
- Joe, H. (2014). *Dependence Modeling with Copulas*. CRC Press.
- Jokar Arsanjani, J., Helbich, M., & De Noronha Vaz, E. (2013). Spatiotemporal simulation of urban growth patterns using agent-based modeling: The case of Tehran. *Cities*, 32, 33–42. <https://doi.org/10.1016/j.cities.2013.01.005>
- Karthikeya, N. V. B. S. S., Tripathi, N. K., Mozumder, C., Pal, I., & Pramanik, M. (2024). Impact of land transition around eastern economic corridor in Thailand in the context of SDG 11.3.1 using urban heat islands, nighttime light intensity and machine learning. *Environmental and Sustainability Indicators*, 24, 100499. <https://doi.org/10.1016/j.indic.2024.100499>
- Kerby, D. S. (2014). The simple difference formula: An approach to teaching nonparametric correlation. *Comprehensive Psychology*, 3, 11.IT.3.1.
- Khan, F., Abbass, K., & Qun, W. (2025). Charting the path to sustainable development: Insights from a review of land use efficiency research and total factor productivity in achieving SDGs. *Journal of the Knowledge Economy*, 16(5), 16187–16210. <https://doi.org/10.1007/s13132-024-02358-1>
- Khatun, A., Nisha, M. N., Chatterjee, S., & Sridhar, V. (2024). A novel insight on input variable and time lag selection in daily streamflow forecasting using deep learning models. *Environmental Modelling & Software*, 179, 106126. <https://doi.org/10.1016/j.envsoft.2024.106126>
- Khosravi, A., Nahavandi, S., Creighton, D., & Atiya, A. F. (2011). Lower upper bound estimation method for construction of neural network-based prediction intervals. *IEEE Transactions on Neural Networks*, 22(3), 337–346. <https://doi.org/10.1109/TNN.2010.2096824>
- Kim, J. H., & Li, X. (2021). Building more housing near transit: A spatial analysis of residential densification dynamics. *Transport Policy*, 114, 15–24. <https://doi.org/10.1016/j.tranpol.2021.08.018>
- Kim, J., & Kim, H. (2024). 3-D land use index: Vertical land use analysis based on multi-dimensional geospatial data cube. *2024 IEEE International Conference on Big Data (BigData)*, 8719–8721. <https://doi.org/10.1109/BigData62323.2024.10825536>

- Kim, J., Kim, H., Kim, H., Lee, D., & Yoon, S. (2025). A comprehensive survey of deep learning for time series forecasting: Architectural diversity and open challenges. *Artificial Intelligence Review*, 58(7), 216. <https://doi.org/10.1007/s10462-025-11223-9>
- Kim, Y., Safikhani, A., & Tepe, E. (2022). Machine learning application to spatio-temporal modeling of urban growth. *Computers, Environment and Urban Systems*, 94, 101801. <https://doi.org/10.1016/j.compenvurbsys.2022.101801>
- Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations (ICLR)*.
- Koomen, E., Rietveld, P., & Bacao, F. (2009). The third dimension in urban geography: The urban-volume approach. *Environment and Planning B: Planning and Design*, 36(6), 1008–1025. <https://doi.org/10.1068/b34100>
- Koroso, N. H., Lengoiboni, M., & Zevenbergen, J. A. (2021). Urbanization and urban land use efficiency: Evidence from regional and Addis Ababa satellite cities, Ethiopia. *Habitat International*, 117. <https://doi.org/10.1016/j.habitatint.2021.102437>
- Krichen, M., & Mihoub, A. (2025). Long short-term memory networks: A comprehensive survey. *AI*, 6(9), 215. <https://doi.org/10.3390/ai6090215>
- Kummaraka, U., & Srisuradetchai, P. (2025). Monte Carlo dropout neural networks for forecasting sinusoidal time series: Performance evaluation and uncertainty quantification. *Applied Sciences*, 15(8), 4363. <https://doi.org/10.3390/app15084363>
- Laituri, M., Davis, D., Sternlieb, F., & Galvin, K. (2021). SDG indicator 11.3.1 and secondary cities: An analysis and assessment. *ISPRS International Journal of Geo-Information*, 10(11). <https://doi.org/10.3390/ijgi10110713>
- 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)*.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159. <https://doi.org/10.2307/2529310>
- Lara-Benítez, P., Carranza-García, M., & Riquelme, J. C. (2021). An experimental review on deep learning architectures for time series forecasting. *International Journal of Neural Systems*, 31(03), 2130001. <https://doi.org/10.1142/S0129065721300011>
- Leyk, S., Gaughan, A. E., Adamo, S. B., de Sherbinin, A., Balk, D., Freire, S., Rose, A., Stevens, F. R., Blankespoor, B., Frye, C., Comenetz, J., Sorichetta, A., MacManus, K., Pistolesi, L., Levy, M., Tatem, A. J., & Pesaresi, M. (2019). The spatial allocation of population: A review of large-scale gridded population data products and their fitness for use. *Earth System Science Data*, 11(3), 1385–1409. <https://doi.org/10.5194/essd-11-1385-2019>

- Li, C., Cai, G., & Du, M. (2021a). 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). <https://doi.org/10.3390/ijgi10020064>
- Li, C., Cai, G., & Sun, Z. (2021b). Urban land-use efficiency analysis by integrating LCRPGR and additional indicators. *Sustainability (Switzerland)*, 13(24). <https://doi.org/10.3390/su132413518>
- 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, Q., & Xu, Y. (2019). VS-GRU: A variable sensitive gated recurrent neural network for multivariate time series with massive missing values. *Applied Sciences*, 9(15), 3041. <https://doi.org/10.3390/app9153041>
- 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. <https://doi.org/10.3390/rs12091391>
- Li, W., & Law, K. L. E. (2024). Deep learning models for time series forecasting: A review. *IEEE Access*, 12, 92306–92327. <https://doi.org/10.1109/ACCESS.2024.3422528>
- Li, X., & Liu, X. (2007). Defining agents' behaviors to simulate complex residential development using multicriteria evaluation. *Journal of Environmental Management*, 85(4), 1063–1075. <https://doi.org/10.1016/j.jenvman.2006.11.006>
- 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 socioeconomic pathways. *Earth's Future*, 7(4), 351–362. <https://doi.org/10.1029/2019EF001152>
- Li, Y., Beeton, R. J. S., Zhao, X., Fan, Y., Yang, Q., Li, J., & Ding, L. (2024). Advancing urban sustainability transitions: A framework for understanding urban complexity and enhancing integrative transformations. *Humanities and Social Sciences Communications*, 11(1), 1064. <https://doi.org/10.1057/s41599-024-03598-x>
- Licuanan, W. Y., Cabreira, R. W., & Aliño, P. M. (2019). The Philippines. In *World Seas: An Environmental Evaluation* (pp. 515–537). Elsevier. <https://doi.org/10.1016/B978-0-08-100853-9.00051-8>
- Lim, B., Arik, 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. <https://doi.org/10.1016/j.ijforecast.2021.03.012>
- Lin, Z. (2018). Vertical urbanism: Re-conceptualizing the compact city. In *Vertical Urbanism: Designing Compact Cities in China*. <https://doi.org/10.4324/9781351206839>

- Ling, Z., Jiang, W., Lu, Y., Ling, Y., Zhang, Z., & Liao, C. (2023). Continuous long time series monitoring of urban construction land in supporting the SDG 11.3.1—A case study of Nanning, Guangxi, China. *Land*, 12(2). <https://doi.org/10.3390/land12020452>
- Liu, J., Jin, X., Xu, W., Gu, Z., Yang, X., Ren, J., Fan, Y., & Zhou, Y. (2020a). A new framework of land use efficiency for the coordination among food, economy and ecology in regional development. *Science of The Total Environment*, 710, 135670. <https://doi.org/10.1016/j.scitotenv.2019.135670>
- Liu, S., Yan, Y., & Hu, B. (2023). Satellite monitoring of the urban expansion in the Pearl River–Xijiang Economic Belt and the progress towards SDG11.3.1. *Remote Sensing*, 15(21), 5209. <https://doi.org/10.3390/rs15215209>
- 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. (2020b). High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015. *Nature Sustainability*, 3(7), 564–570. <https://doi.org/10.1038/s41893-020-0521-x>
- Liu, X., Lu, L., Guo, H., Melchiorri, M., Hafner, S., Ban, Y., Pesaresi, M., Li, Q., & Baqa, M. F. (2025). Assessing urban sustainability in the Belt and Road region: A city-level analysis of SDG 11 indicators using earth observations. *International Journal of Digital Earth*, 18(1), 2522395. <https://doi.org/10.1080/17538947.2025.2522395>
- Liu, X., Wu, X., Li, X., Xu, X., Liao, W., Jiao, L., Zeng, Z., Chen, G., & Li, X. (2024). Global mapping of three-dimensional urban structures reveals escalating utilization in the vertical dimension and pronounced building space inequality. *Engineering*. <https://doi.org/10.1016/j.eng.2024.01.025>
- Liu, Y., Hu, Y., Long, S., Liu, L., & Liu, X. (2017). Analysis of the effectiveness of urban land-use-change models based on the measurement of spatio-temporal, dynamic urban growth: A cellular automata case study. *Sustainability*, 9(5), 796. <https://doi.org/10.3390/su9050796>
- 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. <https://doi.org/10.1016/j.ocecoaman.2022.106201>
- Luo, M., & Chen, Y. (2024). Simulating inter-city population flows based on graph neural networks. *Geocarto International*, 39(1), 2331223. <https://doi.org/10.1080/10106049.2024.2331223>
- Ma, J., Tian, L., Zhang, Y., Yang, X., Li, Y., Liu, Z., Zhou, L., Wang, Z., & Ouyang, W. (2024). Global property rights and land use efficiency. *Nature Communications*, 15(1), 8525. <https://doi.org/10.1038/s41467-024-52859-5>
- Macaspac, K. L., Valenzuela, M. L., & Ramos, R. V. (2024). Land use land cover (LULC) change analysis in Baguio City using GIS-based transition potential modeling. *The International*

- Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-4/W8-2023*, 357–363. <https://doi.org/10.5194/isprs-archives-XLVIII-4-W8-2023-357-2024>
- Mahmoudzadeh, H., Abedini, A., & Aram, F. (2022). Urban growth modeling and land-use/land-cover change analysis in a metropolitan area (Case study: Tabriz). *Land*, 11(12), 2162. <https://doi.org/10.3390/land11122162>
- Mahtta, R., Mahendra, A., & Seto, K. C. (2019). Building up or spreading out? Typologies of urban growth across 478 cities of 1 million+. *Environmental Research Letters*, 14(12), 124077. <https://doi.org/10.1088/1748-9326/ab59bf>
- Marconcini, M., Metz-Marconcini, A., Esch, T., & Gorelick, N. (2021). Understanding current trends in global urbanisation—The world settlement footprint suite. *GI_Forum*, 9(1). https://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. <https://doi.org/10.1038/s41597-020-00580-5>
- Mateus, B. C., Mendes, M., Farinha, J. T., Assis, R., & Cardoso, A. M. (2021). Comparing LSTM and GRU models to predict the condition of a pulp paper press. *Energies*, 14(21), 6958. <https://doi.org/10.3390/en14216958>
- McHugh, M. L. (2013). The Chi-square test of independence. *Biochemia Medica*, 143–149. <https://doi.org/10.11613/BM.2013.018>
- Melchiorri, M., Florczyk, A., Freire, S., Schiavina, M., Pesaresi, M., & Kemper, T. (2018). Unveiling 25 years of planetary urbanization with remote sensing: Perspectives from the global human settlement layer. *Remote Sensing*, 10(5), 768. <https://doi.org/10.3390/rs10050768>
- Melchiorri, M., Pesaresi, M., Florczyk, A. J., Corbane, C., & Kemper, T. (2019). Principles and applications of the global human settlement layer as baseline for the land use efficiency indicator—SDG 11.3.1. *ISPRS International Journal of Geo-Information*, 8(2). <https://doi.org/10.3390/ijgi8020096>
- Mendis, K., Wickramasinghe, M., & Marasinghe, P. (2024). Multivariate time series forecasting: A review. *Proceedings of the 2024 2nd Asia Conference on Computer Vision, Image Processing and Pattern Recognition*, 1–9. <https://doi.org/10.1145/3663976.3664241>
- Metropolis, N., & Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association*, 44(247), 335–341. <https://doi.org/10.1080/01621459.1949.10483310>
- Montero-Manso, P., & Hyndman, R. J. (2021). Principles and algorithms for forecasting groups of time series: Locality and globality. *International Journal of Forecasting*, 37(4), 1632–1653. <https://doi.org/10.1016/j.ijforecast.2021.03.004>

- Mualam, N., Salinger, E., & Max, D. (2019). Increasing the urban mix through vertical allocations: Public floorspace in mixed use development. *Cities*, 87, 131–141. <https://doi.org/10.1016/j.cities.2018.12.027>
- Mudau, N., Mwaniki, D., Tsoeleng, L., Mashalane, M., Beguy, D., & Ndugwa, R. (2020). Assessment of SDG indicator 11.3.1 and urban growth trends of major and small cities in South Africa. *Sustainability (Switzerland)*, 12(17). <https://doi.org/10.3390/su12177063>
- Munn, K., & Dragičević, S. (2021). Spatial multi-criteria evaluation in 3D context: Suitability analysis of urban vertical development. *Cartography and Geographic Information Science*, 48(2), 105–123. <https://doi.org/10.1080/15230406.2020.1845981>
- Nagle, N. N., Battenfield, B. P., Leyk, S., & Spielman, S. (2014). Dasymeric modeling and uncertainty. *Annals of the Association of American Geographers*, 104(1), 80–95. <https://doi.org/10.1080/00045608.2013.843439>
- Nagler, T. (2025). Simplified vine copula models: State of science and affairs. *Risk Sciences*, 1, 100022. <https://doi.org/10.1016/j.risk.2025.100022>
- Nelsen, R. B. (2006). *An Introduction to Copulas* (Second Edition). Springer Science+Business Media, Inc.
- Nicolau, R., David, J., Caetano, M., & Pereira, J. M. C. (2019). Ratio of land consumption rate to population growth rate-analysis of different formulations applied to mainland Portugal. *ISPRS International Journal of Geo-Information*, 8(1). <https://doi.org/10.3390/ijgi8010010>
- OECD. (2018). *Rethinking Urban Sprawl: Moving Towards Sustainable Cities*. OECD. <https://doi.org/10.1787/9789264189881-en>
- Olofsson, P. (2021). Updates to good practices for estimating area and assessing accuracy of land cover and land cover change products. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 1982–1985. <https://doi.org/10.1109/IGARSS47720.2021.9554475>
- 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. <https://doi.org/10.1016/j.rse.2014.02.015>
- 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. <https://doi.org/10.1016/j.rse.2012.10.031>
- O'Neill, B. C., Kriegler, E., Ebi, K. L., Kemp-Benedict, E., Riahi, K., Rothman, D. S., Van Ruijven, B. J., Van Vuuren, D. P., Birkmann, J., Kok, K., Levy, M., & Solecki, W. (2017). The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. *Global Environmental Change*, 42, 169–180. <https://doi.org/10.1016/j.gloenvcha.2015.01.004>

- Ostertagová, E., & Ostertag, O. (2012). Forecasting using simple exponential smoothing method. *Acta Electrotechnica et Informatica*, 12(3). <https://doi.org/10.2478/v10198-012-0034-2>
- 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.
- Pelczynski, J., & Tomkowicz, B. (2019). Densification of cities as a method of sustainable development. *IOP Conference Series: Earth and Environmental Science*, 362(1), 012106. <https://doi.org/10.1088/1755-1315/362/1/012106>
- Pesaresi, M., Huadong, G., Blaes, X., Ehrlich, D., Ferri, S., Gueguen, L., Halkia, M., Kauffmann, M., Kemper, T., Lu, L., Marin-Herrera, M. A., Ouzounis, G. K., Scavazzon, M., Soille, P., Syrris, V., & Zanchetta, L. (2013). A global human settlement layer from optical HR/VHR RS data: Concept and first results. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(5), 2102–2131. <https://doi.org/10.1109/JSTARS.2013.2271445>
- Pesaresi, M., & Politis, P. (2023). *GHS-BUILT-H R2023A - GHS building height, derived from AW3D30, SRTM30, and Sentinel2 composite (2018)*. European Commission, Joint Research Centre (JRC). <https://doi.org/10.2905/85005901-3A49-48DD-9D19-6261354F56FE>
- Pesaresi, M., Schiavina, M., Politis, P., Freire, S., Krasnodebska, 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). <https://doi.org/10.1080/17538947.2024.2390454>
- Pfösch, K. A., Schug, F., Cox, H. R., Pham, V.-D., Van Der Linden, S., & Radeloff, V. C. (2026). Monitoring growth of the wildland-urban interface in 2000 and 2020 in Mediterranean ecosystems with Landsat satellite imagery. *Science of Remote Sensing*, 13, 100387. <https://doi.org/10.1016/j.srs.2026.100387>
- Pickens, A. H., Hansen, M. C., Hancher, M., Stehman, S. V., Tyukavina, A., Potapov, P., Marroquin, B., & Sherani, Z. (2020). Mapping and sampling to characterize global inland water dynamics from 1999 to 2018 with full Landsat time-series. *Remote Sensing of Environment*, 243, 111792. <https://doi.org/10.1016/j.rse.2020.111792>
- Potapov, P., Hansen, M. C., Pickens, A., Hernandez-Serna, A., Tyukavina, A., Turubanova, S., Zalles, V., Li, X., Khan, A., Stolle, F., Harris, N., Song, X.-P., Baggett, A., Kommareddy, I., & Kommareddy, A. (2022). The global 2000-2020 land cover and land use change dataset derived from the Landsat archive: First results. *Frontiers in Remote Sensing*, 3, 856903. <https://doi.org/10.3389/frsen.2022.856903>
- PSA. (2017). *Philippine Population Surpassed the 100 Million Mark (Results from the 2015 Census of Population)*. Philippine Statistics Authority (PSA). <https://psa.gov.ph/content/philippine-population-surpassed-100-million-mark-results-2015-census-population>

- PSA. (2021). *2020 Census of Population and Housing (2020 CPH) Population Counts Declared Official by the President*. Philippine Statistics Authority (PSA). Retrieved July 4, 2024, from <https://psa.gov.ph/content/2020-census-population-and-housing-2020-cph-population-count-s-declared-official-president>
- Rabanser, S., Januschowski, T., Flunkert, V., Salinas, D., & Gasthaus, J. (2020). *The effectiveness of discretization in forecasting: An empirical study on neural time series models* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2005.10111>
- Ren, H., Huang, X., Yang, J., & Zhou, G. (2025). Improving 30-meter global impervious surface area (GISA) mapping: New method and dataset. *ISPRS Journal of Photogrammetry and Remote Sensing*, 220, 354–376. <https://doi.org/10.1016/j.isprsjprs.2024.12.023>
- Robert, C. P., & Casella, G. (2004). *Monte Carlo Statistical Methods* (Second Edition). Springer.
- Robi, R. K., & George, J. K. (2025). Application of machine learning algorithms to predict urban expansion. *Journal of Urban Planning and Development*, 151(2), 03125001. <https://doi.org/10.1061/JUPDDM.UPENG-5466>
- Romano, Y., Patterson, E., & Candes, E. (2019). Conformalized quantile regression. *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*.
- Ruan, L., He, T., Xiao, W., Chen, W., Lu, D., & Liu, S. (2022). Measuring the coupling of built-up land intensity and use efficiency: An example of the Yangtze River Delta urban agglomeration. *Sustainable Cities and Society*, 87, 104224. <https://doi.org/10.1016/j.scs.2022.104224>
- Sajjad, M., Khan, Z. A., Ullah, A., Hussain, T., Ullah, W., Lee, M. Y., & Baik, S. W. (2020). A novel CNN-GRU-based hybrid approach for short-term residential load forecasting. *IEEE Access*, 8, 143759–143768. <https://doi.org/10.1109/ACCESS.2020.3009537>
- Sakib, M., Mustajab, S., & Alam, M. (2025). Ensemble deep learning techniques for time series analysis: A comprehensive review, applications, open issues, challenges, and future directions. *Cluster Computing*, 28(1), 73. <https://doi.org/10.1007/s10586-024-04684-0>
- Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3), 1181–1191. <https://doi.org/10.1016/j.ijforecast.2019.07.001>
- Santé, I., García, A. M., Miranda, D., & Crecente, R. (2010). Cellular automata models for the simulation of real-world urban processes: A review and analysis. *Landscape and Urban Planning*, 96(2), 108–122. <https://doi.org/10.1016/j.landurbplan.2010.03.001>
- Santillan, J. R., Dorozynski, M., & Heipke, C. (2025a). 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. <https://doi.org/10.5194/isprs-annals-X-G-2025-753-2025>

- Santillan, J., Dorozynski, M., & Heipke, C. (2025b). Enhancing land use efficiency assessment through built-up area–built-up volume trajectories: Integrating vertical urban growth into SDG 11.3.1 monitoring. *ISPRS International Journal of Geo-Information*, 14(10), 404. <https://doi.org/10.3390/ijgi14100404>
- Santillan, J. R., Dorozynski, M., & Heipke, C. (2026). A deep learning framework for forecasting built-up area and population to support land use efficiency projections under SDG 11.3.1. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-5/W4-2025, 385–392. <https://doi.org/10.5194/isprs-annals-X-5-W4-2025-385-2026>
- Santillan, J. R., & Heipke, C. (2023). Using GHSL to Analyze Urbanization and land-use efficiency in the Philippines from 1975-2020: Trends and implications for sustainable development. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10(1-W1-2023), 413–422. <https://doi.org/10.5194/isprs-annals-X-1-W1-2023-413-2023>
- Santillan, J. R., & Heipke, C. (2024a). 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*. <https://doi.org/10.1007/s41064-024-00305-y>
- Santillan, J. R., & Heipke, C. (2024b). Visualizing the Philippines’ SDG 11.3.1 status and progress using open-source geospatial software and earth observation datasets. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-5–2024, 173–180. <https://doi.org/10.5194/isprs-annals-X-5-2024-173-2024>
- Schiavina, M., Melchiorri, M., Corbane, C., Florczyk, A. J., Freire, S., Pesaresi, M., & Kemper, T. (2019). Multi-scale estimation of land use efficiency (SDG 11.3.1) across 25 years using global open and free data. *Sustainability (Switzerland)*, 11(20). <https://doi.org/10.3390/su11205674>
- 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. <https://doi.org/10.1080/1747423X.2022.2055184>
- 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. <https://doi.org/10.1016/j.habitatint.2022.102543>
- Schneider, A., & Woodcock, C. E. (2008). Compact, dispersed, fragmented, extensive? A comparison of urban growth in twenty-five global cities using remotely sensed data, pattern metrics and census information. *Urban Studies*, 45(3), 659–692. <https://doi.org/10.1177/0042098007087340>

- Semenoglou, A.-A., Spiliotis, E., Makridakis, S., & Assimakopoulos, V. (2021). Investigating the accuracy of cross-learning time series forecasting methods. *International Journal of Forecasting*, 37(3), 1072–1084. <https://doi.org/10.1016/j.ijforecast.2020.11.009>
- Shen, L., Chen, M., Fu, H., Ren, X., Wang, X. J., Sun, J., Li, Z., & Liu, C. (2025). *The Power of Architecture: Deep Dive into Transformer Architectures for Long-Term Time Series Forecasting* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2507.13043>
- Shi, Y., Sun, X., Zhu, X., Li, Y., & Mei, L. (2012). Characterizing growth types and analyzing growth density distribution in response to urban growth patterns in peri-urban areas of Lianyungang City. *Landscape and Urban Planning*, 105(4), 425–433. <https://doi.org/10.1016/j.landurbplan.2012.01.017>
- Sikder, S. K., Kapoor, M., Parmar, P., Tariqul Islam, Md., & Nahiduzzaman, K. M. (2025). Global open data in local urban development: An actionable framework for adopting Global Human Settlement Layer (GHSL) in the Global South. *Sustainable Communities*, 2(1), 2518428. <https://doi.org/10.1080/29931282.2025.2518428>
- Soltani, A., Azizi, P., Rahimioun, A., & Sedaghatfard, M. (2025). Volumetric urban sprawl: Horizontal and vertical growth in two metropolitans. *Journal of Urban Management*. <https://doi.org/10.1016/j.jum.2025.03.006>
- Song, X., Wang, H., Cao, W., Deng, L., Zhang, Y., & He, Y. (2025). Deep learning-based time series forecasting. *Artificial Intelligence Review*, 58(1). <https://doi.org/10.1007/s10462-024-10989-8>
- Stehman, S. V. (2000). Practical implications of design-based sampling inference for thematic map accuracy assessment. *Remote Sensing of Environment*, 72(1), 35–45. [https://doi.org/10.1016/S0034-4257\(99\)00090-5](https://doi.org/10.1016/S0034-4257(99)00090-5)
- Stehman, S. V. (2009). Model-assisted estimation as a unifying framework for estimating the area of land cover and land-cover change from remote sensing. *Remote Sensing of Environment*, 113(11), 2455–2462. <https://doi.org/10.1016/j.rse.2009.07.006>
- Stehman, S. V. (2013). Estimating area from an accuracy assessment error matrix. *Remote Sensing of Environment*, 132, 202–211. <https://doi.org/10.1016/j.rse.2013.01.016>
- Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLOS ONE*, 10(2), e0107042. <https://doi.org/10.1371/journal.pone.0107042>
- Sun, Y., Jiao, L., Guo, Y., & Xu, Z. (2024). Recognizing urban shrinkage and growth patterns from a global perspective. *Applied Geography*, 166. <https://doi.org/10.1016/j.apgeog.2024.103247>
- Swilling, M., & Hajer, M. (2017). Governance of urban transitions: Towards sustainable resource efficient urban infrastructures. *Environmental Research Letters*, 12(12), 125007. <https://doi.org/10.1088/1748-9326/aa7d3a>

- Taubenböck, H., Mast, J., Lemoine Rodríguez, R., Debray, H., Wurm, M., & Geiß, C. (2025). Was global urbanization from 1985 to 2015 efficient in terms of land consumption? *Habitat International*, 160, 103397. <https://doi.org/10.1016/j.habitatint.2025.103397>
- Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, 54(8), 799–805. <https://doi.org/10.1057/palgrave.jors.2601589>
- Taylor, J. W. (2012). Short-term load forecasting with exponentially weighted methods. *IEEE Transactions on Power Systems*, 27(1), 458–464. <https://doi.org/10.1109/TPWRS.2011.2161780>
- Theobald, D. M., Stevens, D. L., White, D., Urquhart, N. S., Olsen, A. R., & Norman, J. B. (2007). Using GIS to generate spatially balanced random survey designs for natural resource applications. *Environmental Management*, 40(1), 134–146. <https://doi.org/10.1007/s00267-005-0199-x>
- Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *TRENDS in Sport Sciences*, 21(1), 19–25. Retrieved December 23, 2025, from <https://tss.awf.poznan.pl/The-need-to-report-effect-size-estimates-revisited-An-overview-of-some-recommended,188960,0,2.html>
- UN Department of Economic and Social Affairs. (2024). *Sustainable Development Goal 11*. Retrieved April 7, 2024, from <https://sdgs.un.org/goals/goal11>
- UN Population Division. (2025). *UN Population Division DataPortal: Interactive access to global demographic indicators*. Retrieved December 31, 2025, from <https://population.un.org/dataportal/home>
- UN Statistics Division. (2021). *SDG indicator metadata (Harmonized metadata template – format version 1.0). Last updated 2021-03-01*. Retrieved August 28, 2022, from <https://unstats.un.org/sdgs/metadata/files/Metadata-11-03-01.pdf>
- UN Statistics Division. (2025). *SDG indicator metadata (Harmonized metadata template – format version 1.0). Last updated 2025-04-23*. Retrieved October 5, 2025, from <https://unstats.un.org/sdgs/metadata/files/Metadata-11-03-01.pdf>
- UN-Habitat. (2018a). *Metadata on SDGs Indicator 11.3.1 Indicator category: Tier II*. United Nations Human Settlements Program (UN-Habitat). Retrieved May 10, 2025, from https://unhabitat.org/sites/default/files/2020/07/metadata_on_sdg_indicator_11.3.1.pdf
- UN-Habitat. (2018b). *SDG Indicator 11.3.1 Training Module: Land Use Efficiency*. United Nations Human Settlement Programme (UN-Habitat). Retrieved August 28, 2022, from https://unhabitat.org/sites/default/files/2021/08/indicator_11.3.1_training_module_land_use_efficiency.pdf

- UN-Habitat. (2020). *World Cities Report 2020: The Value of Sustainable Urbanization*. United Nations Human Settlements Program (UN-Habitat). Retrieved April 20, 2025, from <https://unhabitat.org/world-cities-report-2020-the-value-of-sustainable-urbanization>
- UN-Habitat. (2024). *Template for compilation of SDG indicator 11.3.1*. Urban Indicators Database; UN-Habitat. Retrieved December 16, 2025, from <https://data.unhabitat.org/documents/GUO-UN-Habitat::template-for-compilation-of-sdg-indicator-11-3-1/about>
- United Nations Statistics Division. (2026). *SDG Indicators Metadata Repository*. United Nations. Retrieved March 11, 2026, from <https://unstats.un.org/sdgs/metadata/>
- US Census Bureau. (2024). Chapter 12: Variance Estimation. In *American Community Survey and Puerto Rico Community Survey Design and Methodology (Version 4.0)*. US Census Bureau. Retrieved April 12, 2026, from https://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/2024/acs_design_methodology_ch12_2024.pdf
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. ukasz, & Polosukhin, I. (2017). Attention is All you Need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc. Retrieved October 21, 2025, from https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
- 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. <https://doi.org/10.1016/j.jclepro.2024.141147>
- Wang, H., Liu, Y., Sun, L., Ning, X., & Li, G. (2025). Assessment of Chinese urban land-use efficiency (SDG11.3.1) utilizing high-precision urban built-up area data. *Geography and Sustainability*, 6(1), 100210. <https://doi.org/10.1016/j.geosus.2024.06.007>
- 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. <https://doi.org/10.1016/j.apenergy.2018.10.078>
- Wang, Y., Huang, C., Feng, Y., Zhao, M., & Gu, J. (2020). Using earth observation for monitoring SDG 11.3.1-ratio of land consumption rate to population growth rate in Mainland China. *Remote Sensing*, 12(3). <https://doi.org/10.3390/rs12030357>
- Wang, Y., Li, B., & Xu, L. (2022a). Monitoring land-use efficiency in China's Yangtze River Economic Belt from 2000 to 2018. *Land*, 11(7), 1009. <https://doi.org/10.3390/land11071009>
- Wang, Y., Sun, B., & Zhang, T. (2022b). Do polycentric urban regions promote functional spillovers and economic performance? Evidence from China. *Regional Studies*, 56(1), 63–74. <https://doi.org/10.1080/00343404.2020.1841147>

- Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., & Sun, L. (2022). *Transformers in Time Series: A Survey*. <https://doi.org/10.48550/ARXIV.2202.07125>
- Wen, R., Torkkola, K., Narayanaswamy, B., & Madeka, D. (2017). *A multi-horizon quantile recurrent forecaster* (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.1711.11053>
- Wilcoxon, W. (1945). Individual comparisons by ranking methods. *Biometrics*, 1, 80–83.
- WorldPop. (2026). *Top-down estimation modelling: Constrained vs Unconstrained*. WorldPop, School of Geography and Environmental Sciences, University of Southampton. Retrieved January 29, 2026, from https://www.worldpop.org/methods/top_down_constrained_vs_unconstrained/
- WorldPop and CIESIN. (2018). *Global High Resolution Population Denominators Project—Funded by The Bill and Melinda Gates Foundation (OPP1134076)*. WorldPop (www.worldpop.org - School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur) and Center for International Earth Science Information Network (CIESIN), Columbia University. Retrieved January 29, 2026, from <https://dx.doi.org/10.5258/SOTON/WP00645>
- Xu, T., Gao, J., & Coco, G. (2019). Simulation of urban expansion via integrating artificial neural network with Markov chain – cellular automata. *International Journal of Geographical Information Science*, 33(10), 1960–1983. <https://doi.org/10.1080/13658816.2019.1600701>
- Xu, Z., Duan, X., Lan, T., Wu, Y., Wang, C., Zhong, Y., & Wang, H. (2024). Tracking the scaling of urban open spaces in China from 1990 to 2020. *Scientific Reports*, 14(1), 11891. <https://doi.org/10.1038/s41598-024-62880-9>
- Yaagoubi, R., Lakber, C.-E., & Miky, Y. (2024). A comparative analysis on the use of a cellular automata Markov chain versus a convolutional LSTM model in forecasting urban growth using Sentinel 2A images. *Journal of Land Use Science*, 19(1), 258–277. <https://doi.org/10.1080/1747423X.2024.2403789>
- Yang, T., & Zhang, Y. (2025). Urban systems science and cross-scale dynamics: A conceptual framework to advance integrated planning and governance. *Transactions in Planning and Urban Research*, 4(1), 3–13. <https://doi.org/10.1177/27541223251325264>
- Yang, W., Xu, Q., Yi, S., Shankar, R., & Chen, T. (2024). Enhancing transit-oriented development sustainability through the integrated node-place-ecology (NPE) model. *Transportation Research Part D: Transport and Environment*, 136, 104456. <https://doi.org/10.1016/j.trd.2024.104456>
- Yang, X., Xu, B., Wang, H., Qin, X., Wang, X., Ren, Z., Yao, Y., Zhou, S., Liu, Y., & Chang, P. (2026). Hybrid high-dimensional vine copula–bayesian network framework for flood risk analysis in reservoir–lake systems: Addressing multisource uncertainties. *Environmental Modelling & Software*, 197, 106818. <https://doi.org/10.1016/j.envsoft.2025.106818>

- Yiannakou, A., & Zografos, G. (2025). Spatial patterns of land take in a Mediterranean City: An assessment of the SDG indicator 11.3.1 in the peri-urban area of Thessaloniki. *Land*, *14*(5), 965. <https://doi.org/10.3390/land14050965>
- Yu, F., & Koltun, V. (2015). *Multi-Scale Context Aggregation by Dilated Convolutions* (Version 3). arXiv. <https://doi.org/10.48550/ARXIV.1511.07122>
- Zambon, I., Colantoni, A., & Salvati, L. (2019). Horizontal vs vertical growth: Understanding latent patterns of urban expansion in large metropolitan regions. *Science of The Total Environment*, *654*, 778–785. <https://doi.org/10.1016/j.scitotenv.2018.11.182>
- 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. <https://doi.org/10.5194/essd-14-1831-2022>
- Zhang, Z., Jiang, W., Ling, Z., Guo, X., Song, J., Xiao, Z., & Yin, X. (2026). Urban land use prediction and construction effectiveness assessment supporting SDG 11.3.1 and regional planning goals. *Land Use Policy*, *162*, 107893. <https://doi.org/10.1016/j.landusepol.2025.107893>
- Zhao, C., Weng, Q., & Hersperger, A. M. (2020). Characterizing the 3-D urban morphology transformation to understand urban-form dynamics: A case study of Austin, Texas, USA. *Landscape and Urban Planning*, *203*, 103881. <https://doi.org/10.1016/j.landurbplan.2020.103881>
- Zhong, C., Chen, J., Yi, B., & Li, H. (2025a). Examining the reliable trend of global urban land use efficiency from 1985 to 2020 using robust indicators and analysis tools. *Habitat International*, *163*, 103477. <https://doi.org/10.1016/j.habitatint.2025.103477>
- Zhong, C., Peng, L., Yu, J., Swan, I., & Li. (2025b). Toward more reliable, complete, and equitable global urban land use efficiency assessments. *Communications Earth & Environment*, *6*(1055). <https://doi.org/10.1038/s43247-025-03033-2>
- Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2020). *Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting* (Version 3). arXiv. <https://doi.org/10.48550/ARXIV.2012.07436>
- Zhou, M., Lu, L., Guo, H., Weng, Q., Cao, S., Zhang, S., & Li, Q. (2021). Urban sprawl and changes in land-use efficiency in the Beijing–Tianjin–Hebei Region, China from 2000 to 2020: A spatiotemporal analysis using earth observation data. *Remote Sensing*, *13*(15), 2850. <https://doi.org/10.3390/rs13152850>
- Zhou, Y., Li, X., Chen, W., Meng, L., Wu, Q., Gong, P., & Seto, K. C. (2022). Satellite mapping of urban built-up heights reveals extreme infrastructure gaps and inequalities in the Global South. *Proceedings of the National Academy of Sciences*, *119*(46). <https://doi.org/10.1073/pnas.2214813119>

- Zhu, X., Li, Y., Zhang, P., Wei, Y., Zheng, X., & Xie, L. (2019). Temporal-spatial characteristics of urban land use efficiency of China's 35mega cities based on DEA: Decomposing technology and scale efficiency. *Land Use Policy*, 88, 104083. <https://doi.org/10.1016/j.landusepol.2019.104083>
- Zięba-Kulawik, K., Skoczylas, K., Mustafa, A., Wężyk, P., Gerber, P., Teller, J., & Omrani, H. (2020). Spatiotemporal changes in 3D building density with LiDAR and GEOBIA: A city-level analysis. *Remote Sensing*, 12(21), 3668. <https://doi.org/10.3390/rs12213668>

Appendix

Table A1: Hyperparameters and search space for the deep learning architectures described in Section 5.4.3.

Architecture	Hyperparameter	Search Space
1D CNN	Conv1D layers	{1, 2}
	Conv1D filters	{32, 64, 128}
	Kernel size	{3, 5}
	Stride	{1}
	Dilation rates (layer 1 / layer 2)	{1}/{1, 2}
	L2 regularization factor	$\{10^{-6}, 10^{-5}, 10^{-4}\}$
	Dropout rate	{0.0, 0.1, 0.2}
	Hidden layer size (hidden units)	{32, 64, 128}
	Optimizer learning rate	$\{10^{-3}, 5 \times 10^{-4}\}$
	Gradient clip threshold	{1.0}
	Batch size	{32, 64}
LSTM	LSTM units	{32, 64, 128}
	Bidirectional	{False}
	LSTM dropout rate	{0.0, 0.1, 0.2}
	Recurrent dropout rate	{0.0, 0.1}
	Dropout rate	{0.0, 0.1, 0.2}
	Hidden layer size (hidden units)	{32, 50, 64, 128}
	L2 regularization factor	$\{10^{-6}, 10^{-5}, 10^{-4}\}$
	Optimizer learning rate	$\{10^{-3}, 5 \times 10^{-4}, 3 \times 10^{-4}\}$
	Gradient clip threshold	{1.0}
	Batch size	{32, 64}
GRU	GRU units	{32, 64, 128}
	GRU dropout rate	{0.0, 0.1, 0.2}
	Recurrent dropout rate	{0.0, 0.1}
	Dropout rate	{0.0, 0.1, 0.2}
	Hidden layer size (hidden units)	{32, 50, 64, 128}
	L2 regularization factor	$\{10^{-6}, 10^{-5}, 10^{-4}\}$
	Optimizer learning rate	$\{10^{-3}, 5 \times 10^{-4}, 3 \times 10^{-4}\}$
	Gradient clip threshold	{1.0}
	Batch size	{32, 64}
1D CNN +LSTM	Conv1D layers	{1, 2}
	Conv1D filters	{32, 64, 128}
	Kernel size	{3, 5}
	Stride	{1}
	Dilation rates (layer 1 / layer 2)	{1}/{1, 2}

Table A1: Hyperparameters and search space for the deep learning architectures described in Section 5.4.3 (continued).

Architecture	Hyperparameter	Search Space
	LSTM units	{32, 64, 128}
	Bidirectional	{False}
	LSTM dropout rate	{0.0, 0.1, 0.2}
	Recurrent dropout rate	{0.0, 0.1}
	Dropout rate	{0.0, 0.1, 0.2}
	Hidden layer size (hidden units)	{32, 50, 64, 128}
	L2 regularization factor	$\{10^{-6}, 10^{-5}, 10^{-4}\}$
	Optimizer learning rate	$\{10^{-3}, 5 \times 10^{-4}, 3 \times 10^{-4}\}$
	Gradient clip threshold	{1.0}
	Batch size	{32, 64}
1D CNN +GRU	Conv1D layers	{1, 2}
	Conv1D filters	{32, 64, 128}
	Kernel size	{3, 5}
	Stride	{1}
	Dilation rates (layer 1 / layer 2)	{1}/{1, 2}
	GRU units	{32, 64, 128}
	Bidirectional	{False}
	GRU dropout rate	{0.0, 0.1, 0.2}
	Recurrent dropout rate	{0.0, 0.1}
	Dropout rate	{0.0, 0.1, 0.2}
	Hidden layer size (hidden units)	{32, 50, 64, 128}
	L2 regularization factor	$\{10^{-6}, 10^{-5}, 10^{-4}\}$
	Optimizer learning rate	$\{10^{-3}, 5 \times 10^{-4}, 3 \times 10^{-4}\}$
	Gradient clip threshold	{1.0}
	Batch size	{32, 64}
Transformer	Model dimension / hidden layer size	{48, 64, 96, 128}
	Attention heads	{2, 4, 8}
	Transformer blocks	{1, 2, 3}
	Feed-forward dimension	{64, 128, 256}
	Projection dropout rate	{0.0, 0.1}
	Attention dropout rate	{0.0, 0.1}
	Feedforward dropout rate	{0.0, 0.1}
	Head dropout rate	{0.0, 0.1}
	Learned positional encoding	{True, False}
	L2 regularization factor	$\{10^{-6}, 10^{-5}, 10^{-4}\}$
	Optimizer learning rate	$\{10^{-3}, 5 \times 10^{-4}, 3 \times 10^{-4}\}$
	Gradient clip threshold	{1.0}
	Batch size	{32, 64}

Table A2: Year-specific forecast performance of candidate models for built-up area (BUA) prediction in 2015 and 2020 across 1,199 test-set urban centers. MedE = median error; MedAE = median absolute error (km^2); MedAPE = median absolute percentage error (%); PICP = prediction interval coverage probability; PIAW = prediction interval average width (km^2); Median WS = median Winkler Score.

Model	MedE 2015	MedE 2020	MedAE 2015	MedAE 2020	MedAPE 2015	MedAPE 2020	PICP 2015	PICP 2020	Median PIAW 2015	Median PIAW 2020	Median WS 2015	Median WS 2020
1D CNN	0.00	0.00	0.02	0.07	0.49	1.67	0.96	0.94	0.14	0.49	0.15	0.52
LSTM	0.00	0.00	0.01	0.07	0.37	1.65	0.96	0.96	0.12	0.53	0.12	0.55
GRU	0.00	0.00	0.01	0.07	0.35	1.66	0.97	0.96	0.12	0.53	0.13	0.55
1D CNN+LSTM	0.00	0.00	0.02	0.07	0.48	1.65	0.96	0.96	0.15	0.52	0.15	0.53
1D CNN+GRU	0.00	0.00	0.02	0.07	0.41	1.62	0.97	0.96	0.14	0.52	0.14	0.53
Transformer	0.00	0.00	0.02	0.07	0.41	1.68	0.97	0.91	0.13	0.43	0.14	0.46
Exponential Smoothing	0.00	0.11	0.13	0.32	3.20	6.41	0.69	0.49	0.46	0.53	0.99	3.25

Table A3: Year-specific forecast performance of candidate models for population (Pop) prediction in 2015 and 2020 across 1,199 test-set urban centers. MedE = median error; MedAE = median absolute error (no. of persons); MedAPE = median absolute percentage error (%); PICP = prediction interval coverage probability; PIAW = prediction interval average width (no. of persons); Median WS = median Winkler Score.

Model	MedE 2015	MedE 2020	MedAE 2015	MedAE 2020	MedAPE 2015	MedAPE 2020	PICP 2015	PICP 2020	Median PIAW 2015	Median PIAW 2020	Median WS 2015	Median WS 2020
1D CNN	−41	−152	1222	3603	1.01	2.99	0.94	0.94	10562	28583	10982	30212
LSTM	−55	−287	1216	4071	1.05	3.32	0.95	0.95	11892	32090	12248	33253
GRU	−73	−222	1165	4029	1.03	3.25	0.95	0.95	12114	32374	12565	33857
1D CNN+LSTM	−52	−197	1169	3772	1.04	3.12	0.94	0.94	11409	30760	11973	32316
1D CNN+GRU	−62	−206	1237	3958	1.06	3.17	0.95	0.94	11491	31175	12058	33169
Transformer	17	−106	1228	3708	1.05	3.16	0.95	0.76	11129	14750	11640	19680
Exponential Smoothing	−389	1328	2693	6165	2.13	4.95	0.86	0.74	16155	24991	19346	37667

Curriculum Vitae

PERSONAL INFORMATION

Name	Jojene Rendon Santillan
Date and place of birth	02.12.1981 in Hamtic, Antique, Philippines

EDUCATION

June 2005–May 2008	<i>Master of Science in Remote Sensing</i> University of the Philippines, Diliman, Quezon City Philippines
June 1999–March 2004	<i>Bachelor of Science in Geodetic Engineering</i> University of the Philippines, Diliman, Quezon City Philippines

WORK EXPERIENCE

Since September 2022	<i>Professor 3</i> College of Engineering and Geosciences Caraga State University, Butuan City, Philippines
July 2019–August 2022	<i>Professor 2</i> College of Engineering and Geosciences Caraga State University, Butuan City, Philippines
May 2017–June 2019	<i>Assistant Professor 1</i> College of Engineering and Geosciences Caraga State University, Butuan City, Philippines
April 2014–March 2017	<i>Chief Science Research Specialist</i> Phil-LiDAR 1.B.14 Project Caraga State University, Butuan City, Philippines
April 2008–March 2014	<i>Research Associate</i> Department of Geodetic Engineering & Training Center for Applied Geodesy and Photogrammetry University of the Philippines-Diliman, Quezon City, Philippines
November 2004– October 2005	<i>Research Associate</i> Institute of Environmental Science and Meteorology University of the Philippines-Diliman, Quezon City, Philippines

