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

Spatial and Socioeconomic Drivers of Land Use Change in High-Density Residential Areas (R-3) in Quezon City: A Predictive Simulation Using the Cellular Automata-Markov Chain Model

Winnie Rose S. Dario[1], Ranielle Adelaine D. David[1], Erica Erin E. Elazegui[1], Alexis Richard C. Claridades[1]

[1]Department of Geodetic Engineering, University of the Philippines, Diliman, Quezon City, Philippines –
{wsdario, rddavid4, eeelazegui, acclaridades}@up.edu.ph

Keywords: Binary Logistic Regression (BLR), Markov Chain (MC), Cellular Automata (CA), CA-Markov, Land Use Change, High-density Residential Areas

Abstract

High-density residential (R-3) areas, characterized by flexible housing typologies and increasing population pressures, are particularly vulnerable to rapid urbanization, especially in highly urbanized cities like Quezon City. However, the complex interplay of spatial and socioeconomic factors driving R-3 zone changes is often overlooked in existing literature. To address this gap, this study employed binary logistic regression (BLR) to determine key drivers of R-3 zone transitions and the Cellular Automata (CA)–Markov model to predict future land use maps, utilizing historical data from 2000, 2009, and 2011. The BLR model, with acceptable discriminative ability (AUC = 0.78; pseudo-R^2 = 0.083), identified 24 statistically significant driving factors at the 95% confidence level. Results revealed that proximity to low-intensity industrial areas, central business districts, roads, informal settlements, vacant lots, population density, and specific soil types positively influenced R-3 transitions. Conversely, elevation, medium-intensity industrial areas, open spaces, cemeteries, reservoirs, waterways, and other soil types affect R-3 changes negatively, with distance from socialized housing and high traffic volume emerging as strong negative drivers. The suitability map, derived from the BLR model applied with a 5 x 5 neighborhood, and the transition areas matrix that accounts for the pixel-level land use shifts from the Markov model successfully projected a 2011 land use map with an acceptable model performance (Overall Kappa = 0.70). By identifying the statistically significant spatial and socioeconomic drivers influencing R-3 changes and predicting future developments, this study offers planners and local governments insights to address overcrowding, housing shortages, and informal settlement growth.

1. Introduction

As of 2023, around 48.29% of the Philippine population resides in urban areas, reflecting the country's rapid trend toward urbanization (World Bank, 2023). This trend is particularly pronounced in Metro Manila, where Quezon City stands out as a highly urbanized city experiencing significant population pressures (Chavez, 2023). However, this progress also brings a pressing challenge in accommodating a growing population while managing the proliferation of informal settlements. Residential zones in Quezon City are classified into low-density (R-1), medium-density (R-2), and high-density (R-3) categories, distinguished mainly by allowable housing types and dwelling unit densities, with R-3 areas accommodating the highest density (over 101 dwelling units per hectare) and a broader mix of housing typologies. These R-3 areas play a pivotal role in addressing this challenge, as their flexible housing typologies allow them to support higher population densities. At the same time, these areas are also vulnerable to unplanned development and spatial pressure, often being associated with the emergence of informal settlements. Understanding land use change (LUC) dynamics within R-3 areas is therefore critical to devising more responsive and sustainable urban development strategies.

Despite this urgency, existing literature often presents significant gaps. First, the complex interplay between biophysical and socioeconomic factors driving LUC in high-density areas remain underexplored. Many studies tend to focus on either natural environmental factors or socioeconomic drivers in isolation, rarely integrating the two. This separation, which can arise from disciplinary silos or technical challenges related to integrating disparate datasets, limits the ability to fully understand how these variables jointly shape land transformation, particularly in densely populated urban zones. Furthermore, there is a lack of existing local literature specifically focusing on zone-specific LUC within the Philippine context. Most studies treat urban areas as homogeneous units, thereby overlooking the unique behaviors of sub-zones like R-3 areas, where rapid vertical expansion and informal developments coexist. The lack of local literature focused specifically on R-3 zones makes it difficult to formulate targeted planning interventions. Hence, addressing the complex nature of urban land use transformation necessitates integrating spatial and socioeconomic factors to avoid oversimplifying urban dynamics and overlooking how physical constraints and human behavior jointly shape land transformation.

To address these gaps, this study employs an integrated modeling approach using Binary Logistic Regression (BLR) to identify significant spatial and socioeconomic drivers of LUC and the Cellular Automata–Markov Chain (CA-Markov) model to forecast future land use scenarios. The CA-Markov model, particularly when combined with regression-based driver analysis, has been shown to improve land use forecasting by effectively capturing both spatial dependencies and temporal dynamics (Mathanraj et al., 2021). By applying this hybrid model specifically to R-3 areas in Quezon City, this study aims to fill the identified research gaps by analyzing the spatial and socioeconomic drivers of LUC through the application of the integrated CA-Markov and BLR models. Through pinpointing key spatial and socioeconomic drivers and projecting potential future scenarios, this study strives to deepen the understanding of urban land use transformation in the Philippines. The insights generated can support more sustainable urban planning and policy decisions. In doing so, the research corresponds with the objectives of Sustainable Development Goal 11 (SDG 11), which emphasizes the importance of fostering inclusive, safe, resilient, and sustainable cities—especially by promoting adequate housing and addressing the rise of informal settlements.

2. Materials and Methods

2.1 Study Area

Quezon City is a central hub for urbanization and population growth in Metro Manila. It is recognized as the top-performing, highly urbanized city in the region. Despite its status, Quezon City also accommodates a significant number of informal settlers, with an estimated family population of 195,875 as of 2020 (ICLEI South Asia, 2021). Analyzing the dynamics of urbanization and the factors driving land use transformation in Quezon City, particularly within its R-3 areas, is crucial for this study. Figure 1 shows the study area.

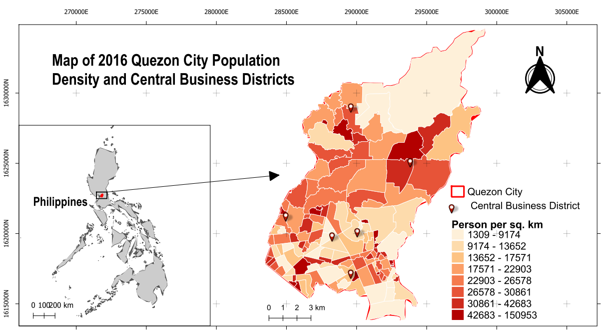

Figure 1. Map of the Study Area

2.2 Materials

2.2.1 Research Workflow: This section shows the processes of the study, including the BLR analysis, Markov chain, CA, and the integration of CA-Markov predictive modelling, as shown in Figure 2.

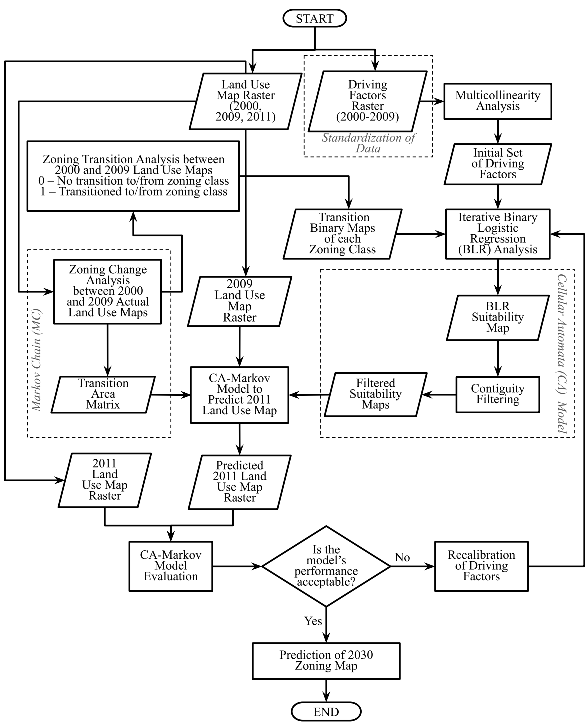

Figure 2. Research Workflow

2.2.2 Data Set: To examine the drivers of LUC in high-density residential areas in Quezon City, an initial set of relevant spatial and socioeconomic datasets was compiled. These datasets were selected based on their relevance to LUC and the availability of consistent and trustworthy data sources. The collected information was sourced from various government agencies and open-access spatial repositories, as shown in Table 1.

Data	Source
Land Use Maps, Soil Map, Universities, Health Centers, Population Density (2000)	Quezon City Planning and Development Department
Boundary Shapefile, Digital Elevation Model (DEM)	Humanitarian Data Exchange
Earthquake and Flood Risk Hotspots (2000)	Department of Human Settlements and Urban Development
Annual Average Daily Traffic (AADT) (2000)	Metropolitan Manila Development Authority
Central Business Districts (CBDs), Protected Areas	Comprehensive Land Use Plan 2011-2025
Train Stations, Bus Stops and Terminals	Department of Transportation
Nighttime Light (NTL) Data (2000)	Global National Polar-orbiting Partnership Visible Infrared Imaging Radiometer Suite

Table 1. Quezon City Data Requirements and Sources

2.2.3 Data Processing: Various software tools were utilized in this study to process, analyse, and visualize spatial and non-spatial data. Using Google Earth Engine (GEE), a web-based platform for scalable geospatial analysis with satellite imagery and remote sensing data, the pre-processing NTL data was conducted. For geolocation tasks, Google Earth was employed to convert tabular datasets into geographic coordinates by placemarking the locations of facilities and transportation hubs. Meanwhile, Google Maps was used to extract traffic route data, particularly for identifying and mapping general traffic routes to support the development of the traffic dataset.

To perform detailed spatial analyses and produce thematic maps, Geographic Information System software such as QGIS and ArcGIS Pro were utilized. These platforms enabled the integration and manipulation of multiple datasets, including zoning, demographic, and environmental layers, to support model development and interpretation. Finally, TerrSet, an integrated geospatial software system, was used for conducting the core predictive modeling tasks. The software facilitated the generation of transition binary maps as inputs of the BLR Model, execution of the CA-Markov model, and validation processes necessary to evaluate the reliability of the model outputs.

2.3 Methods

The methodology employed in this study was divided into six principal parts, integrating various geospatial and statistical techniques to analyse and predict LUC in high-density residential areas in Quezon City.

2.3.1 Data Preparation: The data preparation phase involves the collection and initial processing of datasets critical for analyzing LUC in Quezon City's high-density residential areas. The process began with obtaining actual land use maps for the years 2000, 2009, and 2011, which were initially in JPG format. These maps were pre-processed, including georeferencing to align with geographic coordinates, followed by the creation of a spatial subset to remove extraneous elements. To convert the visual information into usable categorical data, supervised classification using Maximum Likelihood Classification was performed, resulting in classified raster maps with 20 land use classes.

Concurrently, an initial set of 38 relevant spatial and socioeconomic driving factors was compiled from various government agencies and open-access spatial repositories. Each of these 38 driving factors underwent tailored preprocessing steps to convert them into a consistent raster format suitable for the subsequent models. Examples include extracting elevation from a DEM, digitizing and converting categorical soil type data into binary dummy variables and creating distance-based layers using Euclidean distance for zoning classes, essential services, economic centres, and protected areas. Tabular data, such as population density per barangay and AADT volume, were processed and rasterised. NTL intensity data, used as a proxy for economic activity, was also processed from satellite imagery. As a final, crucial step to ensure comparability and consistency for modeling, all 38 raster datasets were standardized using the Minimum-Maximum method to scale their values to a range of 0 to 1 (Getu & Bhat, 2024) and aligned to a common spatial framework by matching extent, resolution, and coordinate system with the processed land use maps.

2.3.2 Markov Chain Analysis: The Markov Chain analysis was utilized to quantify the temporal transitions in the spatial distribution of land use in Quezon City between two specific time points, utilizing the reference land use maps from 2000 ($t1$) and 2009 ($t2$) as primary inputs to generate the transition areas matrix. The transition areas matrix quantified the expected pixel count that would transition between each set of land use classes, integrating the time-related aspects of LUC into the simulation and informing the likelihood of land use conversions over time based on the observed historical trends between 2000 and 2009.

2.3.3 Calibration of Driving Factors: A critical calibration of driving factors phase was undertaken to refine the initial set of drivers of R-3 zone changes, determining the significant spatial and socioeconomic factors driving LUC within Quezon City.

2.3.3.1 Multicollinearity Analysis: Multicollinearity analysis was performed to identify and eliminate predictor variables with high multicollinearity, ensuring the reliability of the subsequent BLR model and preventing distorted coefficient estimates (Getu & Bhat, 2024). The procedure began with filtering out variables that showed strong redundancy, identified through pairwise correlations greater than a threshold of 0.99. When such a pair was identified, the variable with the higher average R^2 when regressed against other predictors was removed. Following this initial filtering, an iterative Variance Inflation Factor (VIF) analysis was conducted. The VIF quantifies the degree to which a variance of a variable has increased due to collinearity. A threshold of VIF < 5 (Kyriazos & Poga, 2023; Shrestha, 2020) was adopted as the criterion for detecting and removing substantial collinearity. In each iteration, the variable with the highest VIF value was systematically removed until all remaining variables had VIF values below 5. This systematic, data-driven reduction process, which involved 11 iterations, resulted in a final set of 27 spatial and socioeconomic variables with acceptable VIF values representing a more independent and non-redundant pool of predictors for the subsequent LUC analysis.

2.3.3.2 BLR Analysis: BLR was employed to estimate the likelihood of land use transitions specifically within the R-3 areas of Quezon City, as well as to quantify the influence of the selected spatial and socioeconomic factors driving these changes. BLR models the relationship between a binary dependent variable, representing the occurrence (1) or absence (0) of LUC, derived from the 2000 and 2009 land use maps, and multiple independent variables, which were the refined set of driving factors. The process involved estimating coefficients for each factor, indicating how strongly and in what direction they affected the possibility of a land use class transitioning. A positive coefficient implied an increased likelihood of change, while a negative one indicates a decreased likelihood.

To ensure that only meaningful variables contributed to the final model, an iterative variable selection process based on the statistical significance (p-values) of the factors' coefficients was conducted. Starting with the 27 driving factors that remained after the multicollinearity analysis, factors with the highest p-value greater than 0.05 (corresponding to a 95% confidence level) were systematically removed in successive iterations, ultimately resulting in a final set of 24 statistically significant driving factors. Using the coefficients from this refined model, the logit equation was applied across the study area to generate a continuous probability surface, also known as a suitability map, which quantified the pixel-level likelihood of land use transitions based on the influence of the identified significant drivers and served as a crucial spatial input for the subsequent CA-Markov modeling framework.

The performance and accuracy of the BLR model were evaluated through the following metrics, including Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) and McFadden's Pseudo R^2 (Shahbazian et al., 2019; Hu & Lo, 2007). The AUC represents the total area under the ROC curve, generated by plotting true positive rates against false positive rates across all probability thresholds, which measures the model's discriminative ability. An AUC between 0.70 and 0.90 indicates an acceptable discriminatory power. McFadden's Pseudo R^2, calculated from the log-likelihoods of the full model versus a null model with only an intercept, evaluated the model's goodness-of-fit. Generally, McFadden's Pseudo R^2 in the range of 0.2 to 0.4 is considered highly satisfactory.

2.3.4 CA-Markov Model: For forecasting future land use transformations, especially in Quezon City's R-3 zones, this study utilized the CA-Markov framework, which combines the Markov Chain, capturing time-based LUC probabilities from past data, with the CA component that accounts for spatial relationships via neighborhood effects and spatial constraints.

The CA-Markov model requires the reference 2009 land use map as the base layer for the simulation, the transition area matrix from the Markov Chain analysis, which incorporates the temporal dynamics into the simulation, and the suitability maps generated from the BLR analysis, which quantified the pixel-level probability of land use transition occurring, based on the influence of the identified significant spatial and socioeconomic driving factors. The CA component of the model then applied these suitability maps as its transition rules, dividing the study area into discrete cells and using spatial filters to determine the influence of neighboring cells on changes in each cell, thereby accounting for local spatial interactions (Taloor et al., 2024; Long et al., 2023; Mondal et al., 2020).

Specifically, a 5x5 contiguity filter was employed, considering the conditions of 24 neighboring cells, indicating that a cell's next state depended on its present condition and the land use categories of surrounding cells. Several runs of the model were

https://doi.org/10.5194/isprs-annals-X-5-W4-2025-205-2026

performed using transition suitability maps generated from iterative BLR analyses. The output consisted of a projected land use map depicting the future. The simulated land use map for 2011 was contrasted with the real 2011 data to evaluate its predictive accuracy prior to forecasting future land use scenarios. If the validation standards were not met, the driving factors would undergo adjustment and refinement.

2.3.5 Model Performance Validation: The validation of the CA-Markov model's performance was a crucial step to evaluate the accuracy of the generated land use projections. The evaluation involved using the real 2011 land use image as the basis and the CA-Markov model's simulated 2011 output as the comparison dataset. The validation was done using the Validate module in Terrset LiberaGIS, a tool designed for systematically comparing two categorical raster maps to assess both their quantity and spatial agreement.

The methodology involved calculating Kappa Indices of Agreement and related statistics to determine how well the simulated results reflected the real-world land use patterns in 2011. This evaluation was performed on both the overall land use map encompassing all classes and the projected R-3 areas map. When kappa statistics exceed the 0.75 threshold, the model validation is considered successful, reflecting a positive and reliable level of agreement between predicted and actual land use outcomes (Beroho et al., 2023; Gasirabo et al., 2023). This validation was conducted iteratively as the driving factors were refined during the preceding BLR analysis.

3. Results and Discussion

3.1 Multicollinearity Analysis

This analysis began with a pairwise correlation filtering step, which resulted in the removal of 'Distance to R3 Areas' after identifying it as highly correlated ($r > 0.99$) with 'Distance to Socialized Housing' and having a higher average R^2 value when regressed against other predictors. Following this initial filtering, an iterative VIF analysis was conducted, systematically removing the variable with the highest VIF in each iteration until all remaining variables had VIF values below the adopted threshold of 5. After 11 iterations, this process resulted in a final set of 27 spatial and socioeconomic variables, shown in Table 2, with acceptable VIF values (VIF < 5), indicating that the dataset was sufficiently independent for subsequent modeling.

A subsequent Pearson correlation matrix for these 27 variables revealed notable strong positive pairwise relationships, such as between distance from roads and distance from waterways ($r = 0.9564$), which reflects the common urban planning practice of aligning road networks with water drainage or river systems. A dense cluster of correlations was observed around distance from informal settlements, which correlated strongly with distance from open spaces, recreational parks, socialized housing, utilities, and health centers. Other significant correlations included relationships between distance from socialized housing and elevation, utilities, and recreational parks, as well as correlations involving elevation, distance from cemeteries, and proximity to health centers and industrial areas. Conversely, traffic data showed a low correlation with all other variables, consistent with its low VIF value near 1. Despite these strong associations, the analysis confirmed that all 27 retained variables met the VIF criterion, ensuring the statistical soundness of the dataset for subsequent modeling.

X_i	Factor	R^2	VIF
1	Elevation	0.78938	4.74780
2	San Luis Clay Soil Type	0.11752	1.13317
3	Escarpment Soil Type	0.13744	1.15934
4	Burgos Clay Soil Type	0.06851	1.07355
5	Built-up Areas Soil Type	0.01536	1.01560
6	Flood Vulnerability	0.12590	1.14403
7	Earthquake Vulnerability	0.23175	1.30165
8	Distance from Socialized Housing Areas	0.77550	4.45428
9	Distance from Light Intensity Industrial Areas (I-1)	0.03134	1.03235
10	Distance from Medium Intensity Industrial Areas (I-2)	0.72976	3.70035
11	Distance from Open Spaces	0.66838	3.01546
12	Distance from Recreational Parks	0.54345	2.19035
13	Distance from Informal Settlements	0.68523	3.17696
14	Distance from Cemeteries	0.79903	4.97587
15	Distance from Vacant	0.37963	1.61195
16	Distance from Reservoirs	0.58622	2.41672
17	Distance from Waterways	0.43142	1.75876
18	AADT Volume	0.00100	1.00100
19	Distance from CBDs	0.54243	2.18547
20	Distance from Health Centers	0.74641	3.94336
21	Distance from Roads	0.46516	1.86970
22	Distance from Bus Stations	0.69667	3.29677
23	Population Density 2000	0.26358	1.35792
24	NTL Data 2000	0.56648	2.30670
25	San Manuel Clay Soil Type	0.07251	1.07818
26	Distance from Protected Areas	0.52464	2.10365
27	Distance from Utilities	0.68951	3.22068

Note. Factors X_1 to X_{24} with p-values < 0.05 were selected and included in the logit equation (1) of the BLR model.

Table 2. Set of 27 Factors After Removing Variables with High Multicollinearity

3.2 Binary Logistic Regression Model

BLR analysis was employed to identify the statistically significant spatial and socioeconomic driving factors that impact land use shifts within R-3 areas in Quezon City. After the multicollinearity analysis, an initial set of 27 spatial and socioeconomic variables (VIF < 5) were retained for the initial BLR model.

The BLR model achieved an AUC of 0.78, falling between 0.70 and 0.90, indicates an acceptable level of discriminative ability. McFadden's Pseudo R^2 was also computed as a measure of goodness-of-fit, particularly because the dependent variable (change vs. no change in R-3 zone) is binary, making traditional R^2 inappropriate. While considered modest, a McFadden's Pseudo R^2 of 0.083 is deemed acceptable for models dealing with complex spatial phenomena like LUC, suggesting that it still explains a meaningful portion of the variation in R-3 transitions, around 8% better than the null model. Furthermore, the confusion matrix for the test set showed that the model was highly effective at correctly identifying areas that did not change to R-3 but also had a notable number of false positives, a common occurrence with imbalanced datasets, resulting in a high recall of approximately 88.89% with a low precision of 0.86%.

https://doi.org/10.5194/isprs-annals-X-5-W4-2025-205-2026 | © Author(s) 2026. CC BY 4.0 License.

An iterative calibration process was then applied to ensure only statistically significant driving factors at the 95% confidence level (p-value < 0.05) were included in the final model. This process began with the initial 27 factors. Factors with the highest p-values above 0.05 were systematically removed in each iteration, with three variables excluded due to their lack of statistical significance at the 95% confidence level: San Manuel Soil Type, Distance to Protected Areas, and Distance from Utilities Zone, retaining 24 statistically significant driving factors. Based on these factors, the final BLR model is represented by the following logit equation:

$$Logit(p) = -3.9040 - 7.7533X_1 - 0.4271X_2 - 0.5729X_3 + 0.3202X_4 + 1.9294X_5 + 0.1285X_6 + 1.7205X_7 - 21.2297X_8 + 0.4697X_9 - 1.4237X_{10} - 0.6500X_{11} + 2.9012X_{12} + 2.8120X_{13} - 1.5372X_{14} + 1.1341X_{15} - 3.5633X_{16} - 5.5192X_{17} - 34.3558X_{18} + 1.6481X_{19} + 1.2371X_{20} + 6.8015X_{21} - 1.4636X_{22} + 2.5690X_{23} + 0.1779X_{24} \tag{1}$$

where X_i = driving factor in Table 2.

The analysis of the coefficients for these 24 factors revealed how each influences the likelihood of R-3 area transitions. Positive coefficients indicate a positive influence, meaning the factor increases the likelihood of R-3 change, while negative coefficients suggest a negative influence, decreasing the likelihood of R-3 change. This logit equation was then used to generate suitability maps indicating the pixel-level probability of R-3 land use change.

3.3 Influence of Driving Factors on R-3 Zone Changes

The final BLR model of 24 statistically significant driving factors revealed distinct influences on the likelihood of LUC within R-3 areas. Approximately 54% of these factors showed a positive influence, increasing the probability of R-3 zone transitions, while 46% exhibited a negative influence, decreasing this probability. Factors positively associated with R-3 change included proximity to I-1, CBDs, roads, informal settlements, and vacant lots, as well as high population density. Specific soil types such as Burgos Clay and Built-up also showed a positive influence, suggesting suitability for construction.

Conversely, factors negatively influencing R-3 changes included elevation, proximity to I-2, open spaces, cemeteries, reservoirs, and waterways. Distance from socialized housing and high traffic volume emerged as particularly strong negative drivers, with socialized housing possibly carrying a social stigma that deters private investment and high traffic volume reducing liveability. Soil types like Escarpment and San Luis Clay also showed negative effects, likely due to physical constraints and geohazards. Minimal positive effects were observed for NTL and Flood Risk, suggesting limited influence despite potential associations with urban activity or strategic location. These results collectively indicate that R-3 development is primarily influenced by accessibility, availability of space, existing urban pressures, and underlying physical suitability of the land.

3.4 Markov Chain Analysis

Following the BLR analysis, the Markov Chain model was employed to quantify the temporal LUC that occurred between the 2000 and 2009 land use maps, providing insights into the possibility of land use types shifting as time progressed. The key outputs included a transition area matrix which explicitly counts the pixels expected to change between different zoning classes. Analysis of transitions, shown in Table 3, specifically involving R-3 areas revealed distinct patterns: six zoning classifications (R-1, C-3, I-1, Recreational Parks, Cemetery, and Reservoir) showed no transition at all during this period, indicating their stability.

Land Use Class	Transition Count	
	From R-3	To R-3
R-2	23	17
R-3	307,440	307,440
Socialized Housing	957	647
C-1	0	1
C-2	50	12
I-2	0	239
Institutional	0	3
Open Spaces	0	7
Informal Settlements	51	277
Military	0	215
Utilities	18	0
Vacant Lands	4	388
Roads	113	543
Waterways	5	20
Note. R-1, C-1, C-3, I-1, Recreational Parks, Cemeteries, and Reservoirs showed no transitions and were excluded.		

Table 3. Land Use Transitions of Other Land Use Class From and To R-3

Five classes (C-1, I-2, Institutional, Open Spaces, and Military) transitioned to R-3 only, suggesting R-3 expansion into these areas. Conversely, only one class, Utilities, showed transitions from R-3 only. Eight other classes (R-2, Socialized Housing, C-2, Informal Settlements, Vacant Lands, Roads, Waterways, and R-3 itself) exhibited transitions both to and from R-3. Despite these shifts, the R-3 class itself was described as very stable when compared to classes exhibiting no change. This transition area matrix, capturing the magnitude of historical land use conversions, served as a crucial input for the CA-Markov model, informing the temporal dynamics guiding future predictions.

3.5 CA-Markov Model

The integrated CA-Markov approach was applied to project future patterns of land development, including a particular focus on R-3 areas in Quezon City. Multiple iterations were conducted, each using transition suitability maps derived from progressively refined sets of driving factors based on statistical significance. Figure 3 showed that the projected maps from the first three iterations remained visually identical. This consistency was attributed to the minimal spatial variation in the transition suitability maps derived from models that included statistically insignificant factors. However, the fourth iteration, which incorporated a refined set of 24 statistically significant factors, produced a noticeably different projected map.

Furthermore, a focused simulation was also conducted to specifically examine how the identified spatial and socioeconomic driving factors influence LUC within R-3 areas. A significant shift in the projected R-3 map occurred only in the fourth iteration where the simulation incorporated a refined set of 24 driving factors that were all deemed statistically significant by the preceding analysis. This finding strongly highlighted that the quality and statistical relevance of the input driving factors are critical for the CA–Markov model to produce spatially distinct and realistic LUC predictions, even when focusing on a specific class like R-3. Results are shown in Figure 4.

https://doi.org/10.5194/isprs-annals-X-5-W4-2025-205-2026

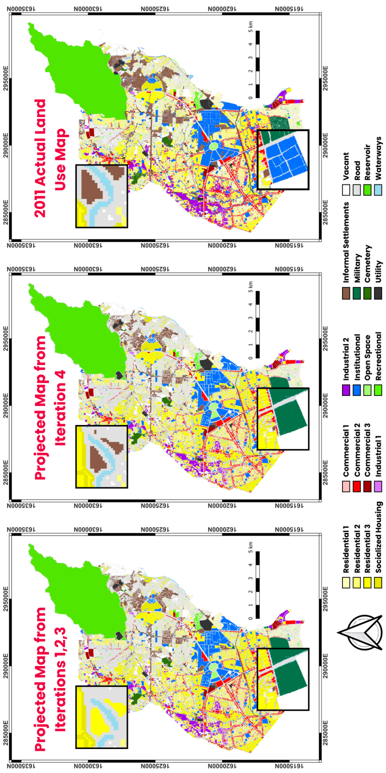

Figure 3. Comparison of Projected 2011 Maps (Iterations 1–3 and Iteration 4) with the Actual 2011 Land Use Map.

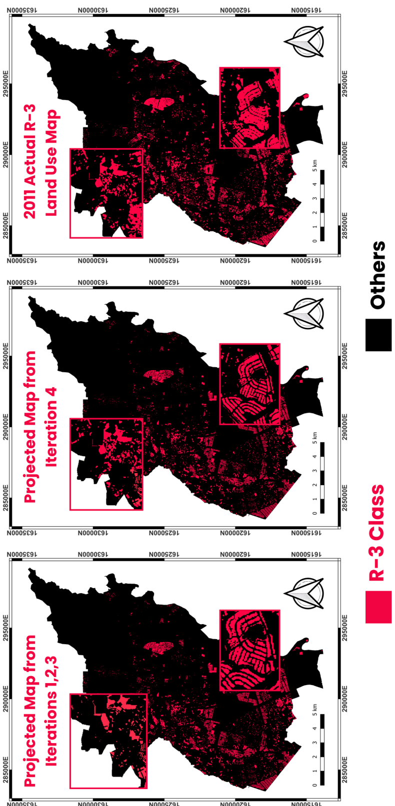

Figure 4. Comparison of Projected 2011 R-3 Maps (Iterations 1–3 and Iteration 4) with the Actual 2011 R-3 Map.

3.6 Accuracy Assessment

The accuracy assessment of the generated land use projections, encompassing all land use classes, was conducted by comparing the CA–Markov model outputs with the actual 2011 land use map, as summarized in Table 4. Iterations 1 to 3 yielded identical metric results, consistent with their identical projected outputs. In contrast, Iteration 4, generated using a refined set of 24 statistically significant driving factors, produced a slightly different map and showed marginal improvements in overall metrics results, including quantity agreement and $K_{location}$. Across iterations, the overall $K_{standard}$ remained around 0.70, indicating acceptable predictive performance at the full-class level.

For the R-3 class specifically, accuracy assessment results provided additional insight. Similar to the projected land use maps encompassing all classes, Iterations 1 to 3 exhibited identical metrics results, while Iteration 4 demonstrated slight improvements. These included higher values for total agreement and Kappa for no information (K_{no}), which exceeded 0.90, suggesting strong model performance in estimating the correct amount of R-3 land use, independent of spatial accuracy. Notably, $K_{location}$ for Iteration 4 (0.5696) was marginally higher than in previous iterations (0.5682), reflecting a modest enhancement in the spatial allocation of R-3 areas. Although the R-3-specific $K_{standard}$ (0.4947) and $K_{location}$ (0.5696) values were lower than those for all-class projections, this does not imply poor model performance. Rather, it underscores the challenge of accurately predicting a spatially sparse class like R-3, where minor misclassifications or positional errors can disproportionately affect spatial accuracy metrics.

Projected Map	All Classes		R-3 Class only	
Metrics	Value			
	Iter 1-3	Iter 4	Iter 1-3	Iter 4
AgreeGridCell	0.4976	0.4976	0.0441	0.0441
AgreeQuantity	0.2434	0.2437	0.4108	0.4109
AgreeChance	0.0476	0.0476	0.5	0.5
Total Agreement	0.7886	0.7889	0.9549	0.9550
DisagreeQuantity	0.0893	0.0917	0.0116	0.0117
DisagreeGridCell	0.1220	0.1193	0.0335	0.0333
$K_{standard}$	0.7019	0.7019	0.4946	0.4947
K_{no}	0.7781	0.7782	0.9098	0.9100
$K_{location}$	0.8031	0.8065	0.5682	0.5696

Table 4. Validation Metrics Result of Projected Land Use Maps (Encompassing All Classes vs Focused on R-3 Areas)

3.7 Prediction of Future Land Use Changes

The integrated CA-Markov approach, specifically the validated configuration from Iteration 4, which incorporated the 24 statistically significant driving factors, was utilized to simulate forthcoming land use changes in Quezon City for the year 2030. Figure 5 shows two simulation outputs generated for 2030: a general land use map encompassing all classes and a focused R-3 zone map.

The general land use projection revealed a recurring pattern of significant interplay and spatial exchange between R-3 areas, socialized housing, and informal settlements. Areas projected to continue attracting urban development, including high-density residential, were those near existing I-1, CBDs, and road networks. Conversely, areas with inherent constraints, such as proximity to cemeteries, reservoirs, or high elevation, were projected to experience minimal to no significant LUC.

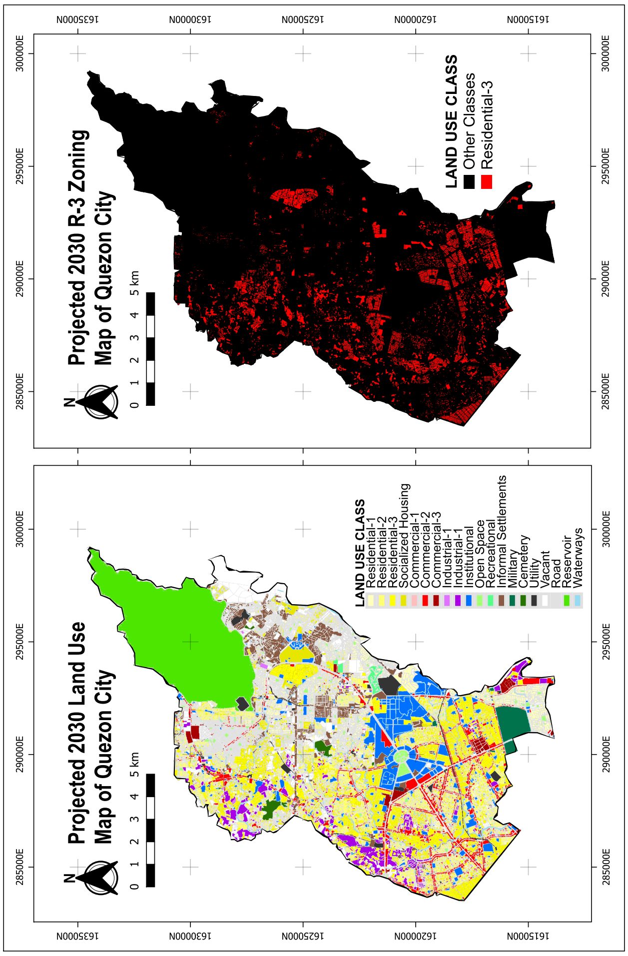

Figure 5. Quezon City's Projected 2030 (a) Land Use Map and (b) R-3 Map

The R-3 focused simulation specifically projected a noticeable increase in high-density residential zone coverage by 2030, consistent with historical trends and urban pressure. This expansion is anticipated particularly in areas adjacent to existing R-3 areas, informal settlements, and underutilized land, reflecting ongoing processes of densification and reclassification to meet housing demand. While the model shows limitations in pinpointing the exact spatial location of change for the specific R-3 class, its high accuracy in predicting the quantity or extent of R-3 change enables LGUs to use the research for strategic urban planning by identifying the likely magnitude of growth and general areas of potential R-3 development (Azabdaftari and Sunar, 2024).

These predicted patterns align with the strategic objectives outlined in Quezon City's CLUP and Development Plan, which aim to concentrate residential development around strategic urban centers and along major transportation routes to enhance accessibility and reduce sprawl. The projection also supports local plans for managing informal settlements and promoting affordable high-density housing, demonstrating the dynamic interplay between R-3, socialized housing, and informal settlement areas. Overall, these land use projections serve as a valuable decision-support tool for policymakers and planners. By identifying future hotspots of R-3 expansion, the simulation can inform proactive planning for infrastructure investment (e.g., transportation, utilities), guide housing policies (e.g., identifying areas for socialized or mid-rise projects, particularly where informal settlements and R-3 expansion overlap), and pinpoint areas potentially requiring zoning updates or additional regulation. This provides a policy-consistent and spatially grounded framework for land use planning aligned with SDG 11.

4. Conclusion and Recommendations

4.1 Conclusion

This study successfully analyzed the spatial and socioeconomic drivers of land use transitions within R-3 areas in Quezon City and projected future transformations using an integrated BLR and CA-Markov approach. By focusing on a zone-specific land use class critical to housing policy and urban planning, the research addressed a significant gap in the Philippine literature and contributed meaningful insights into urban transformation under rapid urbanization.

The findings underscore several key points. First, the intensifying urbanization in cities such as Quezon City has amplified challenges related to housing shortages and the proliferation of informal settlements. R-3 areas, designated for multi-family dwellings, are crucial in understanding and addressing these urban issues. The study demonstrated the utility and feasibility of integrating spatial and socioeconomic data to model urban transformation, especially within zoning categories like R-3, which play a central role in Philippine housing policy and urban planning. The CA–Markov model proved effective in predicting the quantity of changes in R-3 areas but encountered limitations in accurately identifying the precise spatial locations of these changes. This outcome is largely attributed to the intrinsic difficulty of modeling a spatially sparse class like R-3, where even minor locational inaccuracies can significantly penalize spatial accuracy metrics. Moreover, the aggregation of input data at the barangay level may have obscured fine-grained spatial variations, limiting the model's sensitivity to micro-scale patterns of urban change.

While some validation metrics, such as McFadden's Pseudo R^2 and location agreement accuracy indicators, appeared modest, these do not undermine the study's overall utility. The BLR model achieved an AUC of 0.78, indicating acceptable discriminatory power, and the CA–Markov model demonstrated strong performance in predicting the extent of R-3 change ($K_{no} \approx 0.91$), even if pinpointing pixel-level transitions proved challenging. These results affirm the model's practical value, particularly in predicting LUC and identifying statistically significant drivers of said changes.

Additionally, the exclusion of qualitative variables, including policies on land use planning, infrastructure development, and administrative decisions, which play a substantial role in influencing urban form, constrained the model's predictive accuracy. Moreover, the use of temporally unequal input intervals (i.e., 2000–2009 vs. 2009–2011) further complicated the model's ability to generalize change patterns and introduced potential bias in temporal projections.

Despite these limitations, the study contributes to a deeper understanding of Quezon City's urban land use dynamics and provides a critical assessment of the capabilities and constraints of CA–Markov modeling in the context of rapid urbanization and accelerated urban expansion. Ultimately, this research affirms the importance of spatially informed and data-driven approaches in urban planning and aligns with the broader objectives of SDG 11, which promotes inclusive, safe, resilient, and sustainable urban environments.

4.2 Recommendations

Building on the study's findings and limitations, several directions are recommended to improve the forecasting capacity of LUC models. First, improving the selection and calibration of driving factors is essential. This can be achieved by applying stricter statistical thresholds and testing alternative combinations of predictors during the BLR stage, which may yield more accurate suitability maps. Additionally, to address potential class imbalance in the dependent variable, applying class weighting techniques such as Synthetic Minority Over-sampling Technique (SMOTE) could be considered. These methods enhance and adjust the focus on underrepresented class and reduce bias towards the majority class. Furthermore, tuning model parameters and applying regularization techniques could be applied to avoid overestimation in the model, where it overpredicts actual changes.

Enhancing the spatial detail of input datasets, like using census block-level demographics or high-resolution hazard maps, would also allow models to better capture intra-barangay dynamics that may have been obscured by data aggregation. Besides, incorporating more temporally consistent land use maps, or extending the time range between inputs, could reduce calibration bias and improve the model's ability to generalize transition patterns. Alongside this, integrating qualitative data such as records of infrastructure development, zoning policy revisions, or issued permits can enrich the contextual understanding of land use dynamics. While these inputs may not be suited for statistical modeling, they can inform rule calibration and help interpret spatial outcomes more effectively.

Exploring alternative or hybrid modeling approaches is also encouraged. Given the CA–Markov model's reliance on current states and static neighborhood effects, it may fall short in capturing complex feedback loops and irregular development patterns. Agent-based or rule-based models could offer more realistic simulations, particularly for sparse or fragmented land uses like R-3 areas. Moreover, models that allow for mixed land uses within a single cell could provide a more nuanced representation where data resolution permits.

Finally, refining validation methods is crucial, especially for spatially sparse classes. Relying solely on area-wide metrics can obscure performance in specific areas. Implementing stratified or object-based validation approaches, particularly in areas prone to R-3 changes or adjacent to existing R-3 areas, could offer a more meaningful assessment by evaluating spatial patterns and clusters rather than strict cell-level matches. These refinements would collectively improve the model's utility in guiding urban planning and policy decisions in rapidly urbanizing contexts.

References

Azabdaftari, A., Sunar, F., 2024. Predicting urban tomorrow: CA-Markov modeling and district evolution. *Earth Sci. Inf.,* 17(4), 3215-3232. doi.org/10.1007/s12145-024-01340-4

Beroho, M., Briak, H., Cherif, E.K., Boulahfa, I., Ouallali, A., Mrabet, R., Kebede, F., Bernardino, A., Aboumaria, K., 2023. Future Scenarios of Land Use/Land Cover (LULC) Based on a CA-Markov Simulation Model: Case of a Mediterranean Watershed in Morocco. *Remote Sens.*, 15(4), 1162. doi.org/10.3390/rs15041162

Chavez, C., 2023. DILG hails QC Council for bagging NCR Outstanding Sanggunian Legislative Award. Manila Bulletin. mb.com.ph/2023/11/15/dilg-hails-qc-council-for-bagging-the-2023-ncr-outstanding-sanggunian-legislative-award (30 June 2025).

Gasirabo, A., Xi, C., Hamad, B.R., Edovia, U.D., 2023. A CA–Markov-Based Simulation and Prediction of LULC Changes over the Nyabarongo River Basin, Rwanda. *Land*, 12(9). doi.org/10.3390/land12091788

Getu, K., Gangadhara Bhat, H., 2024. Application of geospatial techniques and binary logistic regression model for analyzing driving factors of urban growth in Bahir Dar city, Ethiopia. *Heliyon*, 10(3). doi.org/10.1016/j.heliyon.2024.e25137

Hu, Z., Lo, C.P., 2007. Modeling urban growth in Atlanta using logistic regression. *Comput. Environ. Urban Syst.*, 31(6), 667–688. doi.org/10.1016/j.compenvurbsys.2006.11.001

ICLEI South Asia, 2021. Informality and Resilience: Quezon City. ICLEI South Asia. southasiaoffice.iclei.org/wp-content/uploads/2021/08/Informality-and-Resilience-Quezon-City.pdf (30 June 2025).

Kyriazos, T., Poga, M., 2023. Dealing with Multicollinearity in Factor Analysis: The Problem, Detections, and Solutions. *Open J. Stat.*, 13(03), 404–424. doi.org/10.4236/ojs.2023.133020

Long, Y., Liu, Xi, Luo, S., Luo, T., Hu, S., Zheng, Y., Shao, J., Liu, Xuejun, 2023. Evolution and Prediction of Urban Fringe Areas Based on Logistic–CA–Markov Models: The Case of Wuhan City. *Land*, 12(10). doi.org/10.3390/land12101874

Mathanraj, S., Rusli, N., Ling, G.H.T., 2021. Applicability of the ca-markov model in land-use/ land cover change prediction for urban sprawling in batticaloa municipal council, Sri Lanka. *IOP Conf. Ser.: Earth Environ. Sci.*, 620. doi.org/10.1088/1755-1315/620/1/012015

Mondal, M.S., Sharma, N., Kappas, M., Garg, P.K., 2020. Cellular automata (CA) contiguity filters impacts on CA markov modeling of land use land cover change predictions results. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci*., XLIII-B3, 1585–1591. doi.org/10.5194/isprs-archives-XLIII-B3-2020-1585-2020

Shahbazian, Z., Faramarzi, M., Rostami, N., Mahdizadeh, H., 2019. Integrating logistic regression and cellular automata–Markov models with the experts' perceptions for detecting and simulating land use changes and their driving forces. *Environ. Monit. Assess.*, 191. doi.org/10.1007/s10661-019-7555-4

Shrestha, N., 2020. Detecting Multicollinearity in Regression Analysis. *Am. J. Appl. Math. Stat.*, 8, 39–42. doi.org/10.12691/ajams-8-2-1

Taloor, A.K., Sharma, S., Parsad, G., Jasrotia, R., 2024. Land use land cover simulations using integrated CA-Markov model in the Tawi Basin of Jammu and Kashmir India. *Geosystems and Geoenvironment*, 3. doi.org/10.1016/j.geogeo.2024.100268

World Bank, 2023. Urban Population (% of total population) – Philippines. World Bank Open Data. data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS?end=2023 &locations=PH&start=2%20012 (30 June 2025).