

BUILDING-LEVEL POPULATION ESTIMATION USING LIDAR-DERIVED BUILDING VOLUME DATA

K. A. Vergara

Department of Geodetic Engineering, University of the Philippines, Diliman, Quezon City, Philippines
kpvergara@up.edu.ph

KEY WORDS: population estimation, population, building volume, census, geographic information systems

ABSTRACT:

Population censuses serve as pivotal repositories for demographic and socioeconomic information. Conducted quinquennially in the Philippines, these censuses aggregate population data into administrative units, with the barangay as the smallest unit. However, this aggregation, when used in analyses, often presumes homogeneity within these units, potentially leading to overgeneralized results. With the advent of micro-level data, such as building-specific population counts, more nuanced spatial analyses become feasible. This study leveraged existing mathematical model to estimate residential building populations in Quezon City, utilizing 3D building information derived from elevation models, building footprints, local regulations, and land use types. The study yielded promising results, achieving a normalized absolute error (NAE) of 0.133 and an R^2 value of approximately 0.976, indicating a high degree of model accuracy. However, the model also revealed systematic biases, notably underestimating populations in high-density areas and overestimating in low-density barangays. These findings underscore the complexity of factors influencing the model's performance, ranging from building digitization errors to assumptions about floor height and living area per person. The study thereby elucidates the valuable role that 3D building information can play in disaggregating population data for more granular analyses, while also highlighting areas for further model refinement to mitigate issues of overestimation and underestimation.

1. INTRODUCTION

1.1 Background of the Study

Urban environments are intrinsically dynamic systems, shaped by the complex interactions between human activities, natural elements, and built structures. Serving as vital economic nodes, these cities accommodate large populations and act as catalysts for social and economic progress. These areas are not only distinguished by high population density but also by the proliferation of built infrastructure, signifying economic development. With advancements in technology, the availability of 3D data—acquired through various means such as Light Detection and Ranging (LiDAR)—has opened new avenues for the nuanced understanding of urban landscapes (Chen, et al., 2021; Hanif, et al, 2021; Tomás, et al., 2015)

Amidst rapid urbanization, particularly in developing countries, numerous challenges related to human and infrastructure development emerge. 3D data can provide invaluable geometric and semantic information that allow for intricate spatial analyses in the built environment. To extend these analyses beyond physical structures and into human-centric considerations, it is imperative to understand population distribution and behavior. The geometric characteristics provided by 3D data can enable the estimation of population counts within specific buildings or areas (Lwin, et al., 2009; Qiu, et al., 2010; Zhao, et al., 2017).

Traditional approaches for assessing population have relied heavily on household surveys, which determine the number of individuals per household but lack spatial granularity. These data are typically aggregated to larger administrative units, limiting their utility for fine-grained analyses. To address this shortcoming, various methodologies have been employed, such as areal interpolation and statistical modeling. Areal interpolation reallocates aggregated census data to smaller administrative units and may incorporate additional data

depending on the method employed. However, the efficacy of this approach is constrained by the quality and currency of the census data available (Wardrop et al., 2018). Statistical modeling, on the other hand, uses socioeconomic variables to estimate population numbers, employing existing census data for model training. Factors such as urban area, land use, dwelling units, and image pixel characteristics are correlated to produce population estimates (Wu, et al, 2005).

Considering the limitations inherent in traditional population data, this research investigates an avenue for enhancing the granularity of demographic data: the utilization of 3D information to generate building-specific population estimates. By aligning 3D building data with current census data, the research aims to bridge the gap between broad administrative units and the intricacies of population distribution within individual structures. This strategy promises to furnish a more nuanced and detailed portrait of urban demographics, thereby contributing to the refinement of existing methods for demographic analysis.

1.2 Objectives

The primary objective of this research is to assess the efficacy of utilizing 3D building data to derive building-specific population estimates, thereby enhancing the granularity in the context of urban demographics. Recognizing the limitations inherent in traditionally published census data, which often lack fine-grained spatial resolution and are aggregated to larger administrative units, this research seeks to explore how the integration of available 3D building data—acquired through technologies such as Light Detection and Ranging (LiDAR)—can compensate for these shortcomings.

1.3 Scope and Limitations

The research is constrained by several limitations that warrant discussion. First, it relies on an existing mathematical model to generate population estimates based on building geometry, which itself is derived from building footprint areas and heights extracted from the surface models. The raw data employed in this study were sourced from various government databases, and no validation information was published for these datasets, thus the reliability of the research outcomes is contingent on the quality of the utilized data.

Furthermore, the study assumes that all digitized building footprints corresponding to residential, socialized housing, and informal settlement land uses are occupied. This assumption could introduce a degree of error into the population estimates. Due to privacy concerns, no microdata on individual buildings were available; as a result, the model relies on aggregated data and prescribed standards, potentially affecting the granularity and accuracy of the findings. These limitations should be considered when interpreting the results and could serve as avenues for future research to refine the methodological approach.

2. MATERIALS AND METHODS

2.1 Study Area

The study centers on Quezon City, Philippines, located in the northeastern part of Metro Manila and the largest city in the National Capital Region, covering 171.71 square meters. Politically, it's divided into six congressional districts and 142 barangays, the country's smallest administrative units. According to the 2015 Population and Housing Census, Quezon City had the highest population in the region with 2,936,116 people, nearly 25% of the National Capital Region's total and about 3% of the country's population. Five of its barangays rank among the ten most populous in the region.

Households are estimated to be 683,126 which give Quezon City an average household size of 4.3, slightly lower than the average household size of 4.4 of the country (Philippine Statistics Authority, 2015). According to the reported 2009 actual land use of the city, these households account for 27.43% of its total land area, and continued residential densification is seen as evidenced by construction of subdivisions and condominiums.

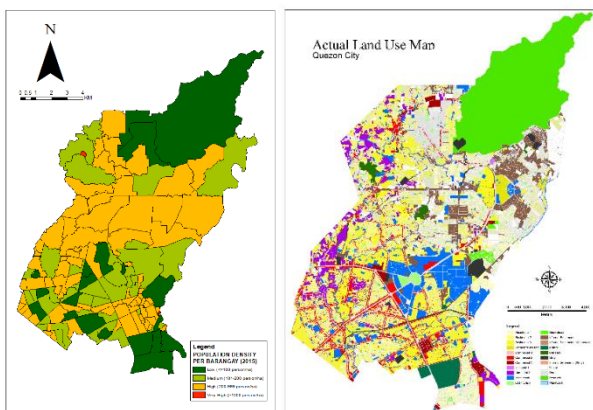


Figure 1. Population Density classification per barangay (left) Actual Land use Map (right) of Quezon City

2.2 Data Used

The data utilized in this study are summarized below. The extents of the administrative boundaries were provided by the Land Management Bureau and are coded in accordance with the Philippine Geographic Standard Code (PSGC). These codes also serve as the basis for the national census.

| Data | Year | Source |
|---|------|----------|
| 1-m resolution DTM and DSM of Quezon City | 2011 | NAMRIA |
| Administrative boundaries | 2016 | OCHA/LMB |
| Population and housing Census | 2015 | PSA |
| Actual Land Use Map | 2009 | QC LGU |
| Building footprints | 2015 | OSM |

Table 1. Data Used

2.3 General Workflow

The general workflow in the estimation and visualization of building population counts is shown in Figure 2.



Figure 2. General Workflow

2.4 Data Pre-processing

During the data pre-processing for the study, a normalized Digital Surface Model (nDSM) was derived from NAMRIA's elevation models by subtracting the Digital Terrain Model (DTM) from the Digital Surface Model (DSM). These 1-m resolution elevation models were obtained from NAMRIA as processed by Fugro Spatial Solutions from the LiDAR survey conducted in Greater Metro Manila Area (GMMA).

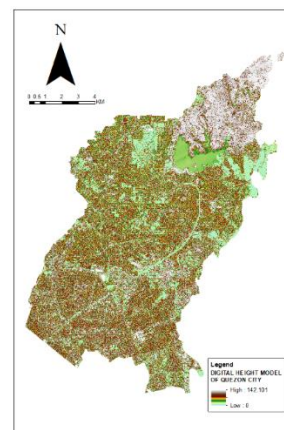


Figure 3. Normalized Digital Surface Model of Quezon City

The building footprints employed in this study were sourced from OpenStreetMap, having been digitized by an array of local volunteer mappers and subsequently verified through both aerial imagery and field validations.

Within the study area, a total of 392,836 buildings were initially identified. However, the research focus is on estimating residential building population counts, necessitating the exclusion of non-residential building footprints. This filtering process was conducted through two primary means: (1) using the attribute building-type tags available in the OpenStreetMap data, and (2) leveraging an actual land-use map provided by the

Quezon City Local Government Unit (LGU). The first method proved to be less effective, as only approximately 2% of the total building count carried such tags. Consequently, the filtering process was primarily guided by the digitized land-use map. Buildings that did not fall within the categories of Residential 1-3, Socialized Housing, and Informal Settlements were removed. As a result, a total of 377,105 building footprints were determined to be residential units.

Measurements of building dimensions were integral to the population estimation process. Specifically, the height, area, and volume of each residential unit were ascertained. The area was derived from the geometry contained in the vector files, with the assumption that these footprints represent the gross area of the respective residential units. Heights, conversely, were obtained from cell values encompassed by the building footprints. To mitigate the influence of outliers on the height values, a median zonal statistic was employed for aggregation. Utilizing these extracted area and median height values, the volumes of the buildings were subsequently calculated.

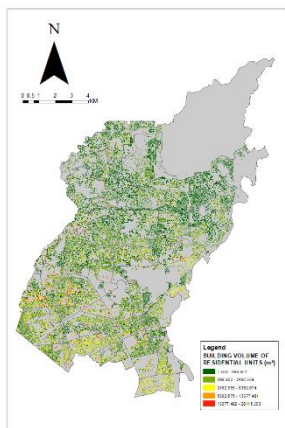


Figure 4. Building volumes of residential units in Quezon City

Population data, sourced from national household surveys conducted every five years by the Philippine Statistics Authority, are aggregated to administrative units and disclosed in summary reports. This population information, when integrated with the three-dimensional attributes of the buildings, is processed through the subsequent models to disaggregate population counts down to the individual building level.

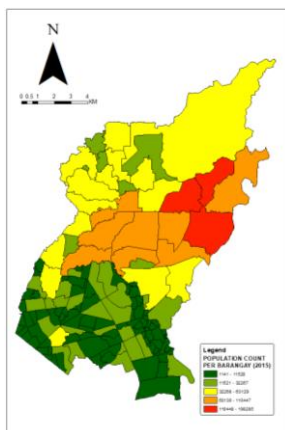


Figure 5. Population Count Per Barangay in Quezon City

2.5 Population Estimation

The mathematical relationship built by Wang, et al. (2016) was utilized to estimate the number of people residing on a dwelling unit spatially identified. The estimated population count per residential unit was computed using the equation below:

$$B_{POP\,MOD} = \frac{B_{AREA}}{B_{LABGY}} \times \frac{B_{HGT}}{B_{AHLU}} + C \quad (1)$$

Where $B_{POP\,MOD}$ is the estimated population inside the residential building, B_{AREA} the floor area of the building, and B_{HGT} the median height of the building from nDSM. Moreover, B_{LABGY} is the living area of each person in the barangay, and B_{AHLU} is the height of the floor based on the land use classification of the residential building. The constant C term is identified by comparing the statistical population sum with the published total population count to account for volume-preserving and pycnophylactic properties identified by Tobler when estimating population counts over irregular-bounded administrative units.

To refine the population estimation, unique values of B_{LABGY} for each barangay were computed.

$$B_{LABGY} = \frac{\sum B_{AREA}}{n} \quad (2)$$

Ave.household size, BGY

where B_{LABGY} living area per person of the barangay, the dividend being the average floor area of the buildings in the barangay and the divisor being the average household size of the barangay based on the published values from the 2015 national census.

Originally, B_{AHLU} shall be computed based on the floor-height relationship characteristics of residential units per barangay. However, since there is no published data on these, assumptions were made depending on the land use of the residential unit. The following B_{AHLU} values were used:

| Land Use | B_{AHLU} (m) |
|----------------|----------------|
| R1, R2, R3, SH | 3.00 |
| IS | 2.20 |

Table 2. Floor height values based on land use classification

Since the model may yield float values of the building population count, in the interest of identifying people as integer units, population counts were rounded down to the nearest integer.

2.6 Assessment

To evaluate the accuracy of the implemented models, several classical performance metrics were utilized:

- RE (Relative Error): Measures the percentage error between the model's population count and the actual values. Barangays are categorized as either poorly overestimated or underestimated if the values exceed or are below 30%, respectively.
- R^2 : Assesses the correlation between model estimates and census values, ranging from 0 to 1. A value of 0 indicates no correlation between the model and actual counts.
- TAE (Total Absolute Error): Calculates the sum of all absolute differences between the model's estimates and census data.

- NAE (Normalized Absolute Error): Transforms the TAE into a relative metric by expressing it as a percentage of the total population.

It's important to note that while public use files containing microdata on household populations are available upon request, these files lack building-level spatial information to safeguard privacy concerns. The absence of ground truth data at the building level necessitates a different approach for validation. Specifically, the absence is largely due to privacy constraints that preclude the collection and publication of such granular spatial data and the considerable manpower and financial resources required to gather such granular information across an entire city.

As a result, the evaluation relies on a comparison between the aggregated estimated counts per barangay and the corresponding published population counts. The rationale for this lies in the presumption that if the model accurately predicts population at the aggregate barangay level, its estimations at the more granular building level are likely to exhibit similar accuracy.

3. RESULTS AND DISCUSSION

3.1 Model Performance

Equations 1 and 2 were applied to all delineated residential structures to compute their respective population counts. These computational estimates were subsequently aggregated at the barangay level and juxtaposed with officially published population figures. Summarized in Table 3 are key performance metrics including the R^2 value, Total Absolute Error (TAE), and Normalized Absolute Error (NAE), which collectively offer a comprehensive assessment of the model's accuracy and reliability.

| Metric | Value |
|--------|---------|
| R^2 | 0.976 |
| TAE | 389,587 |
| NAE | 0.133 |

Table 3. Frequency distribution of relative errors (RE)



Figure 6. Actual vs. model population

The R^2 value of approximately 0.976 suggests a strong linear relationship between the model-estimated and actual population counts, accounting for nearly 97.6% of the variance in the data. This high degree of fit is reinforced by the generally linear trend observed in the scatter plot, where most points closely align with the red line of best fit. However, the scatter plot also reveals some spread around the fitting line, particularly at higher

population counts, indicating variability in the model's performance. Additionally, a few outliers are visible, deviating notably from the fitting line.

The Total Absolute Error (TAE) of 389,587 and the Normalized Absolute Error (NAE) of 13.3% introduce a level of caution in the interpretation of the R^2 value. The TAE, being a summative metric, indicates that the model can have significant discrepancies in specific instances, while the NAE reveals that the model's estimates deviate from the actual values by an average of 13.3%.

The metrics collectively suggest areas for further refinement. The TAE and NAE, although moderate, are significant enough to merit attention, especially if the model is to be deployed in policy-making scenarios that require high precision. Furthermore, the model's performance appears to be variable across different barangays, as evidenced by the spread in residuals and the presence of outliers. This suggests that the model might need to be adapted or fine-tuned to cater to different types of residential areas or varying population densities.

Additionally, a systematic bias toward underestimation was observed in the model. This bias may not merely be a statistical anomaly but could reflect inherent limitations in the model's ability to capture complex socioeconomic contexts. Specifically, the model may tend to underestimate population counts in areas with smaller footprints and/or lower building heights suggests that it may not fully account for scenarios where such areas are densely populated. Similarly, areas with larger footprints and/or higher buildings may not necessarily house more individuals, yet the model's systematic overestimation could lead to inaccurate predictions for these areas as well.

This limitation highlights a fundamental challenge in using primarily geometric and policy-based variables for estimating population. Such variables may not capture the full spectrum of factors influencing population density, such as historical patterns of settlement, cultural factors, or economic conditions that can lead to overcrowding or underutilization of space in different types of buildings and areas. In essence, the model may lack the nuance to understand that smaller structures might be accommodating more people due to economic necessities or cultural preferences, just as larger, taller buildings may not be as populated as their size would suggest, possibly due to economic affluence or different land use policies.

3.2 Relative Errors

To comprehend the nuances of the distribution of relative errors across barangays, both the spatial distribution and the relationship of these errors to each barangay's population density were investigated. Relative errors (RE) between the model-generated estimates and the authoritative population counts for each barangay were calculated to serve this purpose. The frequency distribution of these relative errors is presented in the subsequent analysis:

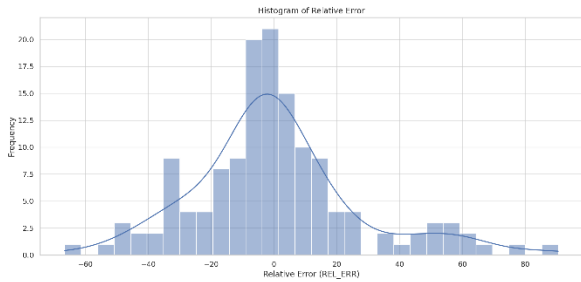


Figure 7. Frequency distribution of relative errors (RE)

Positive relative errors signify overestimations of the actual population, while negative values indicate underestimations. The computed relative errors span a wide range, from -67.67% to 87.30%, with a standard deviation of about 26.18%, suggesting substantial variability in the model's accuracy across different barangays. Notably, a median error of -2.05% and a 25th percentile of -11.97% point to a systematic tendency of the model to underestimate population counts in a majority (82 out of 142) of barangays.

To facilitate a more nuanced understanding, the barangays were segmented into four categories based on their relative errors: (1) poor underestimation, with errors less than -30%; (2) underestimation, with errors ranging from -30% to 0%; (3) overestimation, with errors from 0% to 30%; and (4) poor overestimation, with errors greater than 30%. This classification was adopted from Wang, et al (2016).

Based on the analysis of relative errors (RE), it was found that 16 out of 142 barangays fell into the category of being poorly underestimated. These barangays are geographically dispersed across the city and do not exhibit any discernible spatial clustering. Intriguingly, most of these poorly underestimated barangays are among those with the smallest land areas when ranked accordingly. Conversely, 12% of the barangays, amounting to 17 out of 142, were identified as poorly overestimated. These are predominantly located in the southern part of the city, an area characterized by a high concentration of commercial and low-density residential land use types.

To further explain the discrepancies in population estimation, the relationship between population density and these relative errors was explored. Figure 8 presents a detailed view of each barangay's RE distribution in conjunction with their population density classification.

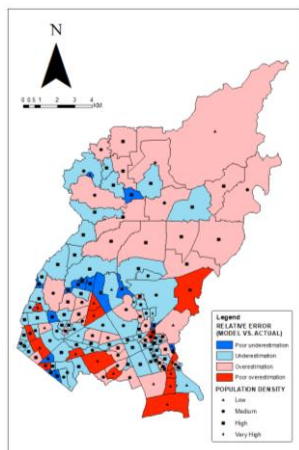


Figure 8. Spatial Distribution of RE

Figure 9, on the other hand, provides a scatter plot illustrating the correlation between barangay population density and their corresponding RE values. These visual representations aim to offer additional insights into the factors contributing to the levels of underestimation or overestimation observed across different barangays.

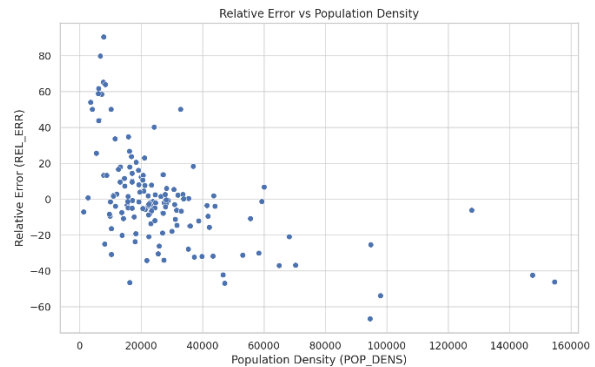


Figure 9. RE vs. population density

Upon examination, poor underestimation predominantly occurs in barangays with high population density, such as Capri, Escopa I, Escopa III, Escopa II, Botocan, and San Isidro. An exception to this trend is Escopa IV, which exhibited a modest underestimation of only 7%.

Several factors could contribute to this observed underestimation in densely populated areas. First, errors in building digitization are more likely in these areas due to the challenges associated with distinguishing individual households in proximity. Second, the use of OpenStreetMap (OSM) building footprints, which are often digitized based on satellite imagery, may result in multiple households being generalized into a single footprint. This would artificially inflate the estimated living area per person, leading to a lower estimated population count. Third, these areas predominantly consist of socialized housing, where the assumed ceiling height of 3.0 meters may be an overestimation. As per BP 220 regulations, the minimum prescribed ceiling height is only 2 meters, making the model's assumption potentially inaccurate for these specific housing types. This overestimation of ceiling height would also contribute to a higher estimated living volume per person, thereby underestimating the building population.

Conversely, poor overestimation is more prevalent in barangays with low population density, ranging from 1 to 100 people per hectare, such as Ugong Norte, Kalusugan, Phil-Am, Blue Ridge A, Lourdes, Mariana, Horseshoe, Saint Ignatius, Damar, and White Plains. In these areas, households are typically spaced more evenly, reducing the likelihood of errors in building digitization. However, these households are often part of subdivisions composed of larger units, both in terms of floor area and ceiling height, compared to the average household size. Despite the prevalence of subdivisions in these barangays, a lack of homogeneity in the area and volume of residential buildings was observed, as indicated by high standard deviations in these measurements. This variability is likely to contribute to the overestimation of population counts in buildings with smaller volumes.

The absence of a clear, overarching geographic pattern in the distribution of relative errors highlights the considerable impact of local factors on the model's performance. This observation strengthens the argument for adopting a more localized approach to model refinement. Such an approach would not only consider the specific physical characteristics of each barangay, such as

population density and building types, but also extend to include socioeconomic factors that have hitherto been absent from the model. These could encompass variables like income levels, employment rates, and educational attainment, which may have an indirect but significant influence on population density and distribution.

Incorporating these socioeconomic factors could offer a more nuanced understanding of the local variations in population density, thereby potentially enhancing the model's accuracy. For instance, lower-income areas might have higher occupancy rates per household, or areas with higher educational attainment might have different residential preferences that affect population density. By considering such factors, the model would be better equipped to capture the complex interplay of variables that contribute to the observed relative errors in population estimates across different barangays.

3.3 Visualization

For enhanced visualization and analysis, a 3D city model incorporating residential buildings and additional data layers was constructed using ArcScene software. This model effectively illustrates the spatial distribution of residential building population counts throughout the city. Aligning with the established relationship between building volume and population count, it is evident that residential buildings with larger footprints and multiple floors tend to house larger populations. The use of advanced analytical tools and query capabilities within the software enables the generation of insights into issues that necessitate micro-level data. Presented below are select visualizations from this 3D city model, with a focus on depicting the phenomenon of building population counts across different areas.

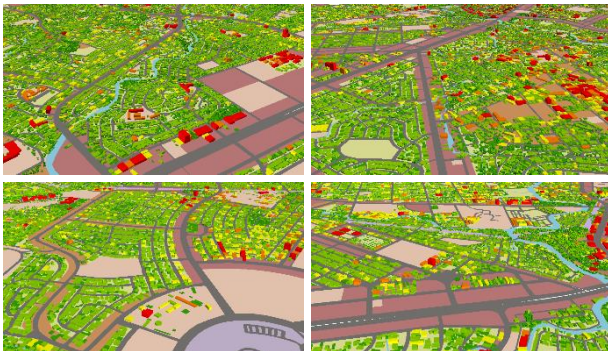


Figure 10. Building population distribution in Quezon City

4. CONCLUSION AND RECOMMENDATIONS

The primary objective of this research was to evaluate the effectiveness of leveraging 3D building data for generating building-specific population estimates, aiming to enhance the level of granularity in urban demographic analyses. The research addressed the limitations of traditional census data, which often lack fine-grained spatial resolution and are confined to larger administrative units. By integrating 3D building data, acquired through technologies such as Light Detection and Ranging (LiDAR), the study sought to provide more precise, localized population estimates.

The results indicate a high degree of overall accuracy in the model, as evidenced by an R^2 value of approximately 0.976. However, the model also manifested systematic biases, notably

a tendency to underestimate population counts in high-density areas and overestimate in low-density areas. The relative errors, varying in magnitude across different barangays, highlight the complexity of factors influencing the model's performance. These include building digitization errors, assumptions related to floor height, and living area per person, among others.

The absence of a clear geographic pattern in the distribution of relative errors underscores the importance of incorporating local factors into the model, extending beyond mere physical attributes to include socioeconomic variables. This more localized and nuanced approach would likely improve the model's performance, particularly in addressing issues of overestimation and underestimation.

To this end, future refinements to the model should aim to incorporate these local factors, including socioeconomic variables that could provide a more comprehensive understanding of population distribution and density. The model's performance can be further enhanced by directly acquiring microdata samples from all administrative units. These samples could inform more accurate estimates of the living area per person and floor heights, which are critical parameters in the current model. Moreover, incorporating microdata on socioeconomic and other demographic factors could offer additional layers of nuance, thereby improving the model's predictive accuracy. Additionally, even though the analysis was conducted at the level of the smallest administrative unit, barangays could be further disaggregated into smaller blocks. This would allow for a more nuanced understanding of the heterogeneity in building characteristics within each barangay, thereby enabling multiple spatial analyses on a micro-level that could provide more insightful solutions to the problems being addressed.

Overall, while the model presents a promising avenue for enhancing the granularity of urban demographic data, it also reveals areas requiring further investigation and refinement. Through targeted improvements, including the integration of microdata and further disaggregation of administrative units, the model has the potential to significantly transform the utility and accuracy of 3D data in generating precise, building-specific population estimates, thereby fulfilling the research's primary objective.

REFERENCES

- Chen, H., Wu, B., Yu, B., Chen, Z., Wu, Q., Lian, T., Wang, C., Li, Q., & Wu, J., 2021: A New Method for Building-Level Population Estimation by Integrating LiDAR, Nighttime Light, and POI Data. *Journal of Remote Sensing* (Vol. 2021). <https://doi.org/10.34133/2021/9803796>
- Clarke, D., Kerr, D., Bengtsson, L., Juran, S., Seaman, V., & Tatem, A. J., 2018: Spatially disaggregated population estimates in the absence of national population and housing census data. *Proceedings of the National Academy of Sciences* (Vol. 115, Issue 14, pp. 3529–3537). <https://doi.org/10.1073/pnas.1715305115>
- Hanif, Y. F., Handayani, H. H., & Nurwatic, 2021: Residential Population Estimation in Small-Area using LiDAR and Aerial Photograph Data. *IOP Conference Series: Earth and Environmental Science* (Vol. 731, Issue 1, p. 012034). <https://doi.org/10.1088/1755-1315/731/1/012034>

Lwin, K., & Murayama, Y., 2009: A GIS Approach to Estimation of Building Population for Micro-spatial Analysis. *Transactions in GIS* (Vol. 13, Issue 4, pp. 401–414). <https://doi.org/10.1111/j.1467-9671.2009.01171.x>

Quezon City government official website. Gov.ph. [accessed 2021 June 30]. <https://quezoncity.gov.ph/about-the-city-government/>

Qiu, F., Sridharan, H., & Chun, Y., 2010: Spatial Autoregressive Model for Population Estimation at the Census Block Level Using LIDAR-derived Building Volume Information. *Cartography and Geographic Information Science* (Vol. 37, Issue 3, pp. 239–257). <https://doi.org/10.1559/152304010792194949>

Tobler, W. R., 1979: Smooth Pycnophylactic Interpolation for Geographical Regions. *Journal of the American Statistical Association* (Vol. 74, Issue 367, pp. 519–530). <https://doi.org/10.1080/01621459.1979.10481647>

Tomás, L., Fonseca, L., Almeida, C., Leonardi, F., & Pereira, M., 2015: Urban population estimation based on residential buildings volume using IKONOS-2 images and lidar data. *International Journal of Remote Sensing* (Vol. 37, Issue sup1, pp. 1–28). <https://doi.org/10.1080/01431161.2015.1121301>

Wang, S., Tian, Y., Zhou, Y., Liu, W., & Lin, C., 2016: Fine-Scale Population Estimation by 3D Reconstruction of Urban Residential Buildings. *Sensors* (Vol. 16, Issue 10, p. 1755). <https://doi.org/10.3390/s16101755>

Wardrop, N. A., Jochem, W. C., Bird, T. J., Chamberlain, H. R., Wu, S., Qiu, X., & Wang, L., 2005: Population Estimation Methods in GIS and Remote Sensing: A Review. *GIScience & Remote Sensing* (Vol. 42, Issue 1, pp. 80–96). <https://doi.org/10.2747/1548-1603.42.1.80>

Zhao, Y., Ovando-Montejo, G. A., Frazier, A. E., Mathews, A. J., Flynn, K. C., & Ellis, E. A., 2017: Estimating work and home population using lidar-derived building volumes. *International Journal of Remote Sensing* (Vol. 38, Issue 4, pp. 1180–1196). <https://doi.org/10.1080/01431161.2017.1280634>