

Engineering and Applied Science Research

 https://www.tci-thaijo.org/index.php/easr/index Published by the Faculty of Engineering, Khon Kaen University, Thailand

Development of small area population estimation models for a developing, densely populated metropolitan area and its applications: A case study of Metro Manila

Monorom Rith^{*1, 2}, Jimwell L. Soliman¹, Neil Stephen A. Lopez¹, Alexis M. Fillone³⁾ and Jose Bienvenido M. Biona^{1,4)}

1)Mechanical Engineering Department, De La Salle University, 2401 Taft Ave, Metro Manila, Philippines ²⁾Research and Innovation Center, Institute of Technology of Cambodia, Russian Conf, Norodom Blvd, Phnom Penh, Cambodia

3)Civil Engineering Department, De La Salle University, 2401 Taft Avenue, Metro Manila, Philippines 4)Enrique K. Razon Jr. Logistics Institute, De La Salle University, LTI Spine Road, Laguna Blvd, Biñan, Laguna 4024, Philippines

> Received 19 August 2019 Revised 29 November 2019 Accepted 3 December 2019

Abstract

Projection of population in a small area is essential for the government to design proactive policies to support a variety of planning processes and making decisions. The private sector can use this information to do customer demand forecasting and market site targets on a small scale. However, most developing countries do not have this level of data for the development of small area population estimation models. The thrust of this study is to develop a linear regression-based small area population estimation model using recent census data of Metro Manila. The R² values of the developed population and household estimation models are 0.975 and 0.994, respectively, while the respective mean absolute errors (MAEs) are 9.76% and 7.98%. The developed models were then applied to project a small area population. The area of Metro Manila with a population density of more than 50,000 persons/km² will increase from 5.78% in 2010 to 9.23% in 2020, 15.14% in 2030, 21.76% in 2040, and 31.31% in 2050. The projected population within Metro Manila will increase from 11.89 million in 2010 to 29.16 million in 2050, with an average annual growth rate of 3.63% from 2010 to 2050. During this time, the population density will rise from 19,137 persons/ km^2 in 2010 to 49,243 persons/ km^2 in 2050. The total number of households is projected to increase from 2.89 million in 2010 to 7.49 million in 2050, which is a 2.59-fold increase.

Keywords: Demographic model, Population projection, Linear regression, Metro Manila, Philippines

1. Introduction

The Philippines is a fast-growing economy, located in the southeastern of Asia. It has had an average annual growth rate of 6.6% during the last three years [1]. Metro Manila, the national capital region (NCR) of the Philippines, consists of sixteen cities and one municipality (see Figure 1), with a total area of 619.54 km². The numbers on the map of Metro Manila in Figure 1 are the codes of traffic analysis zones (TAZ) [2]. The population of Metro Manila was 11.856 million persons in 2010 and increased to 12.877 million persons in 2015, with an average annual increase of 1.722% [3]. The population density was 20,785 persons/km² in 2015, with a density increase of 62% over that in 1990 [3]. However, the projection of Metro Manila's population at a small area level has remained unexplored. Extrapolation of a small area population is essential to support a variety of the local planning processes and decision making by the government and private sector. The government must know the small area population growth beforehand to make proactive local policies and appropriately plan to respond to disasters and distribute sufficient public services and resources to the residents, such as building and modernization of hospitals and schools and development of open spaces, recreational areas, and other public facilities [4- 5]. Furthermore, the regional water and electricity authorities can forecast future demand at the small area level only if their populations are accurately projected. For the transport and infrastructure sector, engineers and researchers can use small-area population projections to do microsimulations of trip generation and attractions from one small area to another to estimate the travel demands and support public transportation policy. The projected population in a small area is also indispensable for the private sector to do customer demand forecasting and market-area delineation.

Various methods have been proposed to develop small area demographic models in previous studies. They include the Component Method II (CM II), the Administrative

Figure 1 The map of Metro Manila (adopted from [2])

Record (AR) method, the Ratio Correlation (RC) method, the Housing Unit (HU) method, and Remote-Sensing and Geographic Information System (GIS) method [5-7]. The CM II method is used to compute the small area-level residents by adding the total of new births, subtracting the number of deaths, and adding the net migration and the net group-quarter populations [5, 7]. This method is best suited for regional population projections and provides an accurate model estimation utilizing recent census data [7]. However, most of the developing metropolitan area does not have official data of net migration at the small area level. The AR method estimates the population based on federal income tax returns. However, taxpayers are typically aged under 60 years in developing countries. The RC method uses school enrollment data, car registrations, the total employment opportunities, and the number of occupied housing units. The HU method multiplies the number of dwelling units by the average household size and adds the group-quarter population [8]. This method is widely applied in the US because of its accuracy and cost-effectiveness but needs building permit data and electric customer information [5, 8- 9]. However, the RC and HU methods cannot generally be applied in the developing metropolitan areas due to limited data on occupied housing units. Remote-sensing and GIS methods or satellite imaginary are relatively accurate and not influenced by census data and population registers [10-11], but the approach is arduous requiring a large amount of time and labor since this technique uses high-resolution aerial photographs to manually count dwellings [5]. A recent study used the Ordinary Least Square (OLS) method to develop a small area population estimation model for depopulating

cities in the US [4]. It employed a small area data of housing units, occupied units, and vacant unit changes, among other information. Each method has its merits and drawbacks. Method selection is assessed based on validity, timeliness, cost, data acquisition, and the purpose of the application, among other factors [12]. Overall, all the methods used in the previous studies are well developed and applied in the developed world. However, the data required at the small area level for those models are probably not available in the developing world.

The current study intends to formulate small area population estimation models for a developing metropolis with a high population density. The models do not use data concerning building permits, school enrollment, car registrations, federal income tax returns, net migration, net group-quarter population, or high-resolution aerial imagery. A case study of Metro Manila was adopted in our study. We used the 2010 and 2015 census data obtained from the Philippine Statistics Authority (PSA). The Traffic Analysis Zone (TAZ) level subdivided by [2] was used as a small scale to represent the smallest areas. Linear regression was applied to develop a TAZ level demographic model. The findings of this study are highly expected to have a significant contribution to practical development by the Philippine government in designing proactive local policies to fulfill the residential demand and by the private sector for market demand forecasting at the TAZ level. Furthermore, the methodology adopted in the study is informative for developing countries with data acquisition difficulties at the small area level.

Figure 2 Population density distribution: 2010 (left) and 2015 (right)

Figure 3 Population distribution: 2010 (left) and 2015 (right)

The remainder of the paper is structured as follows. Section 2 presents data availability and preparation as well as methodology. Section 3 shows the model estimation results and applications of the developed models. Section 4 briefly describes the findings and directions for practical applications as well as future research.

2. Data sources and methodology

2.1 Data sources

The PSA carried out a census in both 2010 and 2015. The population and household data are grouped based on the Barangay level (i.e., community or the smallest administrative division in the Philippines) as the smallest area level. However, the area is extremely varied from one Barangay to another, and some Barangays have no data about their area. The authors decided to use the TAZ level subdivided by [2] as a small area basis of population estimation because the differences between the areas of TAZ levels are smaller than those of the Barangays. MUCEP [2] lists the TAZ ID corresponding to the Barangay names. The area of TAZ-level was previously computed [2].

The authors rearranged the household and population data at the Barangay level into the TAZs. Household population data were used to calculate the number of persons

Figure 4 Distribution of the number of households: 2010 (left) and 2015 (right)

per household (PPH). The distributions of the population density, the population, and the number of households at the TAZ level are illustrated in Figures 2, 3, and 4, respectively. As can be seen from Figure 2, the proportion of Metro Manila with a population density of at least 50,000 persons/km² was 5.78% in 2010. This increased to 6.67% in 2015. Most of the TAZs of at least $50,000$ persons/km² are located in Manila city (the oldest city of Metro Manila).

The TAZs without residents illustrated in Figures 3 and 4 are reclamation, shopping, and airport areas. There are only 250 TAZs where residents live. The distributions of population and the number of households slightly varies from one TAZ to another. There were 21 TAZs with populations of more than 100,000 persons in 2010. This increased to 30 in 2015 (depicted in Figure 3). As can be seen from Figure 4, the number of TAZs having a total of more than 20,000 households increased from 36 in 2010 to 43 in 2015.

2.2 Small area population estimation model

Based on the CM II method, the current census population has a significant effect on the population estimation model. The area and PPH have correlations with the HU-based demographic model [5, 8], and the PPH is not stable [13]. Consequently, we assumed that the TAZ levelpopulation estimation is a function of the population in the previous year, the TAZ area, and PPH. Regression models are typically applied to estimate the parameters of the small area population model [5, 10, 14]. The model used in the current study is expressed as Equation 1:

Population
$$
_{z,y} = \alpha_0 + \alpha_1
$$
 Population $_{z,y-5} + \alpha_2$ Area_z + α_3 PPH_{z,y}

(1)

where the indices *z* and *y* refer to the projected zone and year, respectively, *Area* is the TAZ-level area, and α_0 , α_1 , α_2 , and α_3 are the parameter estimates. PPH refers to persons per household. The average household size was 4.3 in 2010 and 4.1 in 2015 [15]. Consequently, the average household size declined by 0.2 persons over these five years. We assumed that the PPH decreases by 0.2 persons for each TAZ.

The total number of households is estimated using Equation 2:

Household
$$
_{z,y} = \beta_0 + \beta_1
$$
 Population_{z,y} + β_2 DG_{z,y} +
\n β_3 PPH_{z,y} (2)

where Household is the total number of households, and *DG* is the density growth from year *y-5* to year *y. DG* can be calculated using the following equation:

$$
DG_{z,y} = \frac{Population_y - Population_{y-5}}{Area}
$$
 (3)

2.3 Correlation

The Pearson product-moment correlation approach was applied to estimate the correlation coefficients among the independent and output variables. In the case that some independent variables are highly correlated, only the independent variable having the highest correlation with the output variables is chosen. This is a good way to control the multicollinearity problem in interpreting the model estimation results. Table 1 presents the correlation coefficients among the independent variables. The highest correlation coefficient in absolute value between the independent variables (i.e., DG and PPH) is – 0.289, which implies no multicollinearity in the model estimation results.

3. Model estimation and application results

3.1 Model estimation results

Table 1 Correlation coefficients

Table 2 Model estimation results – coefficient (standard error)

Significant codes $(***),$ $(*),$ $(*),$ $(*),$ and $(.)$ are significant at the 0.001, 0.01, 0.05, and 0.1 levels, respectively.

Figure 5 Projection of population density at the TAZ level

The model estimation results are shown in Table 2. Based on the parameter estimates in this table, the developed models for population and total household estimation are expressed as Equations 4 and 5, respectively. The \mathbb{R}^2 and adjusted \mathbb{R}^2 are 0.975 and 0.975, respectively, for Equation 4

with a mean absolute error (MAE) of 9.76% for the year 2015. For Equation 5, the \mathbb{R}^2 and adjusted \mathbb{R}^2 are 0.994 and 0.994, respectively, with an MAE of 7.98% for the year 2015. The R^2 of this study is higher than that of the HU method $(R^2 = 0.718)$ developed by [5] and the adjusted

Figure 6 Projection of population at the TAZ level

 $R^2 = 0.22$ of [4]. The MAE of the population model is slightly higher than the MAE of the HU method (MAE = 8.70%), but the MAE of the household model is marginally lower than that of the HU method, compared to the study by [5]. The \mathbb{R}^2 and adjusted \mathbb{R}^2 are equal for both of the models because all the parameter estimates are significant at the 0.05 level, except the density growth of the total household estimation model. The intercept coefficients have no interpretable meaning, and they are included to improve the model fit and model estimation accuracy. For the population estimation model (see the first two columns), the coefficients of the population in year $y - 5$ and area are positive, which implies that a zone with more population in year $y - 5$ and larger area is associated with a higher population in year *y*. Similarly, for the total household estimation model (see the last two columns), the total number of households increases with the population and density growth. It is not surprising to see that the coefficients of PPH for both of the developed models are negative (see the sixth row) because the PPH declined from 4.3 in 2010 to 4.1 in 2015 on average in Metro Manila, while the population increased [15]. A decrease in

PPH might be due to the preference of start-up young families to separate from their extended families and live in their own houses, therefore increasing the number of households and decreasing the PPH.

Population
$$
_{z,y}
$$
 = 14,230 + 1.084 × Population $_{z,y-5}$ +
632 × Area_z - 3,798 × PPH_{z,y} (4)

Household
$$
{z,y} = 7,780 + 0.237 \times Population{z,y} + 0.122 \times DG_{z,y} - 1,860 \times PPH_{z,y}
$$

(5)

3.2 Model applications

Figure 5 visualizes the spatial distribution of projected population density. The area of Metro Manila with a population density of more than 50,000 persons/km² will increase from 5.78% in 2010 to 9.23% in 2020, 15.14% in 2030, 21.76% in 2040, and 31.31% in 2050. The majority of

Figure 7 Projection of total households at the TAZ level

Table 3 Distribution of the population density		
---	--	--

the metropolis with a high population density of more than $50,000$ persons/ km^2 is located in Manila city, the core of Metro Manila, for all the projected years.

There are 250 TAZs where residents are living. Figure 6 illustrates the spatial distribution of the projected population at the TAZ level. The number of TAZs with a population of more than 100,000 persons will sharply increase from 21 zones in 2010 to 34 zones in 2020, 56 zones in 2030, 86 zones in 2040, and 119 zones in 2050. The majority of TAZs with a population of more than 100,000 persons are located in the eastern, southern, and northern parts of Metro Manila.

Figure 7 presents the spatial distribution of the number of households by the TAZ level. Similar to the population projection, the projected TAZs with more than 20,000 households are located in the eastern, southern, and northern parts of Metro Manila. The number of TAZs with more than 20,000 households will considerably rise from 36 zones in 2010 to 52 zones in 2020, 86 zones in 2030, 122 zones in 2040, and 164 zones in 2050.

No resident | < 20,000
| 20,000 - 40,000 $140,000 - 60,000$ $\begin{array}{c} 60,000 - 80,000 \\ 80,000 - 100,000 \end{array}$ >100.000

The projected population and the total number of households within Metro Manila are illustrated in Figure 8. The population will exponentially increase from 11.89 million in 2010 to 29.16 million in 2050, with an average annual growth rate of 3.63%. The total number of households is projected to increase from 2.89 million in 2010 to 7.49 million in 2050, which translates to an increase of 2.59-fold.

Figure 8 Projection of population and the total number of households

Table 5 Distribution of the total households

The population density will increase to 49,243 persons/km² in 2050 from 19,137 persons/km² in 2010.

The distribution of the number of TAZs for the population density, population, and total households are summarized in Tables 3, 4, and 5, respectively. In Table 3, the proportion of TAZs with the population density of fewer than $50,000$ persons/km² was 84.40% in 2010, and this figure will decline to 40.00% in 2050. In 2050, the projected proportion of areas with 50,000-100,000 persons/km², 100,000-150,000 persons/km² , and > 150,000 persons/km² will be 32.80%, 17.20%, and 10.00%, respectively. From Table 4, the proportion of TAZs with a population of fewer than 100,000 persons was 91.60% in 2010. This will decline to 86.40% in 2020, 77.60% in 2030, 65.60% in 2040, and 52.40% in 2050. As can be seen from Table 5, the proportion of TAZs with fewer than 20,000 households was 85.60% in 2010. This will decline to 34.4% in 2050. TAZs with 20,000- 40,000 households account for the largest share, 43.20%, in 2050.

3.3 Relative error

The relative errors of population and total household estimates are illustrated in Figures 9 and 10, respectively. The relative error is calculated based on an estimated value minus the actual value, divided by the actual value. Very few areas have relative errors higher than 50% and lower than -50%. Most of them are located in the lower quartile representing 25% of the data set. Two proposed options to address such issues are: 1) to set constraints (e.g., lower limit and upper limit) on population change for the 25% of the data set located in the first quartile and 2) classify all TAZs into two different groups (i.e., the 1st quartile for one group and the $2nd$, $3rd$ and $4th$ quartiles for the other group), and then develop small area population estimates for each group.

Table 6 lists the number of TAZs by relative error classifications. Columns 2 and 3 refer to the small area population estimation model. It highlights that 71 TAZs or 28.40% of all the TAZs had relative errors higher than 10% and lower than – 10%. The last two columns present the small area total household estimates. It is apparent that 81 TAZs (32.40%) had relative errors higher than 10% and lower than $-10%$.

Figure 9 Relative errors of population estimates for the year 2015

Figure 10 Relative errors of total household estimates for the year 2015

4. Conclusions and recommendations

The thrust of the current study is to develop small area population and total household estimation models for a developing, densely populated metropolis using linear regression. A case study of Metro Manila is featured. The findings are as follows.

 $R²$ values are 0.975 for the population estimation model and 0.994 for the total household estimation model, while the respective mean absolute errors (MAEs) were 9.76% and 7.98%. The developed models were applied, and the results highlight that Metro Manila areas with a population density of more than 50,000 persons/km² will increase from 5.78% in 2010 to 31.31% in 2050. The proportion of TAZs with a population of fewer than 100,000 persons was 91.60% in 2010 and will markedly decline to 52.40% in 2050. The proportion of TAZs with fewer than 20,000 households was 85.60% in 2010, and this will decline to 34.4% in 2050. The population was 11.89 million in 2010, and it will increase to 29.16 million in 2050, with an average annual growth rate of 3.63%. The total number of households was 2.89 million in 2010 and will increase by 2.59 times to 7.49 million in 2050. The average population density will increase to 49,243 persons/km² in 2050 from 19,137 persons/km² in 2010.

The developed models can also be applied to project the population pyramid by gender, age, and income, among other factors. The potential extension of the small area demographic models of Metro Manila can be applied to simulate and estimate small area private car numbers using the household vehicle ownership model developed by [16], small area vehicular energy consumption and $CO₂$ emissions using the joint household vehicle ownership and energy consumption model developed by [17] and the joint individual vehicle type choice and usage model developed

by [18], and small area vehicle sales and scrapped vehicles using the vehicle survivorship model formulated by [19].

Even though the R^2 values of the models are close to one, the developed models can be applied for projection only. They are not a simulation to study the impact of policy changes. Future work should develop the models to include urban form attributes (e.g., shortest distance from the center point of a TAZ to the nearest train station, multi-criteria accessibility, employment opportunities, and presence of critical facilities). Developed models in terms of the urban form attributes can be applied to simulate the population and total households for a given change in some urban form attributes of interest at a small area level. Particularly, small area population estimation models can be applied to simulate the population based on transit-oriented development planning. The other drawback of the developed models is that unreasonably high and low projections were found in TAZs with low populations. These are areas located in the first quartile representing 25% of the data set. One way to remedy this issue is to set constraints (e.g., lower and upper limits) on population change, specifically for the 25% of the data set located in the first quartile. Another option is to classify all TAZs into two different groups (i.e., the 1st quartile for one group and the 2nd, 3rd and 4th quartiles for the other group), and then develop small area population estimation models for each group.

5. Acknowledgments

The authors are very grateful to the two anonymous reviewers for their expertise and voluntary efforts to enhance this technical article. The outcomes of this research paper are funded by (a) the Japan International Cooperation Agency (JICA) under AUN/SEED-Net project for a Ph.D. Sandwich program at De La Salle University, Philippines and Osaka University, Japan; (b) Commission on Higher Education Development for a Project of Development and Application of Transport Desirability Index for the Assessment and Planning of Transport Systems in the Philippines; and (c) the Big Data Analytics and Applications (MTNN:BDAA) project of the Department of Science and Technology (DOST), Philippines.

6. References

- [1] World Bank. GDP growth (annual %) [Internet]. 2019 [cited 2019 Aug 19]. Available from: https://data.worldbank.org/indicator/NY.GDP.MKTP. KD.ZG?end=2018&locations=PH&start=2015.
- [2] JICA, Department of Transportation and Communications. The project for capacity development on transportation planning and database management in the republic of the Philippines: MUCEP. Japan: Oriental Consultants Global Co., Ltd; 2015.
- [3] PSA. Philippine Statistics Yearbook. Manila: Philippine Statistics Authority; 2018.
- [4] Park Y, Heim-LaFrombois ME. Planning for growth in depopulating cities: an analysis of population
- Deng C, Wu C, Wang L. Improving the housing-unit method for small-area population estimation using remote-sensing and GIS information. Int J Rem Sens. 2010;31(10):5673-88.
- [6] Ghosh M, Rao JN. Small area estimation: an appraisal. Stat Sci. 1994;9:55-76.
- [7] Skirbekk V, Prommer I, Samir KC, Terama E, Wilson C. Report on methods for demographic projections at multiple levels of aggregation. n.p.: PLUREL; 2007.
- [8] Smith SK. A review and evaluation of the housing unit method of population estimation. J Am Stat Assoc. 1986;81(394):287-96.
- [9] Smith SK, Cody S. An evaluation of population estimates in Florida: April 1, 2000. Popul Res Pol Rev. 2004;23:1-24.
- [10] Wu C, Murray AT. Population estimation using landsat enhanced thematic mapper imagery. Geogr Anal. 2007;39:26-43.
- [11] Wardrop NA, Jochem WC, Bird TJ, Chamberlain HR, Clarke D, Kerr D, et al. Spatially disaggregated population estimates in the absence of national population and housing census data. PNAS. 2017;115(4):3529-37.
- [12] Smith SK, Tayman J, Swanson DA. State and local population projections. New York: Kluwer Academic; 2001.
- [13] Smith SK, Nogle J, Cody S. A regression approach to estimating the average number of persons per household. Demography. 2002;39:697-712.
- [14] Li G, Weng Q. Using Landsat ETM þ imagery to measure population density in Indianapolis, Indiana, USA. Photogramm Eng Rem Sens. 2005;71:947-58.
- [15] PSA. Philippines Statistics Yearbook. Quezon: Philippine Statistics Authority; 2017.
- [16] Rith M, Abad RP, Fillone AM, Doi K, Biona JBM. Understanding the impact of urban form attributes on household vehicle ownership and choice in metro manila: Modeling, simulation, and application. Eng. Appl. Sci. Res. 2019;46(3):238-47.
- [17] Rith M, Fillone A, Biona JBM. The impact of socioeconomic characteristics and land use patterns on household vehicle ownership and energy consumption in an urban area with insufficient public transport service – A case study of metro Manila. J. Transp. Geogr. 2019;79:102484.
- [18] Rith M, Lopez NS, Fillone AM, Biona JBM. Determinants of individual vehicle type choice and energy consumption in a heavy traffic metropolis of Southeast Asia featuring the case of Metro Manila. Eng. Appl. Sci. Res*.* 2020;47(1):56-65.
- [19] Rith M, Soliman J, Fillone A, Biona JBM, Lopez NS. Analysis of vehicle survival rate for metro Manila. 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management, HNICEM 2018; 2019; Baguio City, Philippines.