

FORECASTING URBAN POPULATION DISTRIBUTION OF ILOILO CITY USING GIS AND SPATIAL AUTOCORRELATION MODELS

L. Hilario¹, J.A. Duka¹, M.I. Mabalot¹, J. Domingo¹, K.A. Vergara¹, M.J. Villanueva-Jerez¹, K.A. Cabello¹, G.A. Rufino¹, C.J. Sarmiento¹

¹ UP Training Center for Applied Geodesy and Photogrammetry, University of the Philippines Diliman, Quezon City, Philippines - (lhilario, joduka, mdmabalot, jmdomingo3, kpvergara, mvjerez, kecabello, gdrufino, cssarmiento)@up.edu.ph

KEY WORDS: Population Forecasting, Small area, GIS, Spatial Statistics, Spatial Autocorrelation Models, Smart City

ABSTRACT:

Rapid urbanization in localities offers a lot of opportunities but also imposes a lot of challenges due to its direct relationship to population growth. This leads to an increase in the demand for essential goods and services such as food, energy, water among others. Hence, small-area population forecasts have long been an important element in urban and regional planning to aid in the decision-making processes in a locality. The promise of smart cities, through the use of advanced technologies, is to make cities livable and sustainable, preparing more opportunities and addressing challenges on urbanization. This study aims to forecast population distribution in Iloilo city by incorporating GIS techniques and highlighting the use of spatial autocorrelation models. The spatial interaction effects between neighboring barangays are taken into consideration to identify a set of factors affecting the population. The results identified a set of significant explanatory variables and whether it will result in an increase or decrease in population. The study also illustrates the resulting population forecast comparing it to the actual total population of the city.

1. INTRODUCTION

The distribution of the population in any given area can be attributed to multiple factors concerned with human activities (Fang and Jawitz, 2018). Iloilo City, a highly urbanized city in the Western Visayas Region, has a total population of around 447,992 and has a total land area of 7,834 hectares. It consists of 180 barangays and is divided into seven (7) districts (Iloilo City CLUP, 2021). According to the Philippine Statistics Authority's (PSA) 2015 Census, the total annual growth rate of the City is around 447,992, with a 1.02% annual growth rate, relatively lower than the annual population growth rate of the Philippines in general. But in 2000, urban PGR was about 1.84% annually, followed by a fall in 2010 with only 0.51%, and 1.05% in 2015. Iloilo City had a population density per square meter of 5,719 in 2015.

Rapid urbanization brings a lot of opportunities but imposes a lot of challenges in a locality as well. One of which is an increasing population that also increases the demand for essential goods and services such as food, energy, water among others (Fróna et al., 2019). Hence, population projections in small areas have long been an important element in the urban and regional planning processes because of how it drives the change in demands for resources and involves the discussion of smart growth, comprehensive planning, and growth management in urban planning (Chi and Wang, 2017; 2016). Most of the studies, however, were developed and refined on forecasting populations in a larger area. (e.g. provinces/ regions or as a whole country) thus, small area forecasting offers data that is too thin for these methods (Chi and Voss, 2010). In this study, a small area refers to a geographical area that is smaller than countries and regions, which in this case we are referring to a city-level forecast.

A lot of factors could affect future population trends. Traditional methods involve only using population data and its derivatives

(i.e., population density, growth rate). To address these challenges for small-area forecasting, Chi and Wang (2017) suggested taking into account the characteristics of the neighboring units as independent distribution assumption cannot hold, as nowadays interactions among geographic units have increased dramatically. Several studies have also identified and proposed a number of factors to be considered which involves (1) urbanization, due to its bidirectional relationship with the population (Nieves et al., 2017); (2) accessibility to roads and transportation systems, as it plays a vital role in identifying human settlements (Dobson et al., 2000); (3) migration, as population tends to move into cities; (4) topographic suitability, as the human population is more likely to prioritize physiographic parameters such as elevation and proximity to coastlines (Small and Cohen, 2004); (5) social sensing data by taking into account points-of-interests (POIs) in the city - POI types that have included more information associated with human activities are considered to have relatively better livability and higher population density (Yang et al., 2019); and (6) census data containing demographic information which is widely used in population forecasting.

The promise of smart cities, through the use of advanced technologies, is to make cities livable and sustainable, preparing more opportunities and addressing challenges on urbanization. In the context of the creation of smart cities, this simulation study could be the groundwork for future urban models and consequently could help in the planning and monitoring of the city and as well in the development of smarter solutions. Analyzing the factors that could significantly affect the population distribution and projection should be done frequently.

Building an urban geodatabase on population density in a completely dynamic system and inevitably growing population plays a vital role in achieving sustainable social development.

Thus, population forecasting is increasingly becoming more in demand. This study aims to forecast population distribution in Iloilo city through incorporating GIS techniques and different statistical methods through the identified factors affecting population distribution.

2. OBJECTIVES

The study is being implemented in line with the objectives of the Project “A Link-Up of Geomatics and Social Science Research for the Development of Smart Cities” which aims to develop digital twin including targeted simulations in Iloilo City to address present and future challenges of the city. This research aims to relate the identified factors affecting the population distribution of the city by using spatial autocorrelation models applied in GIS. The output of this study also aims to aid the LGU and urban planners in the city in developing socio-economic policies and programs for its citizens.

The specific objectives of this paper are as follows:

1. To identify factors affecting population growth and distribution in Iloilo City;
2. To incorporate the use of GIS techniques with statistical methods to forecast population distribution per barangay;
3. To provide statistical and spatial analysis of the identified factors and their relationship to population; and
4. To produce a report that would relay the results of the study to the LGU which can, in turn, be used as the basis

for the development of new or updated socio-economic policies.

3. METHODOLOGY

The following sections summarize the procedures to be undertaken to meet the objectives of the study. It is based on the hypothesis of forecasting population for small areas with the explanatory variables focusing on urbanization, past population data, accessibility, and migration. This also highlights the incorporation of spatial interaction effects between neighboring barangays to identify a set of factors affecting the population. The workflow of this study can be seen in Figure 1.

3.1 Data and Pre-processing

In pre-processing of initially identified factors, it was found that the slope of Iloilo city was generally flat with 83.87% of the total land area having a slope of <3%. The level of urbanization of Iloilo City was found to have been at 100% since 1980 based on the city's 2021-2029 CLUP. Because of these findings, factors were limited to historical population data including derived population growth rate and density; additional demographic variables based on identified variables by Chi and Wang (2017); migration data per barangay; and the use of POIs to compare the relationship of population distribution with the accessibility to roads and POIs that corresponds to the basic needs and primary sectors (e.g., hospital, schools, markets, transportation terminal, etc.). The factors considered in this study are defined in Table 1.

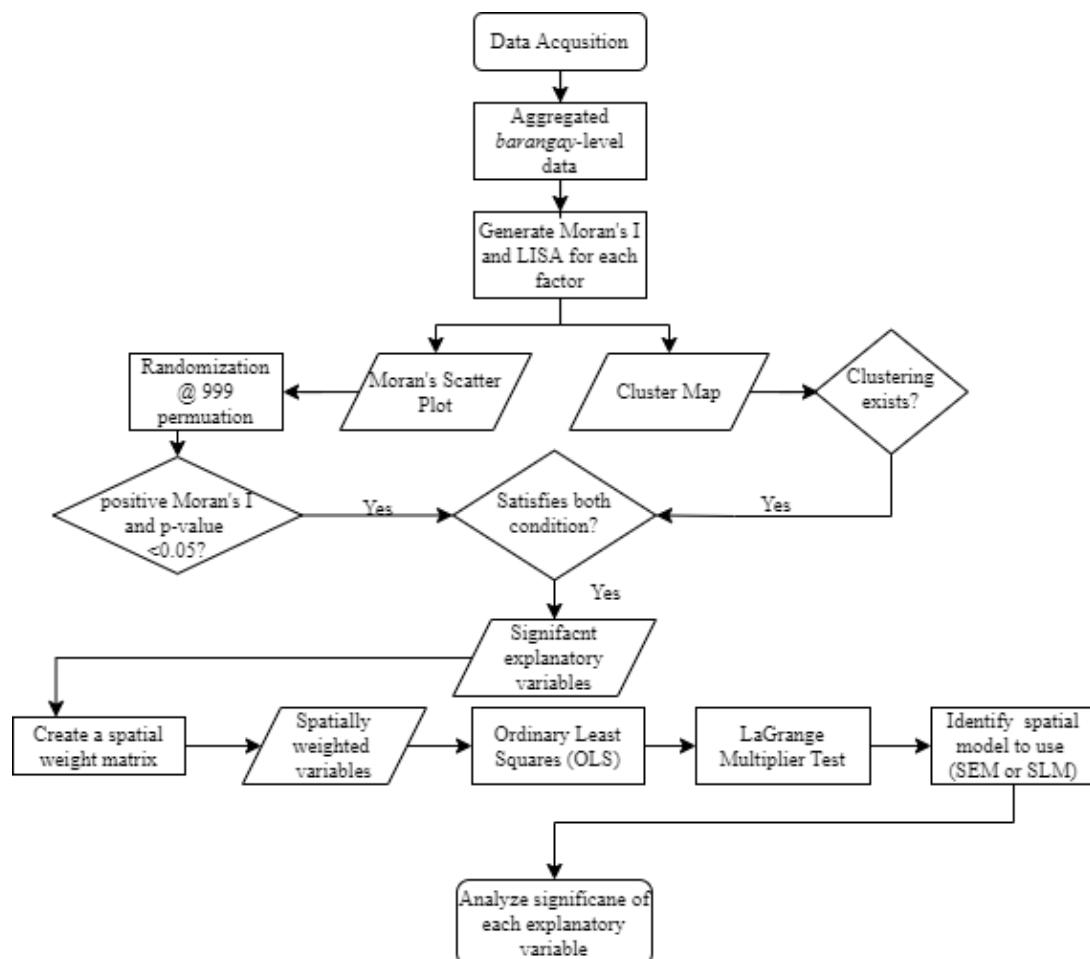


Figure 1. Workflow of the study.

Factor	Mean	Min	Max	Std. dev	Source
Demographics					
Population per brgy 2015	2488.8	75	15150	2393.1368	PSA 2015 census
Population density 2015	13989	296	111443	15285.0530	PSA population derived
Demographic structure (2015)					
% of young population (0-12 years old)	0.1231	0.0342	0.1570	0.0180	PSA 2015 census derived values
% of old population (65+ years old)	0.0642	0.0264	0.1602	0.0214	
% of high school graduates (25+ years old)	0.1425	0.0214	0.3067	0.0397	
% of college graduates (25+ years old)	0.1991	0.0392	0.4239	0.0856	
Migration (percentage of migrated population from 5 years ago)					
Local migration per brgy 2010	47.472	0	100	56.3246	PSA 2010 census
Land Classification					
% of area residential land use	0.3858	0	0.794	0.3469	Iloilo City LULC 2011-2020
Accessibility (inverse nearest distance from centroid of each brgy) in km					
Health center	0.2415	0.0121	1.0486	0.1663	POIs from Iloilo City LGU
Hospital	1.3258	0.0429	6.4383	1.2799	
School (elementary/ high school)	0.8072	0.3696	0.7176	0.8578	
Major/primary road	0.4752	0.0021	2.8874	0.5335	
Transportation terminal/ stop	1.3108	0.1086	3.7682	0.7339	
Public market	0.8569	0.5397	0.9225	0.7717	
Supermarket	0.7026	0.3615	3.0714	0.5848	
Bike lane	0.6629	0.0038	3.4450	0.5243	
Brgy. hall	0.1893	0.0152	0.1966	0.1533	
Police station	0.6912	0.0844	2.7084	0.4894	

Table 1. Descriptive statistics of each factor.

3.2 Spatial Autocorrelation

The histogram for each factor was analyzed to see if it follows a normal distribution. However, the dependent (total population per barangay) and all independent variables resulted in a positively skewed distribution which interprets to having a strong tendency to display negative bivariate correlation (Griffith, 2019). To estimate the variables, we applied a log transformation to scale it and make the distribution more “normal”.

Before testing for spatial autocorrelation, a spatial weight matrix was created. This defined the connectedness of neighboring geographies. In this study, the rook weight matrix with 1st order contiguity was selected as it resulted in the factors having a higher Moran’s I value and a lower p-value in general. The rook contiguity only considers the common sides of the polygons in the neighbor relation (common vertices are ignored) (GeoDa Development Team, n.d.).

In order to assess the spatial relationships of each factor, the local Moran’s I was implemented to measure spatial autocorrelation. In Moran’s scatter plot, randomized at 999 permutations, each factor was evaluated using Moran’s I and p-value. The Moran’s I should be close to 1 and can be interpreted as positive values indicate similarity between neighbors while negative values

indicate dissimilarity. The p-value should also fit the threshold of $p < 0.05$ for it to be statistically significant.

Aside from Moran’s scatter plot, a cluster map was generated in GeoDa to show locations with significant local spatial autocorrelation by type of association. These four types of spatial associations between the values of incidence vs the values of the independent variable are based on the four quadrants of the Moran’s scatter plot: the spatial clustering of similar values (1) high-high, (2) low-low; and spatial association of dissimilar values or spatial outliers (3) high-low, and (4) low-high (Anselin, 2017).

3.3 Application of Spatial Regression Models

Once proved that spatial autocorrelation exists, tests for different spatial regression models were conducted. The standard method in estimating most spatial analyses is to start by using non-spatial linear regression models such as the ordinary least squares (OLS). Additionally, two popular spatial regression models are the spatial error model (SEM) and the spatial lag model. To test which spatial model will be used, a workflow created by Anselin (2005) was followed. The workflow utilizes LaGrange Multiplier (LM) test that is based on the least-squares residuals and calculations involving the spatial weight matrix W. The spatial dependence can either be the result of spatial

autocorrelation of the dependent variable (spatial lag) or spatial autocorrelation in the residuals (spatial error) (Liu and Noback, 2010).

Using OLS, the p-value for each variable was checked if it is statistically significant ($p < 0.05$) in explaining population change. Statistically insignificant variables were removed from the input until it resulted in a set of factors that all produced a statistically significant p-value. For this study, the OLS resulted in the following p-value for each LM test as shown in Table 2. Both the LM error and lag obtained a significant p-value at 3.957×10^{-8} and 2.081×10^{-10} respectively. In the case that both LM error and lag are significant, the robust versions will be assessed on which test resulted in a significant value. Here, we proceeded in using the spatial lag model as it obtained a significant p-value of 0.001009. The significance of each independent variable from the OLS was assessed and projected values for the 2020 population were generated.

LM Test	Value	<i>p</i> -value
LM (error)	30.171	3.957×10^{-8}
LM (lag)	40.389	2.081×10^{-10}
Robust LM (error)	0.59181	0.4417
Robust LM (lag)	10.81	0.001009
SARMA	40.9811	1.262×10^{-9}

Table 2. Diagnostics for spatial dependence.

4. ANALYSIS AND VISUALIZATION

Figure 2 and Figure 3 are sample local cluster maps that demonstrate the difference between variables that show spatial dependence and without spatial dependence, respectively. It can be observed in Figure 2 that local migration in a small area shows no significant local spatial autocorrelation. No clustering occurs as opposed to the population density data where a positive spatial autocorrelation can be seen with the clustering of high-high and low-low values. This may be due to limited data on where migration was considered if only the person is from another city and no data of the migration between barangays was recorded. Another possible reason is the implementation of the methodology in a smaller area and where migration is also observed to be not significantly high since it is not a central city like Metro Manila.

After mapping out the spatial concentration of the explanatory variables, the effects of the independent variables on the population were quantified using OLS and the spatial lag model. Table 3 summarizes the estimates of the two models used in this study and how it affects the population forecast for 2020.

From the initial regression using OLS, 62.1% of the resulting 2020 population in Iloilo City can be explained by six (6) variables which include the population density from 2015, percentage of the old population (composed of 65+ years old),

percentage of residential land classification in each brgy, and the accessibility to POIs of public markets, hospital, and transportation terminal/ stop.

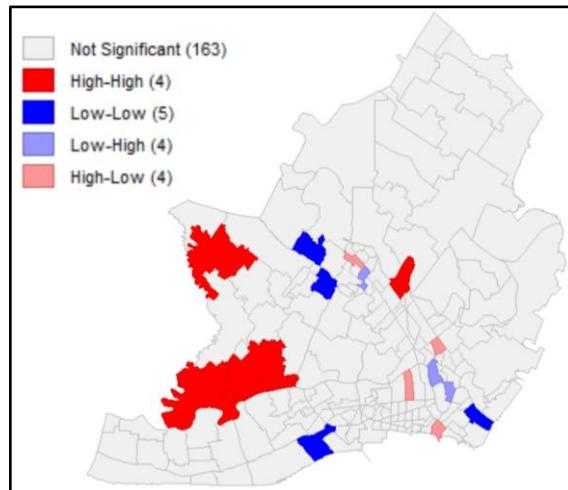


Figure 2. Local migration for each brgy no clustering, Moran's I = -0.048, p-value = 0.173 (significant at $p \leq 0.5$).

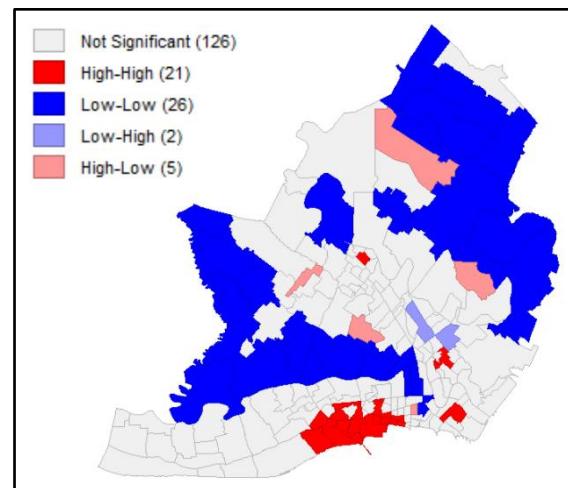


Figure 3. 2015 population density of each brgy with clustering, Moran's I = 0.450, p-value = 0.001 (significant at $p \leq 0.5$).

In the spatial lag model, the R^2 increases from 0.621 to 0.705 indicating the spatial lag as a stronger model. The Akaike Information Criterion (AIC) and Schwarz criterion both decreased from 82.97 and 108.52 to 47.63 and 76.37, respectively, ensuring that the spatial lag is a better test than OLS. This shows that when the spatial weights are taken into consideration in the model, the autoregression becomes noticeably stronger in predicting the dependent variable than a simple OLS Regression (Singh et al., 2020). The multicollinearity condition number also did not exceed 30 therefore the independent variables identified are not strongly correlating to each other.

Independent Variable	Iloilo City Population Forecast 2020	
	OLS	Spatial Error
<i>log</i> population density 2015	0.214*** (0.061)	0.186*** (0.053)
<i>log</i> % old	-0.701*** (0.182)	-0.665*** (0.157)
<i>log</i> % residential	0.451*** (0.079)	0.321*** (0.069)
<i>log</i> nearest distance to public market	0.442*** (0.076)	0.262*** (0.071)
<i>log</i> nearest distance to hospital	0.293*** (0.07)	0.118* (0.066)
<i>log</i> nearest distance to transportation terminal/ stop	0.156** (0.078)	0.069 (0.068)
N	180	180
R ²	0.621	0.705
Multicollinearity Condition Number	28.5	
Akaike Inf. Crit. (AIC)	82.97	47.63
Schwarz criterion	108.52	76.37

Significance level: * $p<0.1$; ** $p<0.05$; *** $p<0.01$

* Standard errors in parenthesis

Table 3. Result of the OLS and spatial lag model in forecasting the population of Iloilo City.

The number of significant independent variables decreased to five (5) where the variable accessibility to transportation terminal is no longer statistically significant. Most of the factors identified have a positive association with the 2020 population whereas a 1% increase in the following variables will result in an increase in population, depending on the resulting value of the coefficients. All the POIs resulted in a positive association (where an increase in the distance from the POI will increase population) as opposed to the assumption that POIs associated with the primary sectors and offering the basic necessities are considered to have relatively better livability and higher population density. To further investigate the resulting relationship of the said variables, the POIs were mapped and analyzed over the annual population growth map of Iloilo City for 210-2020.

In Figure 4, the identified POIs are mostly located in areas with smaller populations and show a population decline. Meanwhile, areas away from the POIs have high population growth rates and where the population is highly concentrated (see Figure 5). This agrees with the spatial association resulting from the OLS and spatial lag model.

The population density of 2015 and the percentage of residential land in each barangay both have a significant positive association with the population while the old population has a negative association. The positive association

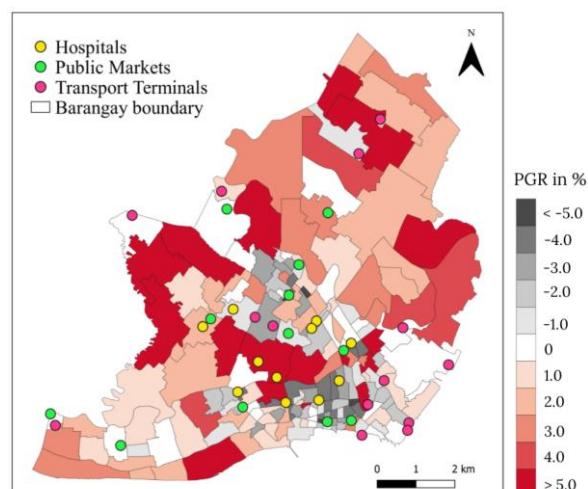


Figure 4. Location of POIs over annual population growth map (2010-2020).

of population density follows its equation where it has a direct relationship with population count. The increase in residential land corresponds to more spaces for people to live in thus the positive correlation.

The map of the predicted values from the spatial lag model in comparison with the actual population of Iloilo City for 2020 is

shown in Figure 5. Figure 6 shows the percent deviation of the projected population. The projected population range for each barangay is significantly comparable visually however, the output forecast resulted in an average percent error of 53.11%. Although the mapped population has a similar pattern of population concentration for the barangays, the predicted values are much lower than the actual population count. This can be observed in Figure 6 where the highly populated barangays

resulted in an underestimation. An increase in population growth in 2010 (as shown in Figure 4) may have contributed to this which can be also associated with the rapid increase of population throughout the years. Further explanations are due to the lack of additional explanatory variables that may highly affect the total population in each barangay such as the economic factors.

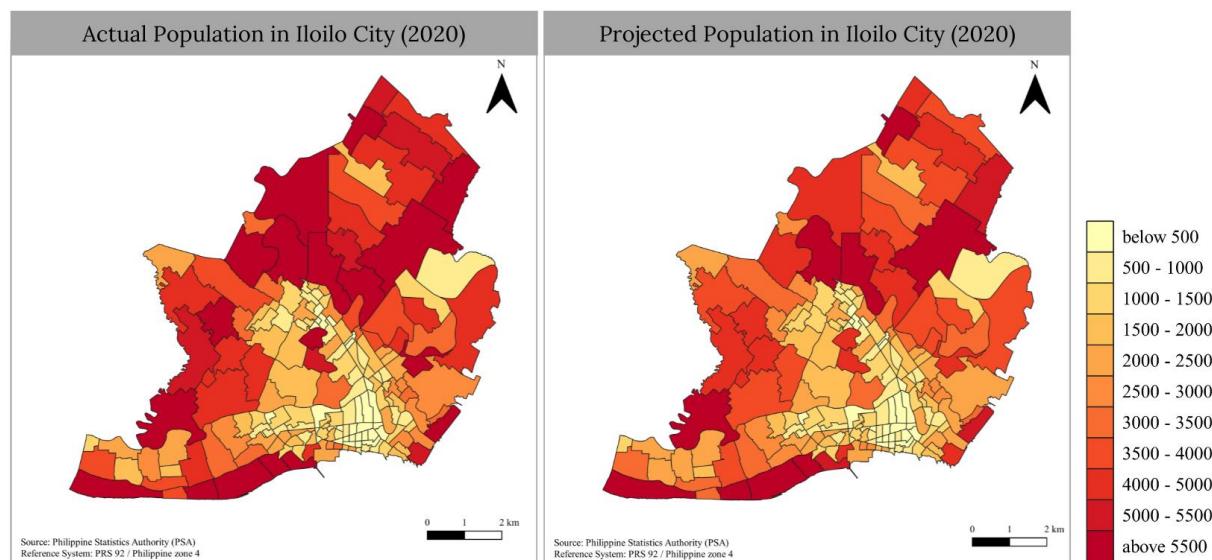


Figure 5. Comparison of actual vs projected total population in Iloilo City for 2020.

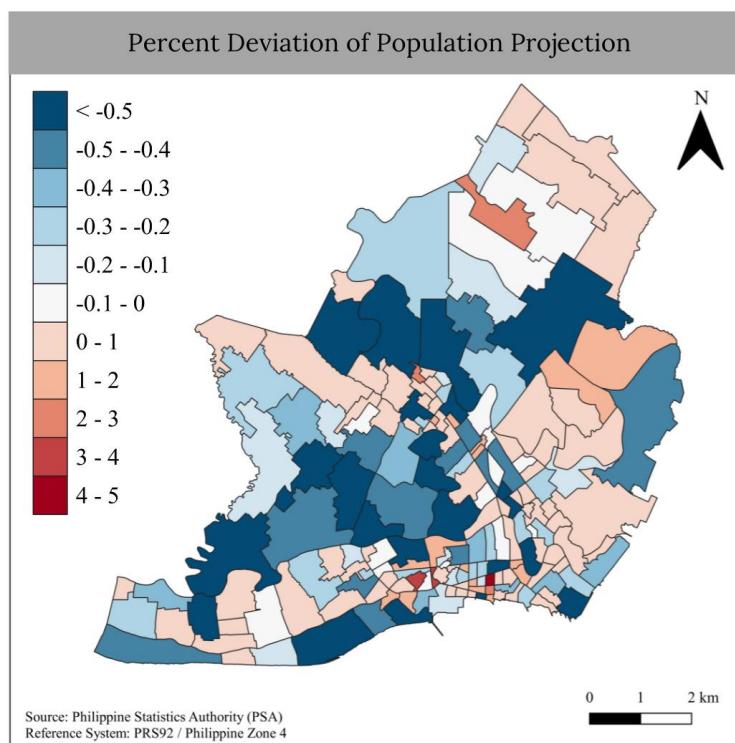


Figure 6. Percent deviation of the projected population from the actual population of Iloilo City for 2020.

5. CONCLUSION

In this study, we investigated the effects of several factors on population and create a forecast of the population in Iloilo City by implementing a spatial autoregressive model using the identified explanatory variables. These factors include the demographic data and its structure, migration, land classified as residential land, and accessibility to POIs that provides basic necessities and part of the primary sectors. The results indicate that the population density, percentage of old population, percentage of lands that are classified as residential land, and access to the POIs of public market, hospital, and transport terminals has a significant positive association with population; and the old population having the biggest impact in this model and have a negative spatial association with the total population count. The methodology of the study was able to demonstrate how the spatial weights affect the autoregression by making the model noticeably a better fit in predicting the dependent variable, by either a Spatial lag model or spatial error model. Furthermore, the identified variables in the model followed the assumed relationship with the total population except for the POIs. The relationship of POIs to the population growth in the city was explained when observed over the map of the population distribution in Iloilo city where the POIs are mostly located in areas with smaller populations and farther from areas with high population distribution. Following the relationship of POIs and the population as suggested by Yang et al., (2019), it can be inferred that since the denser population lies in the outer boundary of the city, more POIs may tend to appear in the area to cater to the needs of the growing population. This can also be studied further in future researches as the analysis of the POIs in relation to the growing population was limited in this study.

5.1 Limitations of the study

The biggest limitation of the study was the availability of the data that can be considered as explanatory variables. The forecasted population values obtained a relatively high percent error even if the model was able to identify a set of factors that are significantly correlated with population. Population forecasts cannot provide absolute predictions of future population change and are even a more difficult task when projected in small areas as errors in projection due to unforeseen events or influential factors may result in important variation in the projection (American Society of Planning Officials, 1950; Chi, Zhou and Voss, 2011; Rayer and Smith, 2010). However, these projections can still be used to help urban planners identify the direction and magnitude of population change to simulate scenarios and rule out unlikely events.

This study may be improved in future related studies by examining a wider range of factors affecting the population. It is important to note that the population is affected by a wide range of factors that often interact with each other and mutually influence population change (Chi and Ventura, 2011). Some of these factors not included in this study are a wider range of describing demography such as ethnicity and race, gender; socioeconomic factors - employment opportunities, crime rate, income growth and distribution, housing conditions, housing prices, household income, poverty rate; more variables explaining transportation accessibility - travel time to work and use of public transportation; and variables describing the natural environment - application of disaster risk maps.

REFERENCES

- American Society of Planning Officials, 1950. Population forecasting. American Planning Association. Information Report No. 17. planning.org/pas/reports/report17.htm (18 October 2021).
- Anselin, L., 2005. Exploring Spatial Data with GeoDataTM: A Workbook. Center for Spatially Integrated Social Science. Urbana, IL. geodacenter.github.io/docs/geodaworkbook.pdf (6 August 2021)
- Anselin, L., 2017. A Local Indicator of Multivariate Spatial Association. [10.13140/RG.2.2.18101.58084](https://doi.org/10.13140/RG.2.2.18101.58084).
- Chi, G., Ventura, S. J., 2011. Population Change and Its Driving Factors in Rural, Suburban, and Urban Areas of Wisconsin, USA, 1970–2000. *International Journal of Population Research*, 2011, 1–14. doi:[10.1155/2011/856534](https://doi.org/10.1155/2011/856534)
- Chi, G., Voss, P. R., 2010. Small-area population forecasting: borrowing strength across space and time. *Population, Space and Place*, 17(5), 505–520. doi:[10.1002/psp.617](https://doi.org/10.1002/psp.617)
- Chi, G., Wang, D., 2016. Small-Area Population Forecasting: A Geographically Weighted Regression Approach. *The Frontiers of Applied Demography*, 449–471. doi:[10.1007/978-3-319-43329-5_21](https://doi.org/10.1007/978-3-319-43329-5_21)
- Chi, G., Wang, D., 2017. Population projection accuracy: The impacts of sociodemographics, accessibility, land use, and neighbour characteristics. *Population, Space and Place*, 24(5), e2129. doi:[10.1002/psp.2129](https://doi.org/10.1002/psp.2129)
- Chi, G., Zhu, J., 2007. Spatial Regression Models for Demographic Analysis. *Population Research and Policy Review*, 27(1), 17–42. doi:[10.1007/s11113-007-9051-8](https://doi.org/10.1007/s11113-007-9051-8)
- Chi, G., Zhou, X., Voss, P. R., 2011. Small-area population forecasting in an urban setting: a spatial regression approach. *Journal of Population Research*, 28(2-3), 185–201. doi:[10.1007/s12546-011-9053-6](https://doi.org/10.1007/s12546-011-9053-6)
- Dobson, J., Bright, E., Coleman, P., Durfee, R., Worley, B., 2000. A Global Population database for Estimating Populations at Risk. *Photogrammetric Engineering & Remote Sensing*, 66.
- Fang, Y., Jawitz, J. W., 2018. High-resolution reconstruction of the United States human population distribution, 1790 to 2010. *Scientific Data*, 5, 180067. doi:[10.1038/sdata.2018.67](https://doi.org/10.1038/sdata.2018.67)
- Fróna, D., Szenderák, J., Harangi-Rákos, M., 2019. The Challenge of Feeding the World. *Sustainability*, 11(20), 5816. doi:[10.3390/su11205816](https://doi.org/10.3390/su11205816)
- GeoDa Development Team, n.d. GeoDa Software. Center for Spatial Data Science, University of Chicago. geodacenter.github.io (18 October 2021).
- Griffith, D. A., 2019. Negative Spatial Autocorrelation: One of the Most Neglected Concepts in Spatial Statistics. *Stats*, 2(3), 388–415. doi:[10.3390/stats2030027](https://doi.org/10.3390/stats2030027)
- Iloilo City Comprehensive Land Use Plan (CLUP) 2021-2029, 2021. Vol.1. Local Development Council of Iloilo City. City Planning and Development Office.

Liu, A., Noback, I., 2010. Determinants of regional female labour market participation in the Netherlands. *The Annals of Regional Science*, 47(3), 641–658. doi:10.1007/s00168-010-0390-8

Nieves, J.J., Stevens, F.R., Gaughan, A.E., Linard, C., Sorichetta, A., Hornby, G., Tatem, A. J., 2017. Examining the correlates and drivers of human population distributions across low- and middle-income countries. *Journal of The Royal Society Interface*, 14(137), 20170401. doi:10.1098/rsif.2017.0401

Rayer, S., Smith, S. K., 2010. Factors Affecting the Accuracy of Subcounty Population Forecasts. *Journal of Planning Education and Research*, 30(2), 147–161. doi:10.1177/0739456x10380056

Singh, S. K., Gupta, J., Sharma, H., Pedgaonkar, S. P., Gupta, N., 2020. Socio-economic Correlates and Spatial Heterogeneity in the Prevalence of Asthma among Young Women in India. *BMC Pulmonary Medicine*, 20(1). doi:10.1186/s12890-020-1124-z

Small, C., Cohen, J. E., 2004. Continental Physiography, Climate, and the Global Distribution of Human Population. *Current Anthropology*, 45(2), 269–277. doi:10.1086/382255

Yang, X., Ye, T., Zhao, N., Chen, Q., Yue, W., Qi, J., Jia, P., 2019. Population Mapping with Multisensor Remote Sensing Images and Point-Of-Interest Data. *Remote Sensing*, 11(5), 574. doi:10.3390/rs11050574