World Academy of Science, Engineering and Technology
International Journal of Urban and Civil Engineering
Vol:13, No:7, 2019

A Remote Sensing Approach to Calculate Population Using Roads Network Data in Lebanon

Kamel Allaw, Jocelyne Adjizian Gerard, Makram Chehayeb, Nada Badaro Saliba

Abstract**—**In developing countries, such as Lebanon, the demographic data are hardly available due to the absence of the mechanization of population system. The aim of this study is to evaluate, using only remote sensing data, the correlations between the number of population and the characteristics of roads network (length of primary roads, length of secondary roads, total length of roads, density and percentage of roads and the number of intersections). In order to find the influence of the different factors on the demographic data, we studied the degree of correlation between each factor and the number of population. The results of this study have shown a strong correlation between the number of population and the density of roads and the number of intersections.

Keywords**—**Population, road network, statistical correlations, remote sensing.

I. Introduction

THE accurate determination of population size and distribution is an important data source for urban planners and decision makers to handle social, economic and environmental challenges [1]. However, the traditional ways of population estimation (e.g. census or population registers) seem to be inefficient in developing countries which do not contain well-organized statistical and technological systems [2]. For this reason, researchers in urban geography have been interested in establishing models to estimate population in a determined area which has been called "social physics" [3]. After that, several equations have been proposed to describe population [4]-[6]. Although those equations could be used to explain population distribution through population density, they were not efficient to estimate population data. Moreover, those equations were not valid world widely; they were suitable for certain locations [7], [8]. For this reason, aerial interpolation has been used as a method for population estimation. This interpolation could be done with or without ancillary information and using several methods such as point-based methods or area-based methods [9], [10]. Moreover, the use of remote sensing data provides a low cost and time consuming procedure for population estimation [11]. In the 1950s, the lack of data as a result of low census frequency [12] and the need to check census results [13] pushed towards the initiation of a new approach based on the combination between remote sensing and statistical modeling to estimate population since it depends from several spatial factors which can be extracted using remote sensing. [5], [6], [14]. In general, several factors have been used to establish the correlation between spatial elements and population data: urban areas [15], [16], land use [17], [12], dwelling units [18]-[21], image pixel characteristics [22], [23] and other physical or socioeconomic characteristics [24]-[27]. In this context, the geographic information system (GIS) constituted a revolution in population estimation since it allows the integration of several types of data (spatial and non-spatial data) [28]. The literature relative to population estimation using statistical modeling and remote sensing reveals a variety of statistical models established to estimate population [29]-[31]. These models fit well in organized areas but they cannot be used in unorganized areas such as rural areas. Lebanon is one of those countries presenting a lack of accurate population data. For a long time, the accurate estimation of Lebanese population is not available. Therefore, it is not possible to conduct a thorough study of many thorny issues. Demographics is one of the most relevant studies because of their implications for political, social and economic life. This state of affairs forces us to look for other factors that can give ideas about demographic data, one of these proposed factors being the road network because of its importance in the life of the population. For this reason, we try in our study to find the form of the relationship between road networks and demographics. Consequently, the use of remote sensing data may constitute an opportunity to fill this data gap. The aim of this study is to quantify the relation between road networks' elements (i.e. length of roads, density and percentage of roads and the number of nodes) and population data.

Kamel Allaw is with the CREEMO, Geography Department, Saint Joseph University and with the Surveying Department, Islamic University of Lebanon (e-mail: Allaw.Kamel@hotmail.com).

Jocelyne Adjizian Gerard and Nada Badaro Saliba are with the CREEMO, Geography Department, Saint Joseph University Beirut, Lebanon (e-mail: Jocelyne.gerard@usj.edu.lb, nada.saliba@usj.edu.lb).

Makram Chehayeb is with the Surveying Department, Islamic University of Lebanon, Beirut, Lebanon (e-mail: makram.ch92@gmail.com).

II. Case Study

The Lebanese Republic is located approximately at 34 °N and 35 °E. It is bordered by Syria to the north and east, and by occupied Palestine to the south, with a total area of 10 452 km². This country is divided into nine governorates (Beirut, Mount Lebanon, Bekaa, North, South, Nabatieh, Aakkar, Baalbak El-Hermel and Keserwan Jbeil) which are subdivided into 26 districts (qadaa or caza). The study area is the whole Lebanese territory but 13 municipalities (e.g. unorganized areas) were chosen arbitrarily to perform the study (Fig. 1). Those municipalities are geographically well distributed over the Lebanese territory.

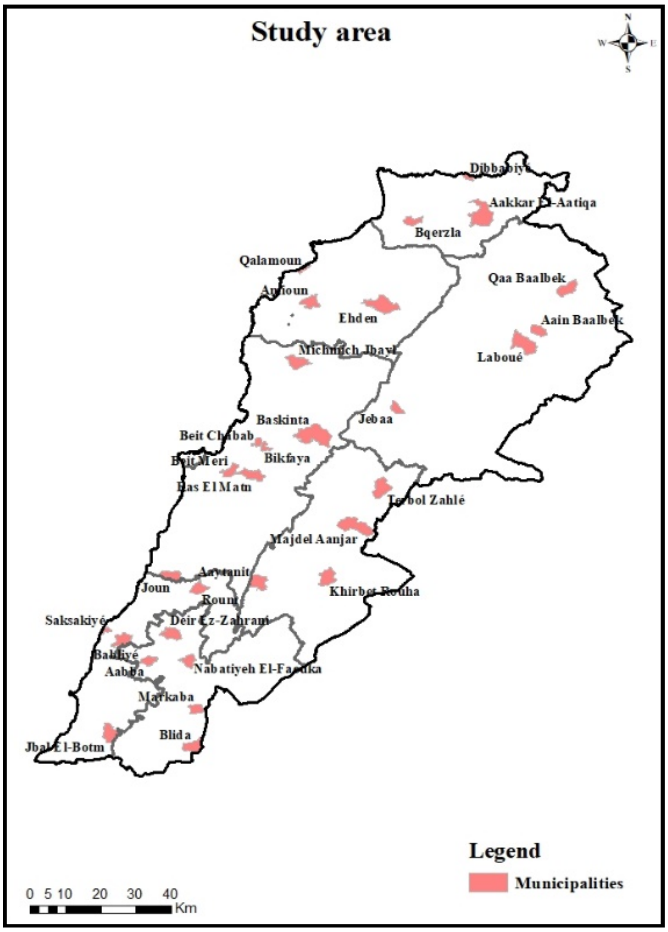

Fig. 1 Distribution of the selected municipalities on the Lebanese map

III. Methodology and Data

The methodology of this research consists in studying the characteristics of road's network in each municipality (total length of roads, length of primary roads, length of secondary roads, density of roads, percentage of roads and the number of nodes). Then, a bivariate correlation has been established between each one of the road's network characteristics and the population data. After that, the factors showing the highest correlations have been chosen and a multivariate regression has been established in order to establish an equation which relates the factors relative to road's networks characteristics and the population count.

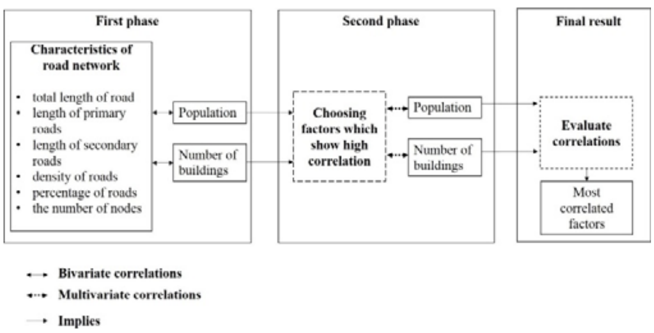

Fig. 2 Principle of the methodology

A. The Characteristics of Roads' Network

1. Length of roads: The length of primary roads, secondary roads as well as the total length of roads has been determined using the data available at [32]. However, many roads were not available in this data set, so, the digitalization of roads on satellite images has been used to cover the missing data.
2. Percentage and density of roads' network: For each municipality, the percentage of roads' network is the ratio between the total area of roads and the area of the municipality. It has been calculated using ArcGIS software by dividing the total area of roads for each municipality by its area. The density of roads' network is defined as the ratio between the total length of roads and the built-up area of the municipality, so, it has been calculated, using ArcGIS, by dividing the total length of the roads with the total built-up area of each municipality (it is expressed in kilometer per kilometer square).
3. Number of nodes: A node can be defined as the intersection points between two roads regardless the road's category (e.g. primary or secondary road) [33]. The number of nodes has been determined using the following method: The intersect tool has been used to determine the points of intersection between roads. However, this will generate many points on the same location, so, using dissolve tool (according to points' coordinates), those points can be replaced by a unique node. Then, the number of nodes for each municipality has been calculated.

B. The Population Count

The data related to the population count have been taken from the municipality.

IV. Results

A. Bivariate Analysis

Figs. 3-6 show the linear regression between the chosen factors (length of secondary roads, percentage of roads, density of roads, number of nodes, length of primary roads and total length of roads) and the population count.

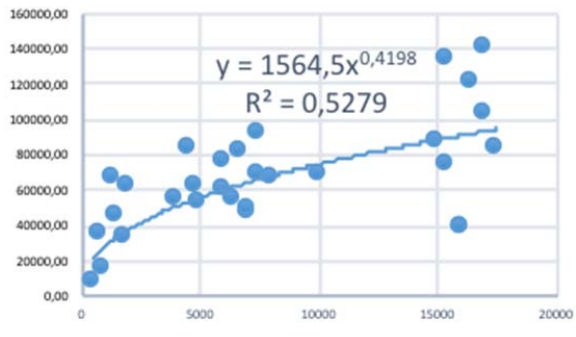

Fig. 3 The relation between the effective of population and the total length of roads

The analysis of coefficient of determination (R-squared) in Table I shows that the total length of roads, the number of nodes, the density of roads and the length of secondary roads are the unique factors which show considerable correlations with the effective of population (Figs. 3-6). As a result, factors whose coefficient of determination (R-squared) is less than 0.5 (e.g. length of primary roads and percentage of roads) might be excluded from the model.

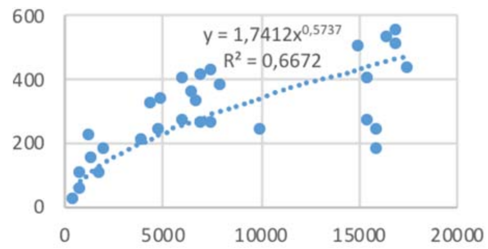

Fig. 4 The relation between the effective of population and the number of nodes

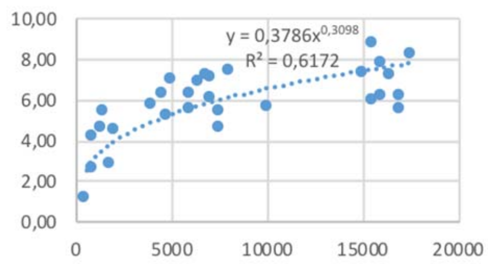

Fig. 5 The relation between the effective of population and the density of roads

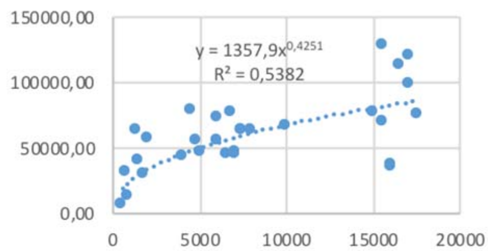

Fig. 6 The relation between the effective of population and the length of secondary roads

TABLE I
EQUATIONS AND COEFFICIENTS OF DETERMINATION SHOWING THE RELATION BETWEEN THE EFFECTIVE OF POPULATION AND FACTORS RELATIVE TO THE CHARACTERISTICS OF ROADS' NETWORK

Independent variable (X)	Equation (y=a.xb)	Coefficient of determination (R2)
Length of secondary roads	y = 1357.9x0.4251	$R^2 = 0.5382$
Percentage of roads	y = 0.0029x0.3125	$R^2 = 0.3551$
Density of roads	y = 0.3786x0.3098	$R^2 = 0.6172$
Number of nodes	1.7412x0.5737	$R^2 = 0.6672$
Length of primary roads	y = 279.48x0.3206	$R^2 = 0.1649$
Total length of roads	y = 1564.5x0.4198	$R^2 = 0.5279$

B. Multivariate Analysis

The factors which show a high correlation with population count have been chosen to construct the model and generate the equation which relates the population to road's network characteristics. The multivariate regression between the effective of population and the factors related to road's network characteristics (total length of roads, number of nodes, density of roads and length of secondary roads) is given by:

$$Y = -6410.6 - 0.00044\ X1 + 6.3048\ X2 + 0.0676\ X3 + 1427.411\ X4 \quad (1)$$

where Y = The effective of population; X1 = Total length of roads; X2 = Number of nodes; X3 = Length of secondary roads; X4 = Roads' density.

The coefficient of determination (R-squared) of this equation is 0.6. Statistically, this value of r-squared reflects a significant relationship between the independent variables (total length of roads, number of nodes, length of secondary roads and roads' density) and the dependent variable (population count or effective of population)

C. Validation

In order to check the reliability of the model, a validation procedure has been conducted. This procedure consists of choosing 5 municipalities with same characteristics as this study's sample (e.g. unorganized area) with a known population count Y. Then, the generated equation will be applied to determine the estimated population count (Y') of each municipality. After that, the population count (Y) and the estimated population count (Y') will be compared and the root squared mean error (RMSE) will be calculated.

TABLE II
THE COLLECTED DATA FOR RESULTS' VALIDATION

n	Total length of roads (X1) [m]	Number of nodes (X2)	Length of secondary roads (X3) [m]	Roads' density (X4) [km/km2]
1	57962.04	189	51499.44	7.04
2	41145.52	134	36558.02	4.98
3	74061.936	241.2	65804.436	8.964
4	37030.968	120.6	32902.218	4.482
5	55546.45	180.9	49353.33	6.723

TABLE III
THE COMPARISON BETWEEN KNOWN AND ESTIMATED POPULATION

Municipality index	Known population	Estimated population	Residual
1	7957	8286	329
2	3891	3996	105
3	12167	12321	154
4	2814	2955	141
5	7500	7638	138

The residual is the difference between the estimated value and the known value (or most probable value). The root mean squared error has been calculated by squaring the residuals, calculating their average, then, calculating the square root of the result (2):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Y' - Y)^2}{N}} \quad (2)$$

where N = Sample size.

The value of RMSE shows that this model has an error of estimation equivalent to 191 which might be considered as acceptable since it constitutes only 2% of difference for a village of 10,000 inhabitants. Moreover, this error may be explained by other factors affecting population.

V. Discussion

As a result of those statistical analyses, a weak relation has been detected between the effective of population and the length of primary roads (R2 = 0.1649) on one hand, and between the effective of population and the percentage of roads (R2 = 0.3551) on the other hand. In fact, according to the used classification, primary roads are the roads relating two municipalities (major roads), for this reason, their lengths are not proportional with the effective of population because the existence of primary roads will not be largely affected by the urban sprawl and population growth. In addition, the weak relation between the percentage of roads and the effective of population can be explained by the adopted definition of roads' percentage which takes into consideration the whole area of the municipality which may contain unpopulated areas. In contrast, a strong relation has been detected between the effective of population and each one of the following factors: the length of secondary roads (R2 = 0.5382), the number of nodes (R2 = 0.6672), the total length of roads (R2 = 0.5279) and the density of roads (R2 = 0.6172). This result can be explained by the internal relations between the road network's elements. To illustrate, the secondary roads are usually existent in case of urban sprawl which means the actual or future population growth. Moreover, the total length of roads is constituted by the combination of the length of secondary roads and the length of primary roads. Furthermore, the number of nodes increases with the rise of roads' ramification (represented by the intersection of primary and secondary roads). At the end, the density of roads, defined as the ratio between the total area of the roads and the sum of built-up areas in a municipality, proves the existence of a strong relation between the effective of population and the built-up areas from one side, and the roads' total area from the other side.

VI. Conclusion

The combination of Remote Sensing data, GIS tools and statistical analysis is profitable to evaluate the correlations between demographic data and characteristics of roads' network. In fact, remote sensing data and GIS tools were essential to extract the characteristics of roads' network and the bivariate analyses were helpful to choose the most correlated factors with the roads' network. As a result, a strong relation has been detected between the effective of population and four elements of roads' network characteristics (length of secondary roads, total length of roads, number of nodes and density of roads). Moreover, the internal relations between those elements have increased the correlation between each one of them and the effective of population. The other elements (length of primary roads and percentage of roads) are not correlated with the effective of population because of the lack of relation between their values and the resultant urban growth. Finally, this study has generated a mathematical relation applicable to rural areas where the knowledge of roads' network elements (easily extracted using remote sensing data and GIS techniques) can help in population's estimation. For future researches, it is recommended to correlate the findings of this research with other morphological characteristics affecting population such as buildings. Moreover, population estimation using remote sensing data needs to be enhanced in order to generate geographic and statistical models compatible with different types of zones and with the highest possible accuracy.

References

[1] Liu, X., and K. C. Clarke, 2002, "Estimation of residential population using high resolution satellite imagery." Proceedings of the 3rd Symposium on Remote Sensing of Urban Areas.

[2] Anderson, D., and Anderson, P., 1973. Population estimates by humans and machines. Photogrammetric Engineering, 39, 147–154.

[3] Stewart, Wartnz J. Q and W., 1958, "Physics of Population Distribution," Journal of Regional Science, 1:99-113.

[4] Clark, C., 1951, "Urban Population Densities," Journal of the Royal Statistical Society, 114:490-496.

[5] Sutton, P., 1997. Modeling population density with night-time satellite imagery and GIS, Computing, Environment and Urban Systems, 21:227–244.

[6] Parr, J. B., 1985, "A Population-Density Approach to Regional Spatial Structure," Urban Studies, 22(4):289-303.

[7] Weiss, H. K., 1961, "The Distribution of Urban Population and an Application to a Servicing Problem," Operations Research, 9(6):860-874.

[8] Newling, B. E., 1965, "Urban Growth and Spatial Structure—Mathematical Models and Empirical Evidence," Annals of the Association of American Geographers,55(4):637-637.

[9] Lam, N. S., 1983, "Spatial Interpolation Methods: A Review," The American Cartographer, 10(2):129-149.

[10] Eicher, C. L. and C. A. Brewer, 2001, "Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation," Cartography and Geographic Information Science, 28(2): 125-138.

[11] Sutton, P., D. Roberts, C. Elvidge, and K. Baugh, 2001. Census from heaven: an estimate of the global human population using night-time satellite imagery. International Journal of Remote Sensing, 1-16, Preview article.

[12] Kraus, S. P., Senger, L. W. and J. M. Ryerson, 1974, "Estimating Population from Photographically Determined Residential Land Use Types," Remote Sensing of Environment, 3(1):35-42.

[13] Clayton, C. and Estes J., 1980, "Image Analysis as a Check on Census Enumeration Accuracy" Photogrammetric Engineering and Remote Sensing, 46:757-764.

[14] Batty, M. and Longley P., 1994, Fractal Cities: A Geometry of Form and Function, London, UK/San Diego, CA: Academic Press, 394 p.

[15] Lo, C. P. and Welch R., 1977, "Chinese Urban Population Estimates," Annals of the Association of American Geographers, 67(2):246-253.

[16] Lee, Y., 1989, An Allometric Analysis of the United States Urban System: 1960-80, Environment and Planning A, 21(4):463-476.

[17] Weber, C., 1994, "Per-zone Classification of Urban Land Use Cover for Urban Population estimation," in Environmental Remote Sensing from Regional to Global

[18] Green, N. E., 1956, "Aerial Photographic Analysis of Residential Neighborhoods: An Evaluation of Data Accuracy," Social Forces, 35:142-147.

[19] Porter, P. W., 1956, Population Distribution and Land Use in Liberia, Ph.D. thesis, London School of Economics and Political Science, London, UK, 213 p.

[20] Hsu, S. Y., 1971, "Population Estimation," Photogrammetric Engineering, 37:449-454.

[21] Lo, C. P. and H. F. Chan, 1980, "Rural Population Estimation from Aerial Photographs," Photogrammetric Engineering and Remote Sensing, 46:337-345.

[22] Iisaka, J. and Hegedus E., 1982, “Population Estimation from Landsat Imagery,” Remote Sensing of the Environment, 12:259-272.
[23] Webster, C. J., 1996, “Population and Dwelling Unit Estimation from Space,” Third World Planning Review, 18(2):155-176.
[24] Dobson, J.E., Bright E. A., Coleman P. R. , Durfee R. C. , and Worley B. A. , 2000. LandScan: A global population database for estimating populations at risk, Photogrammetric Engineering & Remote Sensing, 66:849–857.
[25] Liu, X. and Clarke K. C., 2002, “Estimation of Residential Population Using High Resolution Satellite Imagery,” Proceedings of the 3rd Symposium in Remote Sensing of Urban Areas.
[26] Green, N. E., 1957, “Aerial Photographic Interpretation and the Social Structure of the City,” Photogrammetric Engineering, 23:89-96.
[27] Green, N. E. and Monier R. B., 1959, “Aerial Photographic Interpretation of the Human Ecology of the City,” Photogrammetric Engineering, 25:770-773.
[28] Monmonier, M. S. and Schnell G. S., 1984, “Land-Use and Land-Cover Data and the Mapping of Population Density,” The International Yearbook of Cartography, 24:115-121.
[29] Kaimaris, Dimitris, and Petros Patias. "Population Estimation in an Urban Area with Remote Sensing and Geographical Information Systems." International Journal of Advanced Remote Sensing and GIS (2016): pp-1795.
[30] Karume, K., et al. "Use of Remote Sensing for Population Number Determination." The Open Access Journal of Science and Technology 5 (2017).
[31] Dong, P., Ramesh, S. and Nepali, A., 2010, Evaluation of small-area population estimation using LiDAR, Landsat TM and parcel data. International Journal of Remote Sensing, 31, pp. 5571–5586.
[32] Cycle route planner, https://www.bbbike.org/ Accessible online: November 12, 2018.
[33] Jabbari, S. E., 2016, Node modeling for congested urban networks, 91, 229-249.