A Decision Tree Approach to Estimate Permanent Residents Using Remote Sensing Data in Lebanese Municipalities

K. Allaw, J. Adjizian Gerard, M. Chehayeb, A. Raad, W. Fahs, A. Badran, A. Fakherdin, H. Madi, N. Badaro Saliba

Abstract—Population estimation using Geographic Information System (GIS) and remote sensing faces many obstacles such as the determination of permanent residents. A permanent resident is an individual who stays and works during all four seasons in his village. So, all those who move towards other cities or villages are excluded from this category. The aim of this study is to identify the factors affecting the percentage of permanent residents in a village and to determine the attributed weight to each factor. To do so, six factors have been chosen (slope, precipitation, temperature, number of services, time to Central Business District (CBD) and the proximity to conflict zones) and each one of those factors has been evaluated using one of the following data: the contour lines map of 50 m, the precipitation map, four temperature maps and data collected through surveys. The weighting procedure has been done using decision tree method. As a result of this procedure, temperature (50.8%) and percentage of precipitation (46.5%) are the most influencing factors.

Keywords—Remote sensing and GIS, permanent residence, decision tree, Lebanon.

I. INTRODUCTION

THE determination of population size and distribution using Remote Sensing Data was the objective of several studies since it constitutes an accessible source of data for planners, geographers and decision makers [1]-[5]. One of the challenges for an accurate estimation is the ability to determine accurately the percentage of people who live continuously in a specific region [6]. However, the history of human evolution has shown a continuous change in people's residence. The 21-century has witnessed a rise in urban residents in the third world countries as well as in Lebanon. Despite this tendency, it is clear that some villages, including Lebanese ones, have conserved a high percentage of permanent residents which might be referred to the existence of some common attractive characteristics. In this context, the knowledge of those characteristics may support the orientation of planning policies to reinforce this trend of population distribution. For the previously mentioned reasons, the aim of this study is to evaluate the factors affecting the phenomenon of permanent residence in Lebanon and to generate some

Kamel Allaw is with the CREEMO, Geography Department, Saint Joseph University, Beirut, Lebanon and the with the Surveying Department, Islamic University of Lebanon, 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.

Makram Chehayeb, Ali Raad, Walid Fahs, Ali Badran, Amir Fakherdin and Hussein Madi are with the Surveying Department, Islamic University of Lebanon, Lebanon.

general trends which might be applied in other regions. The determination of those factors might be done using artificial intelligence techniques such as decision tree. Decision tree is one of the predictive modeling approaches used in statistics, data mining and machine learning [7]. This technique has been previously used with GIS and remote sensing data for several geographic applications. To illustrate, it has proved its efficiency to map vegetation [8], to perform land cover classification [9], to detect Spatial Processes in Landscape Transformation [10] and for parcel-based land use change model [11]. For the current study, the adopted methodology involves the use of remotely sensed data and the analysis of data using decision tree method to determine the weight of factors affecting permanent residence phenomenon.

II. STUDY AREA

The Lebanese Republic, occupying an area of 10452 km², is located in western Asia on the eastern shore of the Mediterranean Sea between latitudes 33° and 35° N and longitudes 35° and 37° E. It borders the Sea in the west, Syria in the north and east, and Occupied Palestine in the south. In order to extract the necessary data for this study, a sample of 30 villages (e.g. unorganized areas) have been chosen. Those municipalities are arbitrarily chosen and geographically well distributed throughout the country.

III. METHODOLOGY AND DATA

A. Principle

The used method consists in studying the characteristics (factors) affecting the permanent residence: The slope, the precipitation, the temperature, its proximity to conflict zones, the number of services, and time to CBD. For each one of the first four factors, a map has been established using GIS where the values have been classified into two classes: "appropriate" and "inappropriate" for permanent residence. Regarding the number of services, the number of schools, industries and hospitals has been summarized in a table as well as the time to CBD which has been estimated for each municipality. For each factor, an index, from 1 to 5, has been assigned to the values of the factor where 5 corresponds to the most appropriate condition for permanent residence.

TABLE I
THE INDICES ATTRIBUTED TO EACH MUNICIPALITY ACCORDING TO SLOPE
VALUES

Municipality	Category	Slope value	Area (%)	Index
Deir al zehrani	Appropriate	0-15%	83%	5
	Inappropriate	>15%	17%	1
Nabatiyeh el Fawka	Appropriate	0-15%	96%	5
	Inappropriate	>15%	4%	1
Abba	Appropriate	0-15%	91%	5
	Inappropriate	>15%	9%	1
Blida	Appropriate	0-15%	91%	5
	Inappropriate	>15%	9%	1
Markaba	Appropriate	0-15%	54%	5
	Inappropriate	>15%	46%	1
Saksakiye	Appropriate	0-15%	64%	5
	Inappropriate	>15%	34%	1
Babliye	Appropriate	0-15%	96%	5
	Inappropriate	>15%	4%	1
Roum	Appropriate	0-15%	22%	5
	Inappropriate	>15%	78%	1
Jbeil el botom	Appropriate	0-15%	56%	5
	Inappropriate	>15%	44%	1
Joun	Appropriate	0-15%	45%	5
	Inappropriate	>15%	55%	1
Ras el maten	Appropriate	0-15%	17%	5
	Inappropriate	>15%	83%	1
Mechmech jbeil	Appropriate	0-15%	37%	5
	Inappropriate	>15%	63%	1
Beit Chabeb	Appropriate	0-15%	14%	5
	Inappropriate	>15%	86%	1
Bekfaya	Appropriate	0-15%	23%	5
	Inappropriate	>15%	77%	1
Beit Meri	Appropriate	0-15%	15%	5
	Inappropriate	>15%	85%	1
Terbol Zahle	Appropriate	0-15%	79%	5
	Inappropriate	>15%	21%	1
Majdel Anjar	Appropriate	0-15%	65%	5
	Inappropriate	>15%	35%	1
Kherbet Rouha	Appropriate	0-15%	75%	5
	Inappropriate	>15%	25%	1
Aytanit	Appropriate	0-15%	35%	5
	Inappropriate	>15%	65%	1
Labwe	Appropriate	0-15%	73%	5
	Inappropriate	>15%	27%	1
Al-Qaa	Appropriate	0-15%	100%	5
	Inappropriate	>15%	0%	1
Jebaa	Appropriate	0-15%	75%	5
	Inappropriate	>15%	25%	1
Amyoun	Appropriate	0-15%	89%	5
	Inappropriate	>15%	11%	1
Eheden-Zgharta	Appropriate	0-15%	20%	5
	Inappropriate	>15%	80%	1
Qalamoun	Appropriate	0-15%	66%	5
	Inappropriate	>15%	34%	1
Debabiye	Appropriate	0-15%	58%	5
	Inappropriate	>15%	42%	1
Akkar al atika	Appropriate	0-15%	21%	5
	Inappropriate	>15%	79%	1
Bqerzla	Appropriate	0-15%	51%	5
	Inappropriate	>15%	49%	1
Al-Ain	Appropriate	0-15%	61%	5
	Inappropriate	>15%	39%	1
Baskenta	Appropriate	0-15%	16%	5
	Inappropriate	>15%	84%	1

B. Data Collection

1. The Slope Map

The slope affects the constructability of lands because terrains with high slopes need expensive adjustments, so people prefer to choose flat terrains for their residence. Consequently, the higher the slope of terrains is, the lower the percentage of residents will be. Technically, the slope map has been prepared using contour lines map with an interval of 50 m. Those contour lines have been used to prepare the triangulated irregular network (TIN) which has been used to prepare the digital elevation model (DEM). The DEM has been used to generate the slope map where its values have been classified as follows: The values lying between 0 and 15% have been considered appropriate and the values between 15% and 30% have been considered inappropriate.
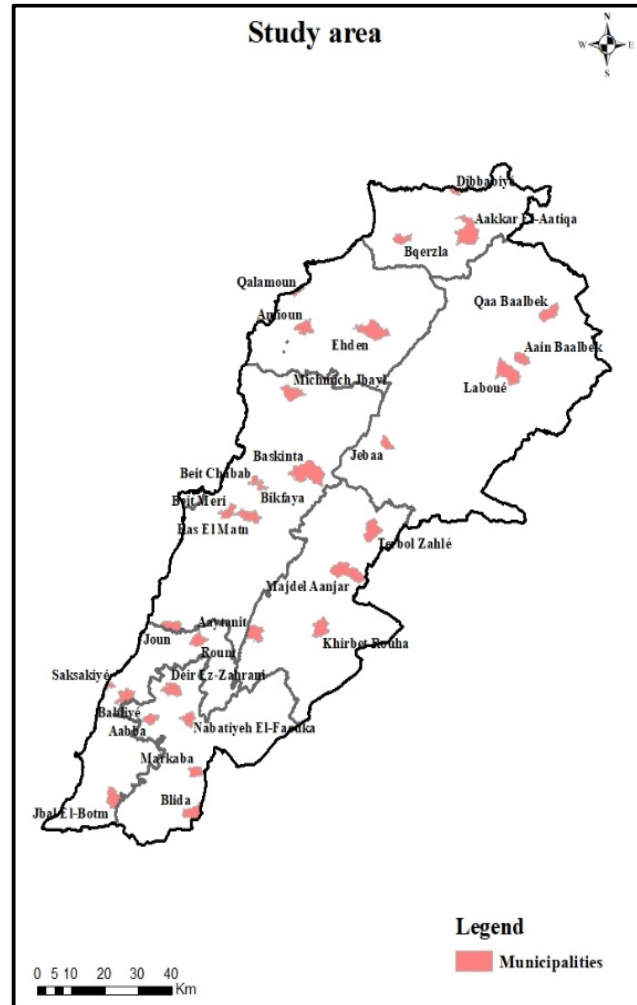

Fig. 1 Distribution of the selected municipalities on the Lebanese map

The precipitation

A huge amount of precipitation might be annoying for inhabitants since it may cause dangerous inundations which might result in decreasing the percentage of permanent residents. However, the existence of a reliable drainage system may reduce the risk of inundations. The precipitation map of the 30 municipalities has been obtained using the precipitation map of Lebanon. After obtaining the amount of precipitation for each municipality, the municipalities have been classified into two categories (appropriate and inappropriate) according

to the amount of precipitation and the existence of an efficient drainage system (Table II). Thus, a municipality is considered as appropriate if the amount of precipitation is inferior to 800 mm/year, or superior to 800 mm/year with an efficient drainage system. So, a municipality with an amount of precipitation superior to 800 mm/year and without an efficient drainage system is considered as inappropriate.

The Temperature

The extreme temperatureA are considered inappropriate for residence. Consequently, the temperature maps representing winter months (December, January, February and March) have been prepared using international temperature maps. The average temperature $T_m = (T_{max} + T_{min})/2$ has been chosen as a threshold to categorize municipalities as appropriate (T > T_m) or Inconvenient (T ≤ T_m) (Table II).

TABLE II
THE INDEX ATTRIBUTED TO EACH MUNICIPALITY FOR THE FOLLOWING FACTORS: PROXIMITY TO CONFLICT ZONES, PRECIPITATION, NUMBER OF SERVICES AND TEMPERATURE

Municipalities	Proximity to conflict zones	Precipitation	Number of services	Temperature
Deir al zehrani	5	1	3	5
Nabatiyeh Fawka	5	5	5	5
Aabba	5	5	3	5
Blida	1	5	1	5
Markaba	1	5	1	5
Saksakiye	5	5	5	5
Babliye	5	5	2	5
Roum	5	1	1	5
Jbeil al botom	1	5	1	5
Joun	5	5	3	5
Mechmech jbeil	5	5	1	1
Baskenta	5	1	6	1
Ras el maten	5	5	2	1
Beit chabeb	5	1	5	5
Bekfaya	5	1	2	1
Beit meri	5	5	3	5
Labwe	5	5	3	1
Al-Ain	5	5	9	1
Al-qaa	5	5	7	5
Jebaa	5	1	0	1
Amyoun	5	5	5	5
Eheden-Zgharta	5	5	16	1
Qalamoun	5	5	3	5
Debabiye	5	5	1	5
Akkar al atika	5	1	7	1
Bqerzla	5	1	3	5
Terbol zahle	5	1	2	1
Majdel anjar	5	5	7	1
Kherbet rouha	5	1	2	1
Aytanit	5	5	2	1

Proximity to Conflict Zones

The political stability is an important factor to be considered while deciding to stay in a zone or to leave it. Since Lebanon is inside a conflict zone (with Israel), the villages have been classified into "stable" and "unstable" according to their location with Israel. Indexes of 1 and 5 have been assigned respectively to stable and unstable villages (Table II).

The Number of Services

Regarding the number of services, the numbers of schools, industries and hospitals have been summarized in Table II.

TABLE III
THE CBD ASSOCIATED TO EACH MUNICIPALITY AND THE TIME TO IT

Municipalities	CBD	Time (per minutes)
Deir al zehrani		12 min
Nabatiyeh el Fawka	Nabatiyeh	0 min
Aabba		20 min
Blida	Bentjbeil	16 min
Markaba		36 min
Saksakiye		23 min
Babliye	Saida	21 min
Roum		24 min
Roum	Jezzine	12 min
Aytanit		45 min
Jbeil el botom	Sour	30 min
Mechmech Jbeil	Jbeil	33 min
Baskenta		51 min
Ras el maten		39 min
Beit Chabeb	Jdeideh	33 min
Bekfaya		30 min
Beit Meri		22 min
Terbol Zahle	Zahle	24 min
Majdel Anjar		32 min
Kherbet Rouha	Rashaya	18 min
Labwe		38 min
Al-Ain	Baalbek	45 min
Al-Qaa		60 min
Jebaa		28 min
Amyoun	Zgharta	30 min
Eheden-Zgharta		0 min
Qalamoun		15 min
Debabiye	Tripoli	58 min
Akkar al atika		89 min
Bqerzla		54 min

TABLE IV
THE RESULT OF DECISION TREE METHOD AND THE ATTRIBUTED WEIGHT FOR EACH FACTOR

Factor	Error	Weight (%)
Temperature	0.0011	50.8
Rainfall	0.0012	46.5
Time to CBD	0.03	1.9
Slope	0.1304	0.42
Number of services	0.2711	0.21
Conflict zones	0.4673	0.12

IV. RESULTS AND DISCUSSION

Using the decision tree method, the error related to each one of the following factors has been determined and the weight of each factor has been calculated using normalization method.

The findings of this study clearly show that the temperature (weight = 50.8%), the rainfall quantity (weight = 46.5%) and the time to CBD (1.9%) are the most influencing factors. Conversely, the slope, the number of services and the political stability have a minimal influence.

Consequently, the results of this study clearly show that the factors related to climate (temperature and rainfall) are the ones which affect mostly the permanent residence phenomenon. However, this study was limited by its inability

to evaluate the factors related to society or the life style of inhabitants.

V. CONCLUSION

The combination of remote sensing data and GIS tools has been used to evaluate the factors affecting the permanent residence in 30 Lebanese municipalities. Moreover, the decision tree method was efficient to combine different types of data in order to calculate the weight of each factor. The results of this study show that factors in relation with climate have the major effect on permanent residence phenomenon. As a result, the previously generated model can be developed and regulated to fit other regions and to include other factors taking into consideration that the determination of permanent residents is a crucial factor for an accurate estimation of population using remotely sensed data.

REFERENCES

[1] Hsu, S. Y., 1971, "Population Estimation," Photogrammetric Engineering, 37:449-454.
[2] Hsu, S. Y., 1973, "Population Estimation from ERTS Imagery: Methodology and Evaluation," Proceedings of the American Society of Photogrammetry 39th Annual Meeting, 583-591.
[3] 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.
[4] Langford, M. and D. J. Unwin, 1994, "Generating and Mapping Population-Density Surfaces within a Geographical Information-System," Cartographic Journal, 31(1):21-26.
[5] Harvey, J. T., 2002, "Population Estimation Models Based on Individual TM Pixels," Photogrammetric Engineering and Remote Sensing, 68(11):1181-1192
[6] Rindfuss, R., Stern, P., 1998, Linking Remote Sensing and Social Science: The need and the challenges, The National Academy Press, Available online at: http://www.csiss.org/SPACE/workshops/2006/OU/reading/Rindfussand Stern.pdf (Accessible: 5/5/2019).
[7] Dahan, Haim, Cohen, Shahar, Rokach, Lior, Maimon, Oded, 2014, Proactive Data Mining with Decision Trees (SpringerBriefs in Electrical and Computer Engineering) 2014th Edition, Springer.
[8] Lees, B., Ritman, K., 1991, Decision tree and Rule-Induction approach to integration of Remotely sensed and GIS data in mapping vegetation in Disturbed or Hilly Environments.
[9] Pal, M., Mather, P., 2003, An assessment of the effectiveness of decision tree methods for land cover classification, Remote Sensing of Environment, 86, pp. 554-565.
[10] Bogaert, J., Ceulemans, R., Eysenrode, D., 2004, Decision Tree Algorithm for Detection of Spatial Processes in Landscape Transformation, Environmental Management, Vol. 33, No. 1, pp. 62–73.
[11] Ballestores Jr, F., Qiu, Zeyuan, 2012, Proceedings of the International Academy of Ecology and Environmental Sciences, 2012, 2(2), 53-69.