

Marios Batsaris, PhD Candidate
Department of Geography, University of the Aegean
m.batsaris@aegean.gr

Dimitris Kavroudakis, Assistant Professor
Department of Geography, University of the Aegean
dimitrisk@aegean.gr

AUTOMATED POPULATION SPATIAL DISTRIBUTION ESTIMATION BASED ON STATISTICAL MODELING METHODS

Abstract

Due to regulations and restrictions census data are usually aggregated at the granularity level of blocks and are referred to night-time population. In some scientific fields such as location analysis, there is a need for more detailed spatial distribution of census data. Scientists have to transform population data from census predefined level of information (spatial/temporal) down to building level. Population distribution estimations in building level can be potentially useful in a number of applications such as disaster/operational management, locational analysis, urban planning and policy making. The aim of this work is the development of an automated population estimation for the city of Mitilini, Lesvos, Greece down to building level. We use statistical modeling methods along with data scraping approaches from web sources and hybrid geolocation techniques in order to populate a spatial database for further analysis. We present initial results of population downscaling at building level for the city of Mytilini and discuss the methodological followed for accurate population estimation. The results of this work are potentially useful to civil protection agencies for cases of natural disaster cases and emergency evacuation plans of the city.

Keywords: *Population Distribution, Statistical Modeling, Web Data Scraping, Geo-spatial Database*

Introduction

Population distribution is a dynamically changing factor that affects city dynamics. Information about the spatial and temporal distribution of a city's/town's population is crucial in many applications such as disaster management/planning and policy making. Each application requires population distribution information in different spatial and temporal levels of detail (scale).

This paper proposes a methodology that attempts to automate the down-scaling/redistribution process of the population of the city of Mytilini under different spatial and temporal scenarios (night and day). We utilize data scraping techniques from several web sources and statistical modeling methods correlated with land use/land cover information and building characteristics (projected area, number of floors) in order to estimate accurate population counts at building level.

Traditional field surveys for obtaining population counts by dwelling unit are resources expensive (time, budget). The methodological approach presented in this paper is relatively less resources expensive and faster and is also reproducible for any other city.

The rest of the paper is structured as follows. Population redistribution/down-scaling methods in Geographical Information Systems (GIS) and Remote Sensing (RS) literature are presented in the next section named Related work (section one). Section two includes the analysis of the methodological approach proposed. The study area and the data used are described in section three. Results are discussed in section four and finally this paper conclude with some concluding remarks.

Background

There are many applications that require population down-scaling such as disaster management, policy making, location analysis and others. Wu et al. (2005) reviews GIS and RS literature and divide population distribution estimation methods in two broad categories; a) areal interpolation methods and b) statistical modeling methods. This paper uses statistical modeling over areal interpolation methods because the latter estimates population density and is designed for zone transformation problems and the former estimates population counts and is designed to calculate inter-censal population or population difficult to encounter.

Mennis (2003) used ArcView GIS and Spatial Analyst extension tool to demonstrate two areal interpolation methods for generating population based surfaces of 100 meter resolution in Pennsylvania region. Freire and Aubrecht (2012) also used areal interpolation methods with ancillary information to model/map at high resolution the spatio-temporal population distribution in order to improve earthquake emergency management in Lisbon Metropolitan Area. On a different basis, Bakillah et al. (2014) explored the usage of Volunteered Geographic Information (VGI) and building footprints as ancillary data in areal interpolation methods to redistribute population at building level. Calka et al. (2017) proposed an improved volume-based method of high resolution population down-scaling that considers residential building information and other statistical parameters with a reliability of over 90%.

Another approach for estimating small area population estimates is the methodological approach of spatial microsimulation. This approach has shown some interesting applications of estimating household-based estimates for taxation, for education and for income related simulations (Kavroudakis et al., 2006, 2007, 2008; Kavroudakis, 2009; Kavroudakis et al., 2012; Kavroudakis, 2015).

On the other hand, statistical modeling methods correlated with building characteristics (count,

area, volume) extracted by LiDAR and Landsat TM data for small-area population estimation are presented in Dong et al. (2010). Statistical modeling methods were also used by Ye et al. (2012) to estimate the spatial distribution of population under different scenarios for earthquake emergency evacuation purposes. Alahmadi et al. (2016) applied both statistical and areal interpolation methods with ancillary information for population down-scaling in Riyadh, Saudi Arabia.

Methodology

This work deals with the urban fabric as a dynamically changing factor that affects and get affected by the population activities and mobility. Spatial analysis of the population distribution under different temporal scenarios is based on small scale land use/land cover information. This information is difficult to retrieve because in many cases it costs enough by means of time and money.

The proposed methodology is an attempt to semi-automate the process of (micro) simulation of the population mobility under different day-time/night-time scenarios and consists of two main tasks. Most of the population estimation related literature is based on areal interpolation methods either with ancillary information or without. This paper uses statistical modeling methods correlated with land use/land cover information and building characteristics. In order to accurately estimate population counts, we need to make some assumptions to reflect the custom behavior of some specialized population groups in the study area.

We assumed that population of Mytilini in both day-time and night-time scenarios population is being inside buildings and is divided in four categories:

- Individuals over 65 and above 4 years old are considered as *home-stayers* – 5832 individuals were distributed in residential buildings
- Students (University – School) are aged between 5 and 24 years old – 6139 individuals were distributed in Schools and Universities
- Working population is 9003 individuals were distributed in non-residential buildings
- The rest 4740 individuals are distributed in several POIs

GIS and Geocomputation techniques play very important role in the population modeling process. In the first task, *data scraping techniques* from several official and non-official web sources were implemented to acquire available/open information about businesses and work places, public services/agencies, Points Of Interest (POIs) and selling/renting properties as shown in Table 1. Scraped data are stored in a spatial database (PostgreSQL with PostGIS extensions by utilizing geo-coding techniques.

Table 1. Open information collection

Type	Information	Source
------	-------------	--------

Businesses and work places	Title Activity Location/address	Lesvos Chamber/Various Catalogues
Public services/agencies	Title Location	11888/Various Catalogues/OSM
POI	Title Location	OSM
Selling/Renting property	Location	Real estate agencies (Spitogatos.gr/Spitimou.gr, Xe.gr)

The presented methodology is based on statistical modeling methods to achieve inter-census population estimations in the granularity level of the building under night and day scenarios. The night-time population down-scaling and day-time population redistribution was based in equation (1).

$$p_{ijk} = \frac{(n_{ijk} \times s_{ijk}) \times P_{kj}}{S_{kj}} \quad (1)$$

i: building id

j: census block id

k: land use/land cover id

p_{ijk} : population of building *i* that belongs to *j* census block with *k* lulc

n_{ijk} : number of floors of building *i* that belongs to *j* census block with *k* lulc

s_{ijk} : projected area of building *i* that belongs to *j* census block with *k* lulc

P_{jk} : population of *j* census block

S_{jk} : total living space in *j* census block

Data and study area

The data used to accomplish scenario-based population modeling come from several resources. Population data and information are provided by 2011 general census of building and population-housing Greek Statistical Authority (ELSTAT – <http://www.statistics.gr/>). Due to restrictions and regulations about personal data census data are aggregated at the block spatial level of detail mainly in order to avoid individuals-targeting and identification. Information about land use/land cover of the urban fabric are acquired from Municipal business catalogs such as Lesvos Chamber and unofficial sources such as digital phonebooks such as 11888 and volunteering geographic information Database such as Open Street Map (OSM). Table 2 shows

the data used to spatially down-scale/redistribute population under day and night scenarios.

Table 2. Data format and sources

Data	Format	Source
Population divided by gender and age in building blocks 2011 – Census	Excel Table	ELSTAT
Digital Cartographic representation of building blocks 2011 – Census	Shapefile	ELSTAT
Digital Cartographic representation of building outlines incl floor number 2001 – Census	Shapefile	ELSTAT
Main occupation 2011 – Census	Excel Table	ELSTAT
Points of interest (business, museums etc)	Text/HTML/XML document	Lesvos Chamber
Sales/renting of residential property	Text/HTML/XML document	Spitogatos.gr, Spitimou.gr Xe.gr
Public services/agencies	Text/HTML/XML document	OSM

Study area

The city of Mytilini, Lesvos Island, Greece was selected as a study area to illustrate the proposed methodology mainly due to its size, extent and natural disasters experience. It has a population of approximately 25500 according to 2011 census and is the capital of Lesvos Prefecture and the North Aegean Province. Mytilini is also the port and airport of the Island of Lesvos and includes a significant portion of the economical, political, cultural and other resources of Lesvos Island.

Results

To estimate population distribution, statistical modeling methods were implied correlated with land use/land cover information and dwelling unit characteristics. Two population distribution scenarios were evaluated to demonstrate and validate the accuracy and the reliability of the

proposed methodology.

Figure 1 depicts the comparison of results obtained between the two scenarios. This is the spatial and temporal analysis of the population distribution. Residential population counts of the area of interest was successfully down-scaled and is presented in Figure 1 (right). The south part of Mytilini is more densely populated than the north part of the city because of the multi-apartments/multi-family houses and lack of non-

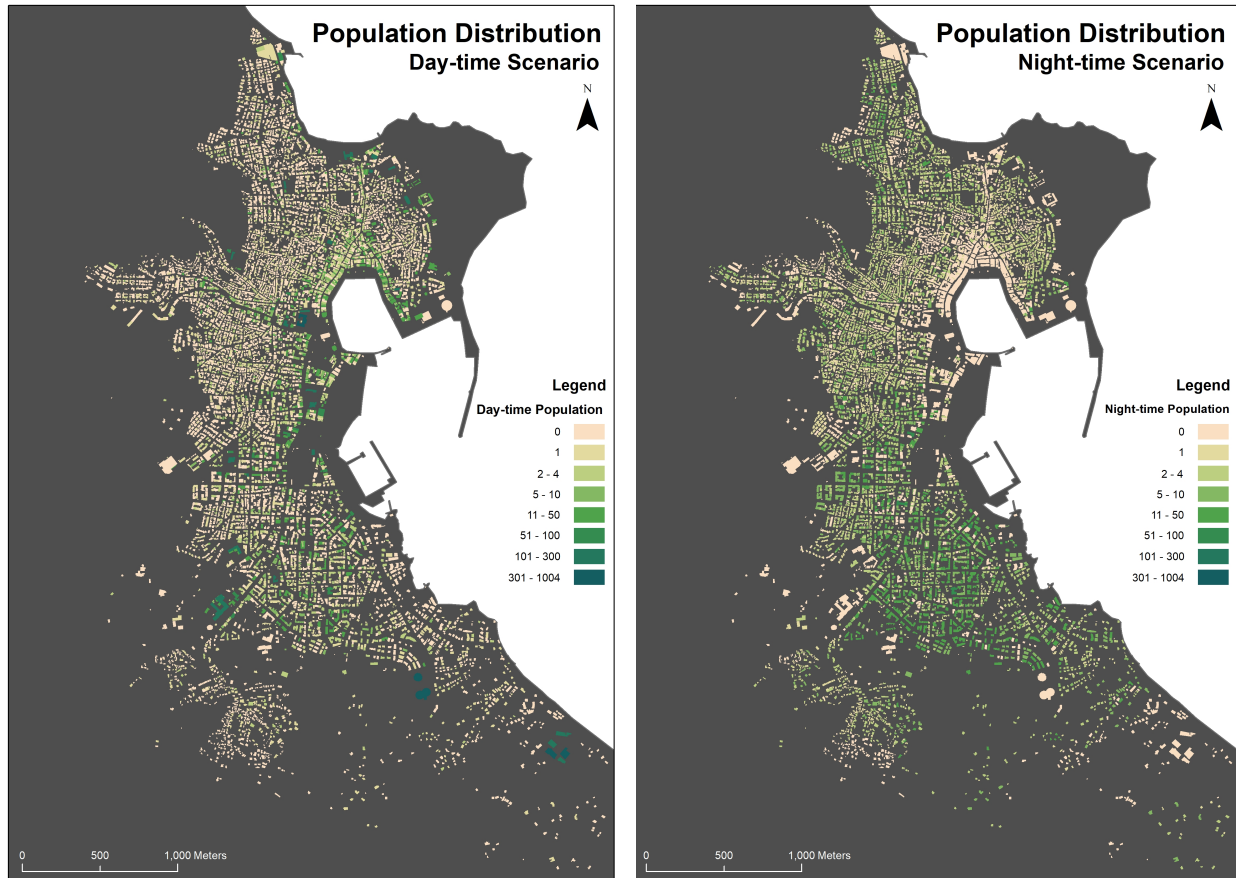


Figure 1. left: Day-time spatial Population distribution, right: Night-time (residential) population distribution

Residential dwelling units. The left map of Figure 1 illustrates the day-time population distribution in the city of Mytilini. The redistribution of the day-time population refer to land use/land cover categories different from residential. There is a significant difference between the two spatial distributions of population. From the one hand this difference is lying on the fact that the commercial activity is gathered in the city's down town and from the other hand

activities related to working environments or other services (University, Hospital, Public services and others) are casually distributed within the urban fabric.

Conclusion

The objective of the methodology presented in this paper is to estimate population counts in dwelling units under different scenarios according to building characteristics (projected area, number of floors) and land use/land cover data acquired from several web sources. Two population distribution scenarios were evaluated in order to test the speed, reliability and demonstrate the results of the proposed methodological approach. The results of this work are giving a meaningful insight about the spatial distribution of population in a medium size city in the Aegean Sea in Greece. We have not encountered similar results for Greece before.

The implementation of the methodological approach and the results obtained of the analysis of this paper indicate that web data scraping techniques combined with statistical modeling methods are capable of offering flexibility in terms of time and money and ensure the reliability of the results. An interest finding of this work is the run-time performance that shows that can be utilized also in large cities or even provinces/prefectures etc. Another advantage of this method is that can easily be repeated/automated at any time with zero costs.

Further expansion of this work could be the use of various other methods of population down-scaling and the in-situ evaluation of the results. The use of the results could be potentially useful for civil protection agencies mainly on the basis of emergency evacuation plans.

References

- Alahmadi, M., Atkinson, P.M., Martin, D., 2016. A Comparison of Small-Area Population Estimation Techniques Using Built-Area and Height Data, Riyadh, Saudi Arabia. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9, 1959–1969.
- Bakillah, M., Liang, S., Mobasheri, A., Jokar Arsanjani, J., Zipf, A., 2014. Fine-resolution population mapping using OpenStreetMap points-of-interest. *International Journal of Geographical Information Science* 28, 1940–1963.
- Calka, B., Nowak Da Costa, J., Bielecka, E., 2017. Fine scale population density data and its application in risk assessment. *Geomatics, Natural Hazards and Risk* 8, 1440–1455.
- Dong, P., Ramesh, S., Nepali, A., 2010. Evaluation of small-area population estimation using LiDAR, Landsat TM and parcel data. *International Journal of Remote Sensing* 31, 5571–5586.
- Freire, S., Aubrecht, C., 2012. Integrating population dynamics into mapping human exposure to seismic hazard. *Natural Hazards and Earth System Science* 12, 3533–3543.
- Kavrouidakis, D., 2009. Spatial microsimulation approach to the analysis of social and spatial

- inequalities in higher education attainment (PhD.). University of Sheffield, Sheffield, England, UK.
- Kavrouidakis, D., 2015. sms : An R Package for the Construction of Microdata for Geographical Analysis. *Journal of Statistical Software* 68.
- Kavrouidakis, D., Ballas, D., Birkin, M., 2006. Spatial microsimulation modelling for public policy analysis. In: 36th Annual Conference of Regional Science Association International: British and Irish Section, Jersey, Channel Islands,(Theme: Social and Public Policy)16 August-18 August 2006. Presented at the 36th annual conference of Regional Science Association International: British and Irish Section, Jersey, Channel Islands,(theme: Social and Public Policy)16 August-18 August 2006, 36th annual conference of Regional Science Association International: British and Irish Section, Jersey, Channel Islands,(theme: Social and Public Policy)16 August-18 August 2006, 36th annual conference of Regional Science Association International: British and Irish Section, Jersey, Channel Islands,(theme: Social and Public Policy)16 August-18 August 2006.
- Kavrouidakis, D., Ballas, D., Birkin, M., 2007. Spatial microsimulation: Enriching the Census of Population. In: 15th European Colloquium of Theoretical and Quantitative Geography. Presented at the 15th European Colloquium of Theoretical and Quantitative Geography, 15th European Colloquium of Theoretical and Quantitative Geography, 7-11 September 2007 in Montreux Switzerland.
- Kavrouidakis, D., Ballas, D., Birkin, M., 2008. Using Spatial Microsimulation for the analysis of social and spatial inequalities. In: Studying, Modeling and Sense Making of Planet Earth International Conference. Presented at the Studying, Modeling and Sense making of Planet Earth International Conference, Department of Geography, University of the Aegean, Mytilene Lesvos Greece, 1-6 June 2008.
- Kavrouidakis, D., Ballas, D., Birkin, M., 2012. Using Spatial Microsimulation to Model Social and Spatial Inequalities in Educational Attainment. *Applied Spatial Analysis and Policy*.
- Mennis, J., 2003. Generating Surface Models of Population Using Dasymetric Mapping*. *The Professional Geographer* 55, 31–42.
- Wu, S., Qiu, X., Wang, L., 2005. Population Estimation Methods in GIS and Remote Sensing: A Review. *GIScience & Remote Sensing* 42, 80–96.
- Ye, M., Wang, J., Huang, J., Xu, S., Chen, Z., 2012. Methodology and its application for community-scale evacuation planning against earthquake disaster. *Natural Hazards* 61, 881–892.