reviewed paper

Modelling Day and Night Time Population using a 3D Urban Model

Monika Kuffer, Richard Sliuzas

(Dipl.-Geogr. Monika Kuffer, ITC, University of Twente, P.O. Box 217,7500 AE Enschede, The Netherlands, m.kuffer@utwente.nl)
(Dr. Richard Sliuzas, ITC, University of Twente, P.O. Box 217,7500 AE Enschede, The Netherlands, r.sliuzas@utwente.nl)

1 ABSTRACT

Dasymetric methods are commonly used to redistribute or disaggregate (census) population data, using either simple binary or multi-layer models. Most models show limitations in high density built-up areas as they commonly ignore the 3D dimension (meaning buildings height) of multi-story urban environments. For example, simple dasymetric models only allocate the population counts to built-up areas, without considering differences between areas of multi-story and single-story buildings. Furthermore, such models only allow the disaggregation of 'night-time' population data, while for many urban applications such as transport, health or hazard, the location of 'day-time' population is of interest. This research presents an initial approach to model day and night-time population using as case study an Indian city (Kalyan-Dombivli). For most Indian cities, census population data is only available for wards, while day-time population data is either not available or of very poor quality. Besides census data and ancillary spatial data, this research uses a 3D urban model, extracted from Cartosat stereo-images. First, the extracted height from the stereo-image is used in combination with building footprints to disaggregate census population data at wards to 'night-time' population per building. Second, a classification of economically active areas is constructed based on the 3D urban model in combination with other spatial layers (e.g. transport layers) to model the day-time population. The result shows different concentration of population during day and night-time across ward boundaries as well as it confirms the potential of 3D data to disaggregate population data.

2 INTRODUCTION

2.1 Context

In the past years many researchers (e.g. Briggs et al., 2007, Langford, 2007, Maantay et al., 2007, Mennis and Hultgren, 2006 and Su et al., 2010) explored dasymetric methods to essentially disaggregate (census) population data, using either simple binary or multi-layer models. Due to high costs, low frequency and problems related to the aggregation levels of census population data the use of remote sensing for estimating and/or disaggregating population data have been explored since the 1950s (Liu, Clarke, & Herold, 2006). However, most models show limitations in high-density built-up areas as they commonly ignore the third dimension (meaning building heights) of multi-story urban environments (Aubrecht et al. 2009). Furthermore, such models commonly only consider the 'night-time' population (e.g. extracted by census counts or registers), while for many applications such as transport, health or hazard the day-time population would be important (Briggs, et al., 2007). Part of the day-time population is the workforce, which is sometimes captured by economic census statistics as number of employees per areal unit (job location). Most developed countries and some emerging economies (e.g. China) have information on job locations (either by economic census or public registers). However, in many countries the employment counts and therefore the location of jobs are not very accurate as they exclude often small and medium size companies, as well as the informal sector. Generally, in most developing countries and emerging economies employment statistics are of poor accuracy if available at all.

2.2 Population Data

Census statistics provide the amount of residential population at night-time (where people are living). Such census data are publically available only at aggregated administrative units, which are often very heterogeneous in terms of socio-economic characteristics. Furthermore, census wards can be partially built-up and non-built-up. Thus the main problems of census population data are that the administrative units used are commonly "geographically meaningless" (arbitrary) as well as they "smooth local variability" (Wu & Murray, 2005). Furthermore, modifications of (ward) boundaries are happing frequently, making temporal analysis difficult (Langford, 2006). To avoid such problems several researches (e.g. Mennis and Hultgren 2006, Maantay et al. 2007) have been conducted to disaggregate population data.

The concept of day-time population is rather complex because of the complex daily agenda of people. An important source to locate day-time population is the location of jobs. Such data can be extracted in some countries from business or economic census. An economic census is also conducted in India, but data have severe problems with full coverage and temporal accuracy. Such problems even exist in countries like the USA, where the economic census is held every 5 years. Micarelli (1998) indicates two major problems of the economic census; first, not all companies are included and second, the self-reporting is prone to bias. Thus, for many countries employment locations are not available or of very limited accuracy, as well as models covering the complete day-time population (including various daily activities beyond labour) hardly exists.

2.3 Dasymetric Methods

Dasymetric mapping is a technique for disaggregating spatial data (e.g. census ward population) from larger to smaller spatial units using ancillary data (e.g. land use data). These techniques have been employed since more than 200 years (Maantay et al. 2007). The results of dasymetric methods aim at more homogeneous areas and therefore at better representation units of the modelled phenomena, e.g. avoiding population data to be allocated to areas that are not inhabited. Dasymetric models avoid displaying data with arbitrary (census) boundaries, which are often very heterogeneous and therefore a poor representation level. Higgs and Langford (2009) stated that dasymetric techniques have a great potential for a more "accurate representations of the spatial distribution" of population data. For instance, Holt et al. (2004) used Landsat images to extract land-use/-cover data to disaggregate population densities using dasymetric methods.

Commonly used binary dasymetric methods divide an area into inhabited and non-inhabited and assign the population (of a census ward) to the inhabited areas. Whereas multi-class weighted dasymetric model use a subdivision of areas into additional categories which are e.g. related to different population densities using additional data layer e.g. land use, zoning, land value, accessibility, infrastructure density, home living style (Su, et al., 2010). Common methods to disaggregate population data are according to Biggs et al. (2007), Langford (2006), Maantay, et al. (2007) the following:

- Simple (e.g. binary) and filtered area-weighting/interpolation, which is often used to compare and assess more advanced dasymetric techniques (as benchmark),
- Global regression, which builds a global relationship between e.g. population data and land cover information,
- Regional regression, which uses independent fitting for different zones,
- Multi-class dasymetric method, which uses different density zones based e.g. on socio-economic variations, different housing types,
- Dasymetric mapping in combination with an expert system using e.g. cadastral data.

Most of these techniques (with the exception of the expert system) show limitations in high density built-up areas with multi-story buildings as population is simply redistributed to built-up areas without considering the building height. Therefore Bajat et al. (2011) suggests the use of Lidar data to allow modelling the 3D dimension, which is believed to have the potential to improve population mapping. Aubrecht et al. (2009) showed that building footprint and Lidar data in combination with company information gave the best result to disaggregate population data. However Lidar data are for many developing countries and emerging economies not available due to high costs and other data access restrictions. Therefore, this paper explores the use of high resolution stereo-images (Cartosat) for dasymetric mapping. Stereo-images (e.g. Cartosat-1 data) are easily available at much lower costs compared to Lidar data. Thus, the focus of the research is on disaggregating population data to the spatial unit of individual buildings by taking the building height into account. Further, the population estimation per building is used to generate day and night-time population density maps that do not use census ward boundaries.

3 METHODOLOGY

3.1 Data set and case study area

The city of Kalyan-Dombivli, one of the major urban agglomerations of the Indian state of Maharashtra, is situated in close proximity to Mumbai. The twin city has according to the 2011 census about 1.2 million inhabitants. Large parts of the city are densely built-up, with vast areas of slums (44% according to the city development plan).

Sensor/layer	Date	Attributes/spatial resolution
Resourcesat	12.11.2007	5.8 m
Cartosat-1 (stereo pair)	26.02.2008	2.5 m
Building footprints	2007/8	Use of buildings, partially building height
Ward boundaries	2001	Census ward ID
Population census	2001	Population and workforce per ward
Road network	2007/8	Main and minor roads
Public transport stops	2007/8	Location of bus and railway stops

Table 1: Spatial and census dataset

Building footprints were available from the KDMC (Kalyan Dombivli Municipal Cooperation). However, for many slum areas not individual buildings are digitised but a slum area is delineated as an entity. Reason for not having individual buildings of slums is most slums are of such high density that even a visual delineation of individual buildings is difficult. Building footprints are classified into 3 main use classes (residential, mixed and economic use). The footprint data were checked and updated (e.g. omitted buildings) using Resourcesat and Cartosat images (Table 1). A digital surface model was created using a Cartosat-1 stereo pair (2.5 m resolution) using Erdas LPS. This dataset was available from an earlier research (Mishra et al. 2011). Ancillary spatial data were available, including the road network and location of public transport stops. In addition to the spatial data, census statistics from 2001 were available (unfortunately the more recent census of 2011 is not available at this moment). The census gives the number of inhabitants and total workforce per ward. The temporal consistency between the spatial and statistical data is not optimal. Thus, a central ward with little change is selected to develop the model (ward No. 17) (Figure 1).

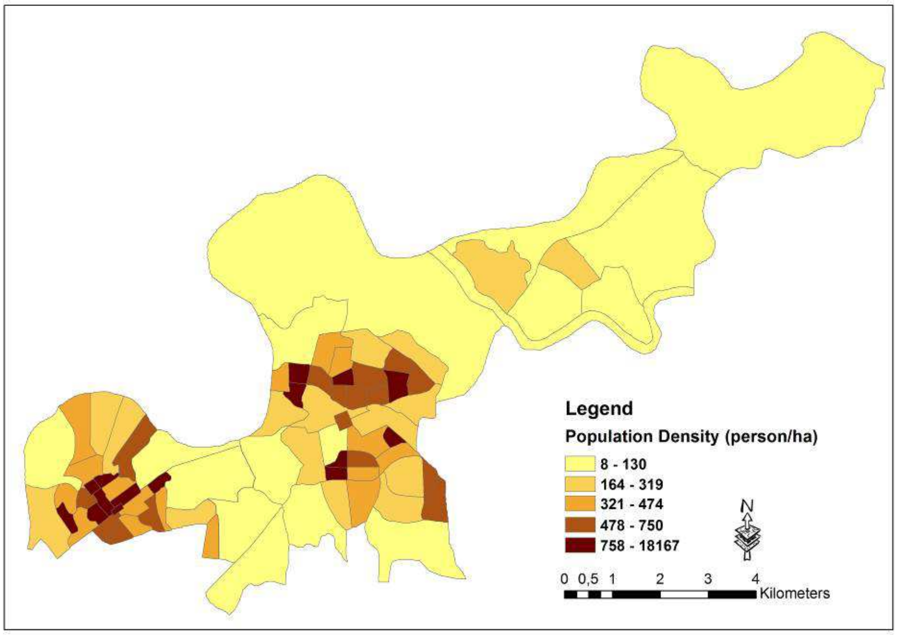

Figure 1. Population density (person/ha)

Figure 1 shows the population density distribution for the entire city of Kalyan Dombivli using the ward boundaries. Many of the wards are a mix of built-up/non-built-up. Moreover, in many wards a mix of single versus multi-story buildings can be found.

3.2 Methodology

In order to disaggregate the census population data to the units of individual buildings the model is set-up in two major steps (Figure 2). In a first step, the night-time census population (available per ward) is disaggregated using the height information in combination with the building footprints. Here the assumption is made that an average floor area ratio is occupied per person. This assumption has two major limitations, first it does not differentiate between different socio-economic areas (e.g. slum and non-slum areas) (see Figure 3); second it assumes that people stay at home during night. Thus, for extracting the night time (census) population the average height per building is modelled using the height extracted from the Carotsat-1 stereo-pair (see figure 4). In order to calculate the number of floors, the assumption is made that in average 3 meter height represents one floor. This allows estimating the number of floors per buildings. In a next step a ratio is built by dividing the floor area of a specific ward by the number of inhabitants (of the ward). The result gives an average m²/person ration. This ratio is applied to the individual buildings to estimate the number of persons per building.

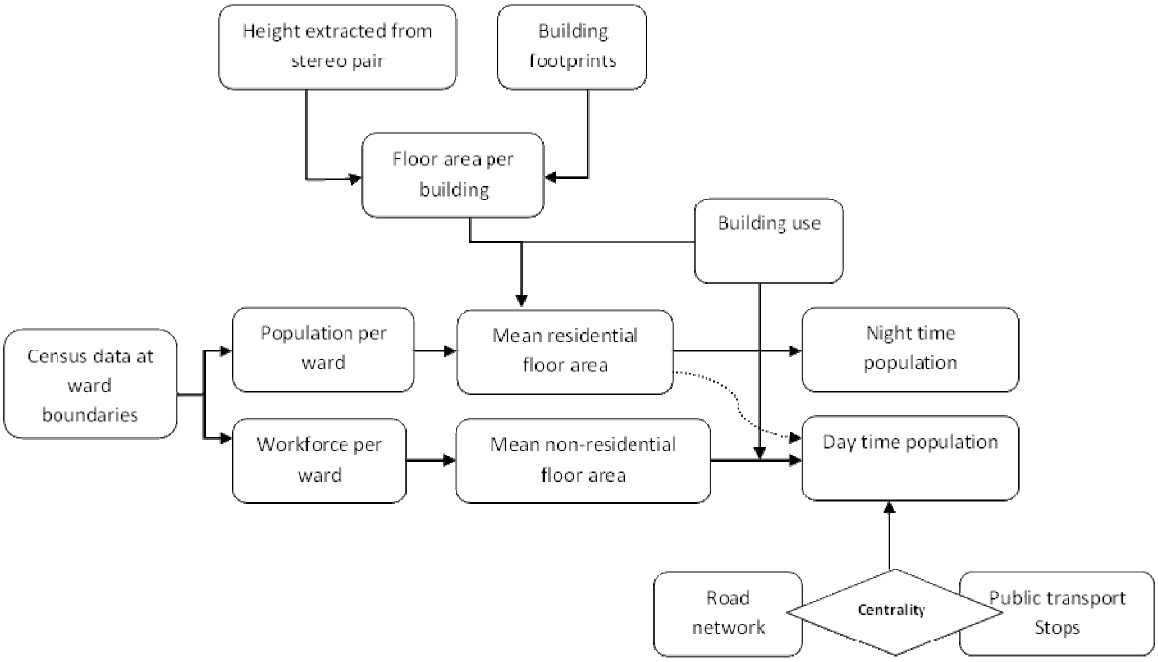

Figure 2. Overview of Methodology

The second step models the day-time population. This includes extracting buildings where economic activities are performed and redistributing the workforce to areas of such activities, as well as modelling the location of the 'remaining population' (people not counted as workforce).

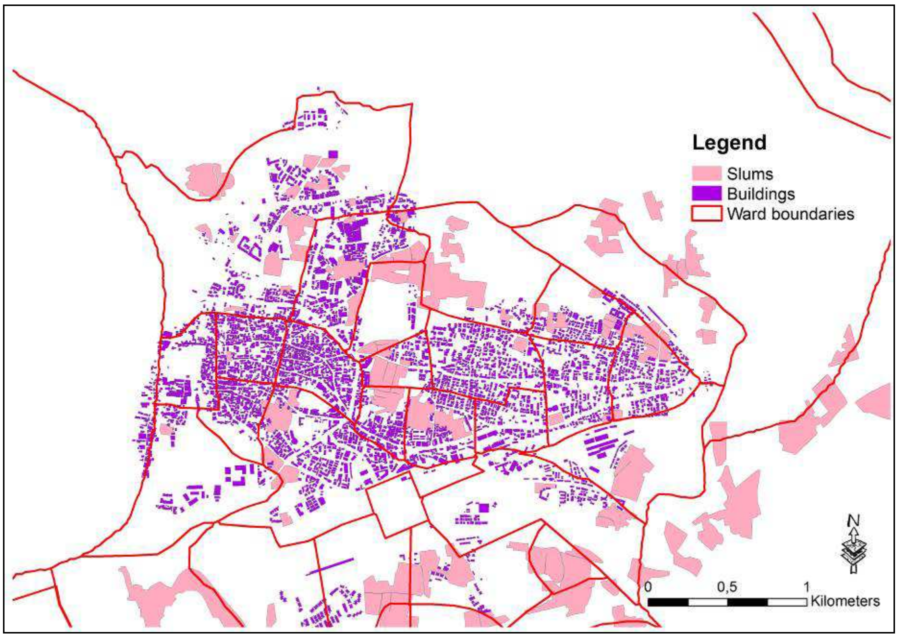

Figure 3. Buildings and slum areas of part of Kalyan-Dombivli

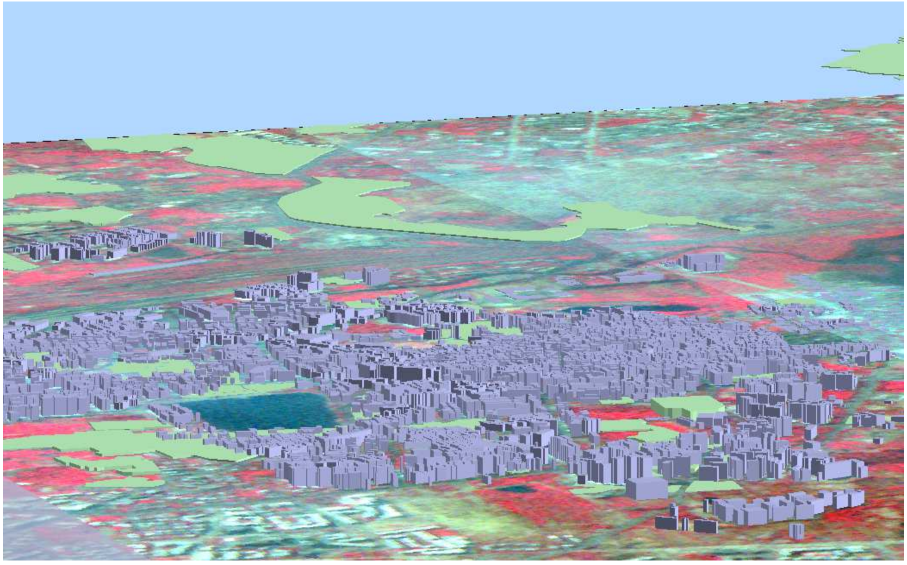

Figure 4. Extruded buildings using the height information from Cartosat-1 images (displayed on the Resourcesat image)

Thus, for extracting the day-time population the data available from the census is used. The data give the workforce and non-workforce population. The registered workforce population is redistributed towards areas of economic building use. Here the module works again with an economic floor area/workforce ratio in m2 per employee. The remaining day-time population (not part of the workforce) is redistributed to residential buildings as well as to buildings that are,

- located along roads,
- located in proximity to public transport modes.

The reason for including besides the buildings classified as economic use also other buildings is that some inhabitant will be at their residence but also to include economic and other activities that are not included into the workforce statistic. For example informal economic as well as shopping and leisure activities commonly occur along roads, public transport points and at central locations. Therefore, this is also an attempt to include informal economic activities into the model. Therefore buildings with high likelihood of having economic activities in its surrounding are selected using the above mentioned criteria. This part of the day-time population (which is not registered as workforce) is assigned to buildings with high likelihood of economic activities on the street. The remaining population is assumed to be at residential buildings. This assumption is again a limitation of the model as optimally also other activities e.g. educational activities need to be included.

4 RESULTS AND DISCUSSION

The results have two major outputs, first a model of the night-time population per building and second a model of the day time population per building. In a final step population density maps (day and night time) are generated that do not rely on ward boundaries.

Difference in Floors (reference – extracted)	**Frequency**	**Percentage**
3+	92	4
2	91	4
1	90	4
0	774	33
-1	638	28
-2	329	14
-3+	301	13
SUM	2315	100

Table 2. Extracted floor number compared with municipal reference data

ISBN: 978-3-9503110-6-8 (CD-ROM); ISBN: 978-3-9503110-7-5 (Print)
Editors: Manfred SCHRENK, Vasily V. POPOVICH, Peter ZEILE, Pietro ELISEI

4.1 Modelling the height of buildings

The extracted surface model using the Cartosat-1 (stereo pair) allows an estimation of the building height information. Using an average floor height of 3 m, the number of floors per building is calculated. The results are compared with a municipal data set containing for parts of the buildings the number of floors (unfortunately without information about temporal consistency). Comparing the extracted number of floors with the reference, 33% have an optimal match, while 64% of the buildings have a match of +/-1 floors. The major problem of the extracted floor heights using the stereo module is an overestimation of the number of floors (Table 2).

4.2 Modelling the night and day time population per building

For the example ward No. 17 (displayed in figure 4) the ratio used for calculating the night time population is 19.5 m2 floor area per person. Thus lager numbers of night-time inhabitants are allocated to multi-story buildings while buildings of fewer floors have fewer inhabitants (Figure 5).

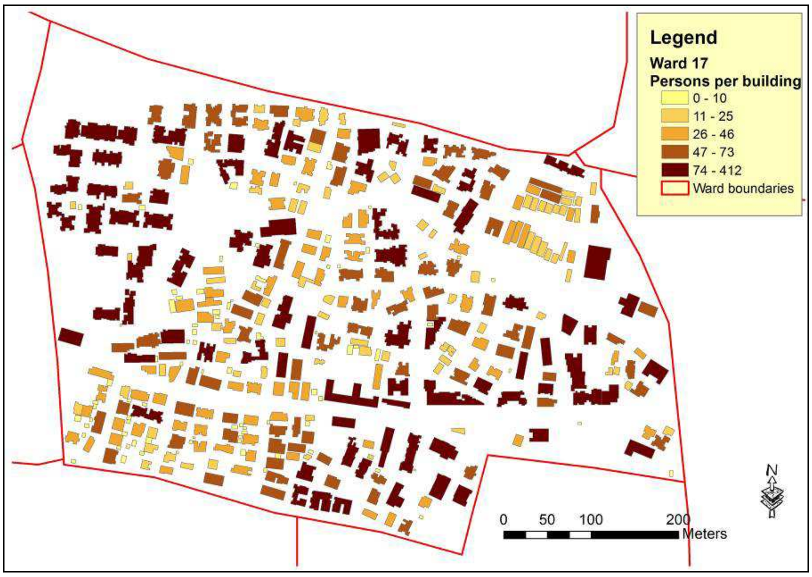

Figure 5. Model of night time (census based) population distribution

The ratio for ward 17 to allocate the workforce is 16 m2 floor area per employee. The day-time population model shows higher number of population in areas of economic activities and along major roads while part of the population (non-workforce) is allocated to residential areas (Figure 6). Validation of both models is still missing.

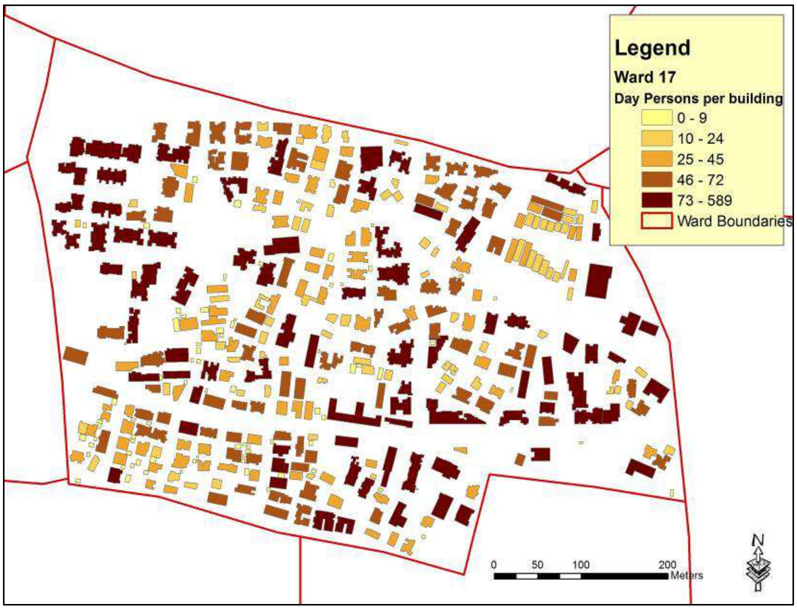

Figure 6. Model of day-time population

4.3 Modelling the night and day time population density

In order to illustrate the different day and night time density distribution of the ward No. 17 (a high density area), kernel density maps (weighted by the number of inhabitants per building) are generated. The resulting maps of night time (Figure 7) and day-time (Figure 8) density distribution illustrates the heterogeneity of the ward, which is commonly neglected by displaying density maps aggregated at ward boundaries. Moreover, the density maps illustrate differences between night and day-time, where during day-time more persons can be found in economically active areas.

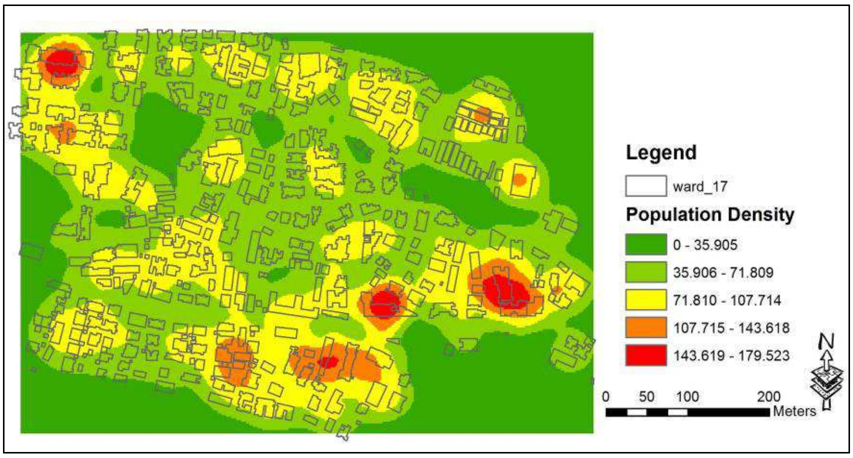

Figure 7. Night time population density distribution of ward No.17

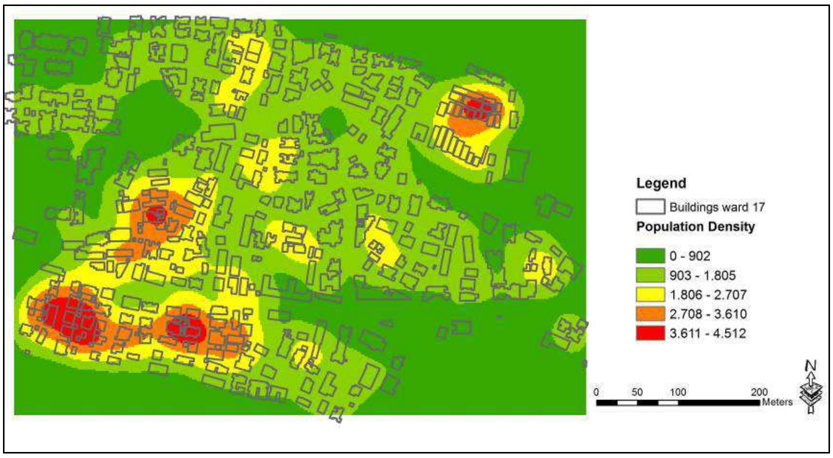

Figure 8. Day time population density distribution of ward No.17

One visible problem of the example ward No. 17 in the case of the night time population is that an area covered by ‘Chawls’ (small multi-storied dwelling units used for workers) in the North-East part is accounted with the same ratio of m2/persons as better-off areas. This definitely allocates too less people to the Chawls area (slum like conditions). However besides this problem (planned to be addressed in upcoming research) the results give a general impression of difference between day and night-time population density distribution.

For a selected larger part of the city of Kalyan-Dombivli the difference between day and night-time population was displayed across ward boundaries (Figure 9 and 10). This illustrates the advantage of not aggregating data at ward boundaries, which allows displaying cross-ward clusters of high versus low population densities. The selected area is covering a relatively stable part (with little development dynamic) to minimize problems of temporal inconsistency of the data used for this research.

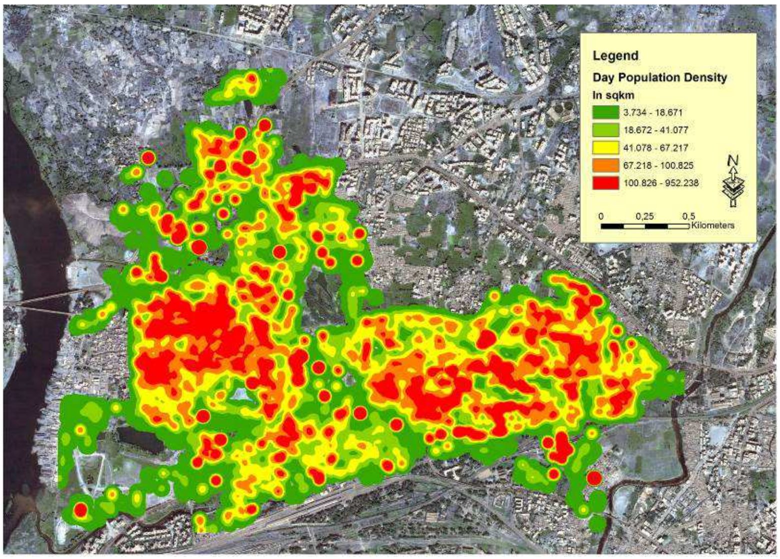

Figure 9. Day time population density distribution of part of Kalyan-Dombivli

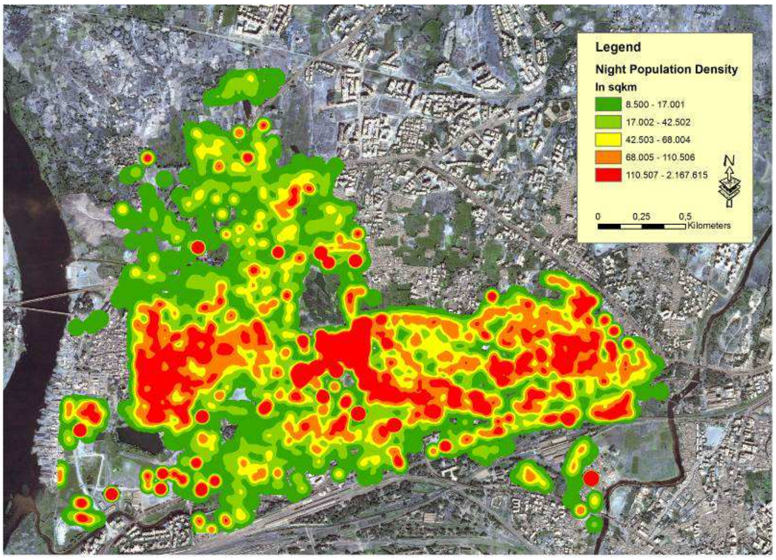

Figure 10. Night time population density distribution of part of Kalyan-Dombivli

Several difference between the day and night-time population distribution can be observed, e.g. some areas in the North-West have higher day time densities caused by the location of areas of economic use, while some of the residential areas have during daytime less density. Areas of high density during day and night time were in particular densely built-up residential areas, which includes in particular areas of multi-story buildings and slum areas.

5 CONCLUSION

Models that address differences between day and night-time population at disaggregated level are not much found in published research. Such models are of high complexity caused by the complex daily activity patterns of urban population, to model this complexity is a challenge. This research aimed at a first exploration of the use of height information in such models coming from high-resolution stereo-images. The presented research output illustrated that such height information allowed disaggregating night-time population data (census) to individual buildings using here a simple m2/person ratio for the individual wards. Day-time population models are more complex as urban inhabitants have diverse activities patterns. Part of the population is working while others participate in educational activities, shop or stay at home etc. The presented model used a simple approach, dividing the population in registered working and non-working population and redistributing the working population into areas of economic use. While the “officially” non-working population was partially assigned to areas of economic activities (streets and central location) and partially to residential locations.

This research aims at opening a first thought towards providing more temporal detailed and disaggregated population data which is of high demand as input information for urban models (e.g. transport models). Relatively simply built and available stereo-images have a potential to be used in areas were more accurate Lidar data is not available for extracting building height information. An accuracy of 64% of +/-1 floor could be achieved; this allowed distinguishing between single and multi-story buildings within the limitation of the obtained accuracy. The extracted model of day and night-time population for a central part of the study area gave an indication of density variations without being limited by ward boundaries, while including also the 3rd dimension (showing e.g. high densities in areas of multi-story buildings).

The future research focus will be on refining the day-time model of population distribution allowing the modelling of other activities as well as the refinement of the night-time population by considering areas of different socio-economic classes. Also validation of this simple model presented in this paper needs to be still conducted.

6 REFERENCES

ALARCÓN, D. & ZEPEDA, E., 1998. Employment trends in the Mexican manufacturing sector. The North American Journal of Economics and Finance, Vol. 9, Issue 2, pp. 125-145.

AUBRECH, C., STEINNOCHER, K., HOLLAUS, M. & WAGNER, W., 2009. Integrating earth observation and GIScience for high resolution spatial and functional modeling of urban land use. Computers, Environment and Urban Systems, Vol. 33, Issue 1, pp. 15–25.

BAJAT, B., HENGL, T., KILIBARDA, M. & KRUNIĆ, N., 2011. Mapping population change index in Southern Serbia (1961–2027) as a function of environmental factors. Computers, Environment and Urban Systems, Vol. 35, Issue 1, pp. 35-44.

BRIGGS, D. J., GULLIVER, J., FECHT, D. & VIENNEAU, D. M., 2007. Dasymetric modelling of small-area population distribution using land cover and light emissions data. Remote Sensing of Environment, Vol. 108, Issue 4, pp. 451-466.

HOLT, J. B., LO, C. P. & HODLER, T. W., 2004. Dasymetric estimation of population density and areal interpolation of census data. Cartography and Geographic Information Science, Vol. 31, Issue 2, pp. 103-121.

LANGFORD, M., 2006. Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method. Computers, Environment and Urban Systems, Vol. 30, Issue 2, pp. 161-180.

LANGFORD, M. 2007. Rapid facilitation of dasymetric-based population interpolation by means of raster pixel maps. Computers, Environment and Urban Systems, Vol. 31, Issue 1, 19-32.

LIU, X. H., CLARKE, K. & HEROLD, M., 2006. Population density and image texture: A comparison study. Photogrammetric Engineering & Remote Sensing, Vol. 72, Issue 2, pp. 187–196.

MAANTAY, J. A., MAROKO, A. R. & HERRMANN, C., 2007. Mapping population distribution in the urban environment: The cadastral-based expert dasymetric system (CEDS). Cartography and Geographic Information Science, Vol. 34, Issue 2, pp. 77-102.

MENNIS, J. & HULTGREN, T., 2006. Intelligent dasymetric dapping and its application to areal interpolation. Cartography and Geographic Information Science, Vol. 33, Issue 3, pp. 179-194.

MICARELLI, W. F., 1998. Evolution of the United States economic censuses: The nineteenth and twentieth centuries. Government Information Quarterly, Vol. 15, Issue 3, pp. 335-377.

MISHRA, S., MARTINZ, J., KUFFER, M. & PFEFFER, K., 2011. An exploration of natural capital in the context of multiple deprivations, presented at the JURSE 2011, Munich (Germany).

SU, M.-D., LIN, M.C., HSIEH, H.I., TSAI, B.W., & LIN, C.H., 2010. Multi-layer multi-class dasymetric mapping to estimate population distribution. Science of the Total Environment, Vol. 408, Issue 20, pp. 4807-4816.

WU, C., & MURRAY, A. T., 2005. A cokriging method for estimating population density in urban areas. Computers, Environment and Urban Systems, Vol. 29, Issue 5, 558-579.