CALCULATION OF A NIGHTTIME-POPULATION MAP BASED ON FREE DATA

Morten Fuglsang
SWECO
MortenWinther.
Fuglsang@sweco.dk

When the first authoritative data was released, free access was given to registries like the building registry (BBR). However, statistical data like the nighttime-population is still not free. This analysis shows how we can use dasymetric mapping to dis-aggregate municipality population data into BBR housing units, producing an estimated nighttime-population map. The results produced show that the proposed model has a general overrepresentation of the population, but despite that, this free dataset calculated here, should be usable in many cases where an estimate is adequate.

Keywords: dasymetric mapping, nighttime population, registry data, housing units.

When the first authoritative data was released in 2011, free access was given to registries like the building registry (BBR) and the company registry (CVR). This has meant that the use of these data has increased in many different companies, helping them to make informed decisions. With this increased use, a desire to be able to use population data has grown. Unfortunately, the first release did not include data from the statistical office in Denmark (DST), which still required payment.

This paper will focus on the calculation of a nighttime-population dataset. Nighttime-population is defined as the number of inhabitants that resides at a location – in contrast to the daytime-population map stating where people work. The mapping effort here will focus on calculation of point data, with results presented summarized in the standardized Danish grid net. This format has become popular as agglomerative unit for socioeconomic data in recent years, and is therefore a well proven unit of measure.

A country-wide nighttime-population dataset from DST is currently valued at about 59.000 dkk, which is a significant price to pay in order to use

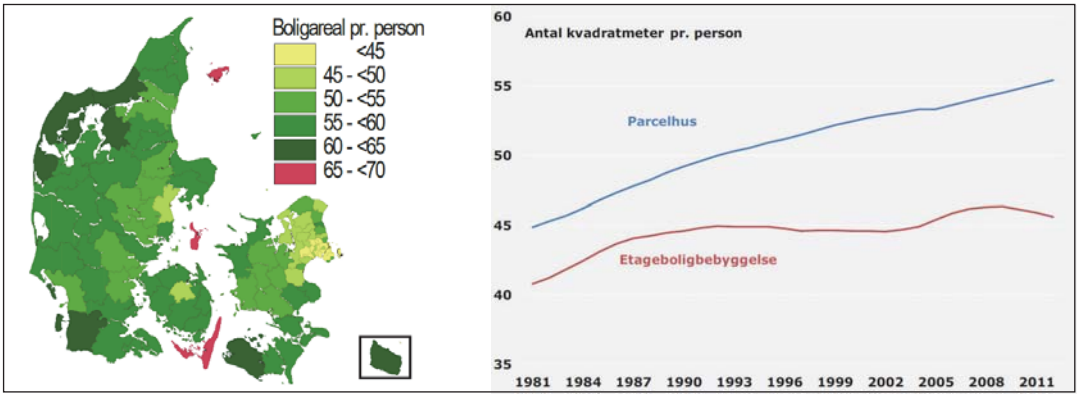

Figure 1. Area pr. individual from DST 2011 (Left), and Development of area pr. Individual from BRF (Heining 2013) (Right).

these data in small and medium scale companies. The method proposed here uses the municipality population statistics – which is freely available from DST, combined with the housing-unit table from BBR. Based on dasymetric mapping the population of each municipality is distributed into the housing units. The result will be evaluated against the official nighttime-population map from DST in order to assess the usability of the map produced.

THEORY

The conceptual idea behind the conducted analysis was based on the calculation of how many square meters of living space each individual statistically have in different municipalities. This has been investigated in many different cases. The case that formed the idea for this analysis was from DST in 2011. The topic has remained in focus since – as late as in September 2016 by Gadberg 2016, where it is documented that the area pr. Individual has increased by about 20% since 1981.

The work conducted by (DST 2011) focuses on the calculation for each municipality individually. This was determined as the offset for this analysis, since it was shown by (DST 2011), that there are quite large regional differences.

The theoretical framework for the analysis was based on the concept of dasymetric mapping. This is a tool for dis-aggregating data, creating more detailed choropleth maps by distributing observations into smaller classes (Tapp 2010). It is an often used framework for distributing population measurements, examples include research by (Sleeter & Gould 2007) and (Eicher & Brewer 2001).

$$1.\hat{y}_{t=\sum_{s=1}^{n}\frac{y_s A_{s\cap z}}{A_s}}$$

Where :
$\hat{y}_t$= Estimated count in target zone
y_s= Count of the source zone
$A_{s\cap z}$= Area of intersection between source and t
A_s= The area of the source zone
n = The number of zones which z overlaps
(Mennins & Hultgren 2006)

The main mathematical expression of the concept can be seen in eq.

The classic result of a dasymetric calculation can be seen in figure 2, where the rough population in districts is dis-aggregated into smaller units based on land use types.

The analysis extends the concept, by not looking at land use zones as targets for the dis-aggregation, but on housing units which are located by their address points. We use the unit table from BBR, where relevant information about usage, size etc. are present. In order to only get units where people live, we disregard industry, service and other classes, and only use the types specified in table 1.

The work conducted here is as a combination of the dasymetric methodology, the statistical measure of 'square meters pr. Individual' and the

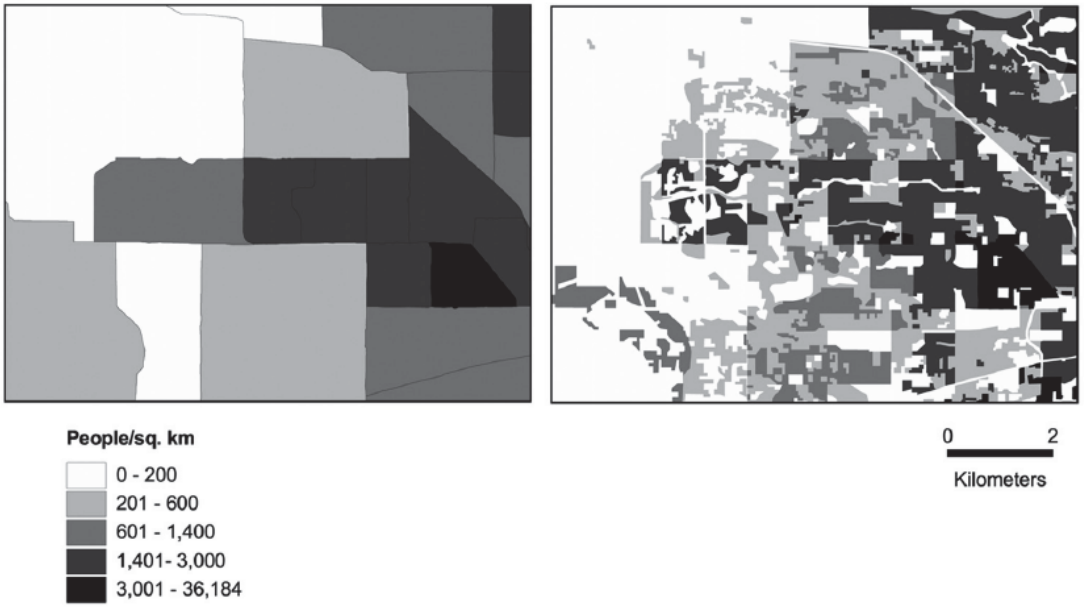

Figure 2. Standard dasymetric analysis result, from general districts to smaller units of measure. (Mennis & Hultgren 2006).

110	Stuehus til landbrugsejendom
123	Fritliggende enfamilieshus (parcelhus).
130	Række-, kæde- eller dobbelthus (lodret adskillelse mellem enhederne).
140	Etageboligbebyggelse (flerfamiliehus, herunder tofamiliehus (vandret adskillelse mellem enhederne)).
150	Kollegium
160	Døgninstitution (plejehjem, alderdomshjem, børne- eller ungdomshjem).
190	Anden enhed til helårsbeboelse

Table 1. BBR usage codes used for the calculation (BBR.dk 2016).

housing registry. When this is put together, we create an estimated nighttime-population map based solely on open data.

METHODOLOGY

The first task of conducting this analysis was to gather the relevant data. The population data was acquired through 'Statistikbanken' from DST, where the series 'FOLK1A: Folketal den 1. i kvartalet efter område, køn, alder og civilstand' was downloaded and stored in a PostgreSQL database for further processing. Secondly, the housing unit table was collected from Sweco, where these data are available with a weekly update cycle. Some auxiliary data was required for conducting the calculation and preparation of the final product – this was the municipality map from 'Danmarks Administrative Geografiske Inddeling' (DAGI), and the standardized Danish grid net (DST 2012) created with the 'Gridfactory' application. After collecting these data, it was all loaded and stored in the database, and processing could begin. This was computed in the following steps:

- The total sum of housing square meters for the selected codes was calculated for each municipality – providing a total sum pr. municipality.
- This municipality sum was then divided by the population of the municipality, based on

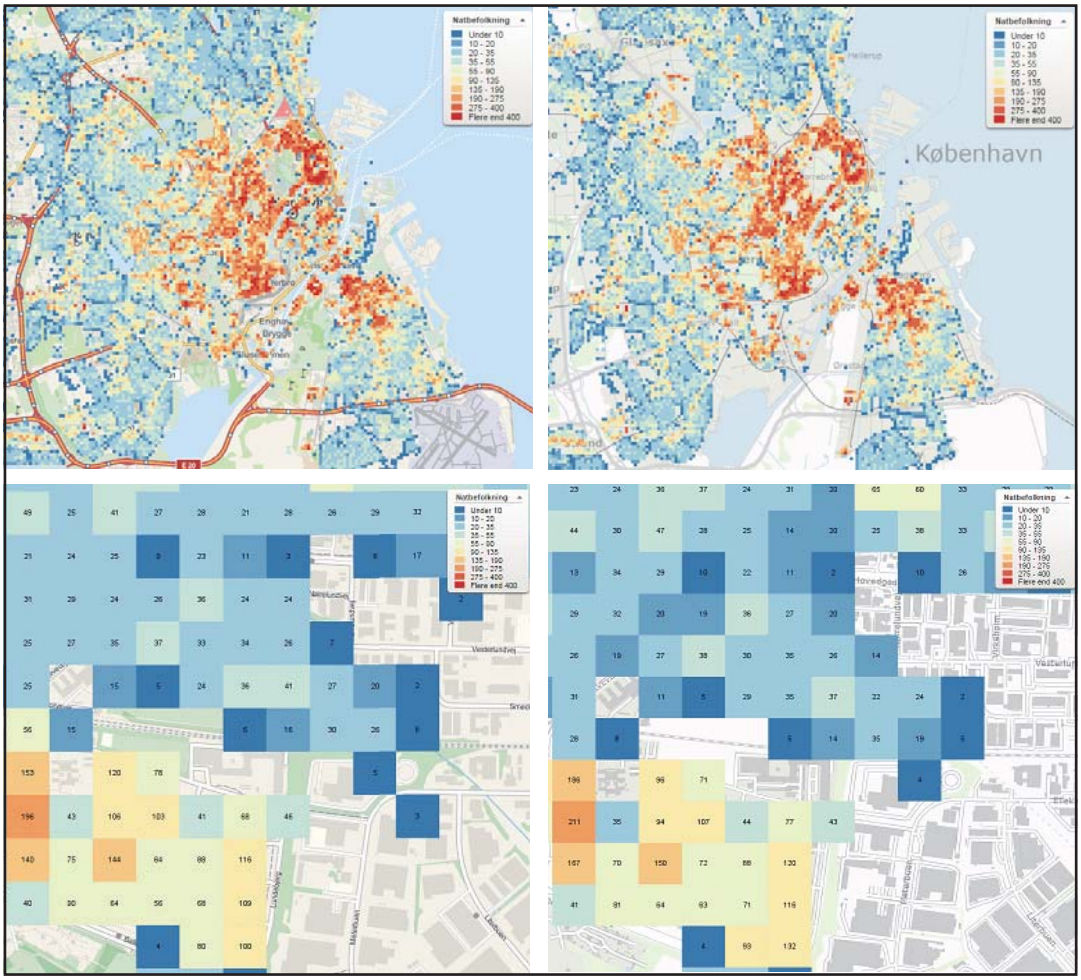

Figure 3. DST night population (Right column), calculated results (Left column).

the statistics from DST. This calculates the square meters pr. Individual in each municipality.

- The calculated value pr. Individual was then compared to each housing unit. If the value pr. individual was 50 square meters, and a given house was 120 square meters, the model assigns 2,4 individuals to the house.
- Using these values calculated for each house, the total pr. Grid cell is summarized using a spatial join between the point based house data and the grid-net. This creates a total population pr. Grid cell result, that is comparable to the 'Natbefolkning' product from DST.

RESULTS

In figure 3, a comparison of the obtained results from the dasymetric mapping analysis and the official nighttime-population from DST is presented. Data is classified into the same classes between the two datasets, making them visually comparable. It can be seen, that the patterns observed, are quite equal between the two sets of maps.

DISCUSSION

In order to analyze the accuracy of the calculation, the official nighttime-population map from DST was used. There is about 8 months of time difference between the DST map and this calculation, meaning

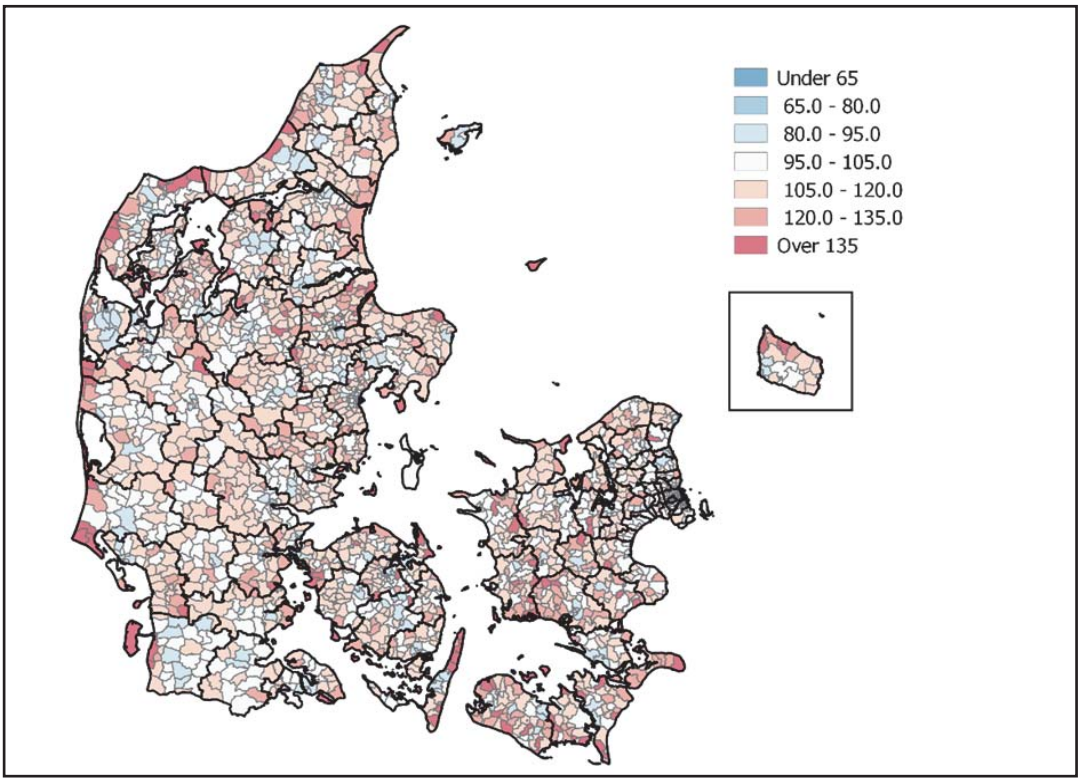

Figure 4. Accuracy assessment – low values indicate underrepresentation in the result, high values an overrepresentation.

that the results are produced using more recent data, than what was used for the DST map.

Since the calculation uses municipality totals, calculation of uncertainty at this level is useless – The accuracy pr. Municipality was < +/- 1 %. It was evident that this was not correct, since the uncertainty is expected to be higher based on visual comparison. Therefore, a third reference dataset was used – the church district map, also from DAGI, which provides smaller units of measurements than the municipalities.

The values from the grid net centroids for both the result and the official map from DST where intersected with the church district map, to calculate an accuracy percentage. This was based on the population sum for both datasets. The result of this assessment can be seen in figure 4.

If the sum of all districts is calculated, it shows an overall accuracy of the result of about + 10%, meaning that the model over-represents the population. As it can be seen in figure 4, the districts, where the high overrepresentation is present, are in the coastal regions, where this simplistic model appears to have its biggest problems.

If we focus on the urban areas, it can be seen from Copenhagen in figure 5, that the results here are more accurate. It is visible, that the greatest part of the districts all falls in the 95 – 105 % class, meaning that the deviation of the result from the observed data is small. Therefore, it appears that the model proposed has a better accuracy in the denser populated areas.

CONCLUSION

We can conclude based on the work conducted, that it is possible to create an estimated nighttime-population density map using the dasymetric technique on housing unit level.

With a general uncertainty of about 10 %, it is clear, that the result can never be used as anything

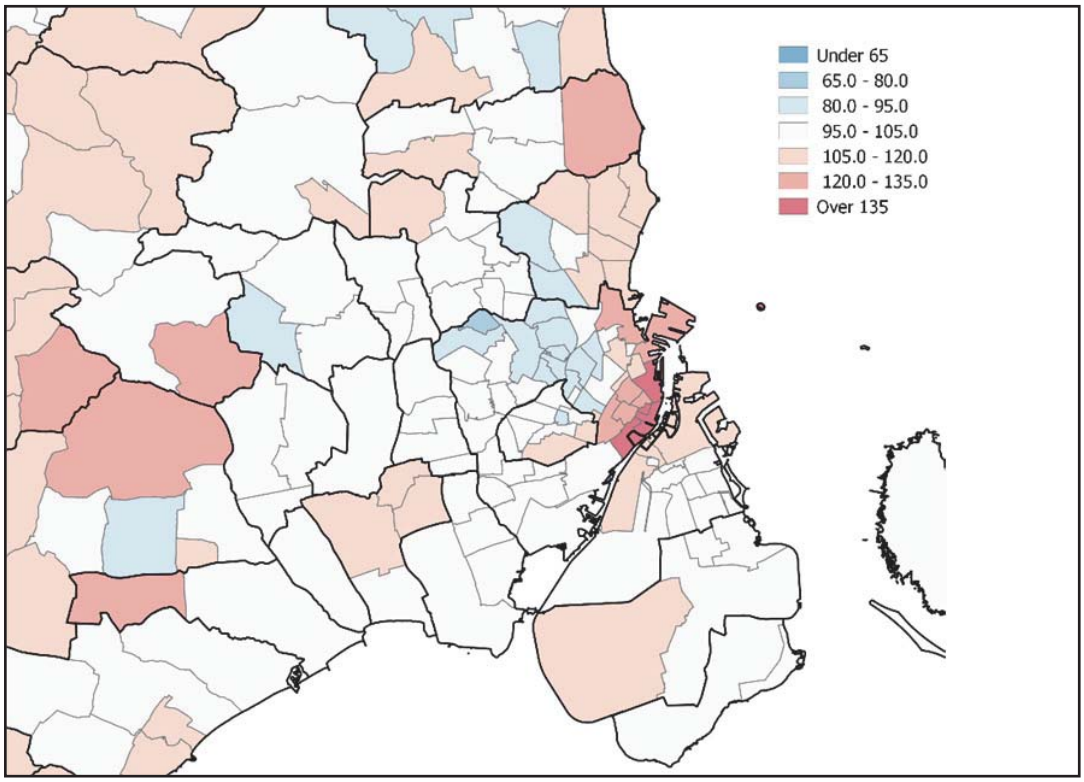

Figure 5. Subset of the accuracy assessment from the greater Copenhagen area.

else than a rough estimate on a local population. It is also evident that an extension of the model incorporating housing unit types might be a logical next step in the development. Moreover, the regional differences in the country could be addressed, since the uncertainty is unevenly distributed. Therefore, the model could be improved in numerous ways, making the results better and more accurate.

In relation to the general usability of the map, it can be concluded that the proposed method gives an estimate that is quite accurate, and might very well be usable enough for simple analyses, where an estimate is sufficient, without having to pay for the official nighttime-population dataset.

ACKNOWLEDGEMENT

Consent for using the nighttime population map from DST in the comparison, was given from a Spatial Suite customer who have purchased these data. The customer wishes to remain anonymous.

References

- BBR (2016): *"Enhedsanvendelse".* http://bbr.dk/enhedsanvendelse/0/31 [Accessed online 15.09.2016]
- DST (2011): *"Mest plads på landet"* Nyt fra Danmarks Statistik nr. 354. August 2011
- DST (2012): *"Faktaark-om-Kvadratnettet-V-2"* [Accessed online 13.09.2016]
- Eicher, C & Brewer, C (2001): *"Cartography and Geographic Information Science, l1JI.28, No.2, 2001, pp.125-138"*
- Gadeberg, L (2016): *'Vores boliger vokser – familierne skrumper' in Bolius,* Bedre hjem nr. 5 2016 [Accessed online 10.09.2016]
- Heinig, C.H. (2013): *"Vi sluger flere og flere kvadratmeter i boligen"* Realkredit Danmark, Risikostyring [Accessed online 13.09.2016]
- Mennis, J & Hultgren, T (2006): *"Cartography and Geographic Information Science, Vol. 33, No. 3, 2006, pp. 179-194"*
- Sleeter, R & Gould, M (2007): *"Geographic Information System Software to Remodel"*
- Population Data Using Dasymetric Mapping Methods. USGS Techniques and Methods 11–C2
- Tapp, A (2010): *"Cartography and Geographic Information Science, Vol. 37, No. 3, 2010, pp. 215-228"*