

Article (refereed) - postprint

Li, Sen; Juhász-Horváth, Linda; Harrison, Paula A.; Pintér, László; Rounsevell, Mark D.A.. 2016. **Population and age structure in Hungary: a residential preference and age dependency approach to disaggregate census data.**

© 2016 Sen Li

This version available <http://nora.nerc.ac.uk/515510/>

NERC has developed NORA to enable users to access research outputs wholly or partially funded by NERC. Copyright and other rights for material on this site are retained by the rights owners. Users should read the terms and conditions of use of this material at <http://nora.nerc.ac.uk/policies.html#access>

This is an Accepted Manuscript of an article published by Taylor & Francis Group in Journal of Maps, December 2016, available online: <http://www.tandfonline.com/doi/pdf/10.1080/17445647.2016.1237898?needAccess=true>

Contact CEH NORA team at
noraceh@ceh.ac.uk

Title

Mapping population and age structure in Hungary: A residential preference and age dependency approach to disaggregate census data

Authors and affiliations

Sen Li^{a, †}, Linda Juhász-Horváth^b, Paula A. Harrison^{a,c}, László Pintér^{b,d}, Mark D.A. Rounsevell^e

Affiliations

a. Environmental Change Institute, University of Oxford, South Parks Road, Oxford OX1 3QY, UK.

b. Department of Environmental Sciences and Policy, Central European University, Nádor u. 9, Budapest 1051, Hungary

c. Centre for Ecology & Hydrology, Library Avenue, Lancaster LA1 4AP, UK

d. International Institute for Sustainable Development, 325-111 Lombard Avenue, Winnipeg MB R3B 0T4, Canada

e. School of GeoSciences, University of Edinburgh, Drummond Street, Edinburgh EH8 9XP, UK

† Author for correspondence: S. Li (sen.li@ouce.ox.ac.uk)

Abstract

We present a simple model to disaggregate age-structured population census data to a 1-km grid for Hungary. A dasymetric approach was used to predict the spatial distribution of population in different age groups by distinguishing residential preferences (in relation to accessible social, economic and green amenities) for working age groups (15-29, 30-49 and 50-64) and population dependencies for children and the elderly (aged 0-14 and 65+). By using open-access land cover data and fine level population census data as inputs, the model predicts the likely spatial distribution of population and age structure for Hungary in 2011. The resulting map and gridded data provide information to support spatial planning of residential development and urban infrastructure. The model is less data-demanding than most existing approaches, but provides greater power for describing population patterns. It can also be used to create scenarios of future demographic change.

Keywords

Age structure, Dasymetric mapping, Land cover, Population distribution, Residential preference, Population dependency

1 Introduction

The spatial distribution of population, and age structures, are important socio-economic indicators that can support decision making in locating residential and infrastructure development (Alegana et al., 2015; Linard et al., 2012; Lyberaki et al., 2013). Population distribution datasets are usually not directly available at a fine resolution, but are estimated by disaggregating coarse level census data. Current disaggregation methods focus either on population surface generating (Martin et al., 2011; Martin et al., 2000; Mennis, 2009; Yoo et al., 2010), or, more recently, on empirical associations between population and multiple environmental variables, including land use and cover (Gallego et al., 2011; Linard et al., 2013), light emission (Briggs et al., 2007), and mobile phone usage (Deville et al., 2014; Stevens et al., 2015). Most of these recent, empirical association-based studies rely heavily on specific datasets that may not be available in other regions and/or may not be updated in the future. An approach that targets the most readily and widely available environmental data can, therefore, provide greater flexibility and reproducibility in downscaling population data.

Age structures are rarely studied in existing population distribution studies as population census data that includes age information is often not accessible. Age structure data is, however, important because it effects

residential behaviour, in different ways, with working ages (e.g. 15-60) being the most mobile and selective (Hagan et al., 1996). During their working life, people experience different transitions in their situation (e.g. employment, marriage and having children) and this leads to changes in their residential preferences for the social, economic and environmental amenities a property and its neighbourhood can provide (Fontaine & Rounsevell, 2009; Fontaine et al., 2014). For example, consumption behaviour (i.e. accessibility to shops) is highly related to age, being less intensive for older age groups (Erlandsen & Nymoen, 2008). Access to residential public transportation has also been shown to be more important to older employees for non-work purposes and to younger employees for work purposes (Hensher & Reyes, 2000). Furthermore, preferences for access to urban green space near a residence has been found to increase with age especially in poor and noisy urban environments (Gidlöf-Gunnarsson & Öhrström, 2007). Finally, the residential distribution pattern of children (aged 0-14) and the elderly (retired and aged 65+) is dependent to some extent on the working age population. Children are most likely to live with their parents and hence do not have a direct influence on where they live, although school location may affect adult decision-making. The elderly may live alone, but typically are highly dependent on the working age population who are the main providers of care and support (Haynes et al., 2010). Capturing these types of preferences and dependencies in an age-specific way would make population distribution models more representative of real world patterns and thus more applicable in the development of future demographic scenarios.

This paper aims to improve understanding of the mechanisms that underpin residential patterns in rural and urban environments. The objectives are to (i) develop a simple population distribution model with low requirements for input data; and (ii) account for age-specific preference and demographic dependency, in order to spatially model the residential patterns of life stages. The model integrates land cover-based approximations of residential preferences originating from regional economic theories and takes advantage of recent dasymetric approaches to disaggregate census-based, age-structured population at the 1 km² cell level. We parameterise and apply the model with open-access land cover data, i.e. the CORINE land cover (version 2012) (Bossard et al., 2000) and the OpenStreetMap datasets (accessible at www.geofabrik.de), and fine level population census data to map the distributions of population and age structures for Hungary in 2011.

2 Methods

2.1 Local residential preference for the working age groups (15-29, 30-49 and 50-64)

The local residential preference (P) is considered to be a collection of amenities that are accessible from a particular residential location and its neighbourhood (Caruso et al., 2007; Caruso et al., 2005). This is based on (i) social benefits and the availability of health, education and transport infrastructure, i.e., the social preference (S), (ii) employment and consumption opportunities, i.e., the economic preference (C), and (iii) urban environmental quality for leisure and recreational activities, i.e., the preference for urban greenspace (G). Here, P is calculated for each cell as:

$$P = S^{w1} * C^{w2} * G^{w3} \quad (1)$$

where $w1$, $w2$, and $w3$ represent the relative importance of residential preference for social (S), economic (C), and urban greenspace and leisure (G) when making residential location decisions. These weights are assumed to be age-specific to reflect the situation that people at different life stages have different preferences for access to social, economic and environmental benefits from a residential location, e.g. public transportation, schools, shops and leisure amenities (Erlandsen & Nymoen, 2008; Gidlöf-Gunnarsson & Öhrström, 2007; Hensher & Reyes, 2000).

The residential preferences are estimated using an exponential function $e^{\hat{p}}$ where \hat{p} denotes the concerned land covers' density in the cell's neighbourhood (Schindler & Caruso, 2014). Thus equation 1 extends to:

$$P = e^{w1*\hat{p}_S+w2*\hat{p}_C+w3*\hat{p}_G} \quad (2)$$

where $\hat{\rho}_S$, $\hat{\rho}_C$ and $\hat{\rho}_G$ are the neighbourhood densities of residential land covers, commercial and industrial land covers, and urban green and leisure land covers.

The residential preference for a location is influenced by the neighbouring locations (Strange, 1992). For example, frequent social interactions between neighbouring households can stimulate residential maintenance behaviour and residential stability (Ioannides, 2002). In contrast, dissatisfaction with neighbouring residents can be stimulated by negative externalities of urban areas such as noise and poor sanitation (Howley et al., 2009). In the model presented here, for a cell i , its neighbourhood density of a land cover type ($\hat{\rho}_i$) is estimated as (Caruso et al., 2007): $\hat{\rho}_i = \sum_{n \in N} (f(d_n, \tilde{d}) \cdot U(n))$, where N is the collection of cells belonging to the (circular or rectangular) neighbourhood of the cell i . A distance decay function $f(d_n, \tilde{d})$ is used to estimate the weight given to the cell n ($n \in N$) for its importance on residential decisions in cell i : $f(d_n, \tilde{d}) = 1 - ((d_n - 1)/\tilde{d})^2$, with d_n being the distance from cell n to the cell i , and \tilde{d} being the maximum distance considered within the neighbourhood. The function $U(n)$ returns the proportion of the land cover type in cell n . Since the distance an individual is willing to travel for different types of activity is different, the neighbourhood size ($\tilde{d} * \tilde{d}$) used to estimate the three local preference types needs to be defined. The economic preference, \tilde{d} is set according to the observed distance ranges of daily work trips. In this study, multiple values of \tilde{d} (0-5, 5-15, 15-25 and 25-50 km) were considered for the economic preference based on census data quantifying the frequency of commuting distances (Table 1, source: the Hungarian Central Statistical Office 2008), by applying the frequencies as additional weights. The social and urban greenspace preferences are more closely related to the amenities for personal activities (e.g. social interaction, shopping, healthcare, religious services, and daily leisure) which are normally associated with a shorter travel distance (Li et al., 2015; Li et al., 2016; Moudon et al., 2006; Schafer, 2000), hence \tilde{d} was set to 3 km.

2.2 Population dependency for children (aged 0-14) and the elderly (aged 65+)

Dependency is important in representing the location of the non-working population. Thus, it would not be appropriate to use the preference-based approach for predicting the distribution of dependents such as children and the elderly. Children (aged 0-14) are assumed to live with their parents who are represented in the working age population. The elderly (aged 65+) are also highly dependent on the working age population for care and support (Haynes et al., 2010). Despite empirical evidence which suggests that a proportion of previously urban retirees choose to move to the countryside (Brown et al., 2005; Kok, 1999), residential mobility of the European elderly is generally low (Angelini & Laferrère, 2012). Given that the average retirement age in Hungary is 62.5 and lower incomes represent an economic constraint on relocation, elderly mobility after 65 is assumed to be low and residential preferences are assumed to be largely in line with those of the 50-64 age group (Gobillon & Wolff, 2011; Stockdale & MacLeod, 2013). This is further supported by empirical findings which suggest that residential preferences of the elderly are more likely to be influenced by their financial situations and/or social bonds with neighbours than by proximity to social, economic and green amenities (Angelini & Laferrère, 2012; Hansen & Gottschalk, 2006; Temelová & Dvořáková, 2012). In this study, the strengths of nine dependencies were considered for both children (aged 0-14) and the elderly (aged 65+) to support their potential association with each of the working age groups (15-29, 30-49 and 50-64) and urbanisation levels (capital, town and village).

2.3 Population redistribution by dasymetric modelling

The method for population redistribution follows the dasymetric modelling approach used in Briggs et al. (2007) and Stevens et al. (2015), who fit a regression function of aggregated cell-level parameters to a higher, administrative region and use the resulting function to redistribute the higher-level population to each inhabitable 1 km cell pycnophylactically. The dasymetric modelling approach is described in more detail by Batista e Silva et al. (2013).

For the working age groups (15-29, 30-49 and 50-64), regression functions were first built based on aggregated values of the local preferences in the capital (P^C), towns (P^T) and villages (P^V) at the administrative region level: $f_w = a_1 * P^C + a_2 * P^T + a_3 * P^V$, where a_1 , a_2 and a_3 denotes the weight of

the local preference effects on population distribution. For an inhabitable cell i in a region m , the number of residents (W) of each working population age group is calculated following the approach used in Briggs et al. (2007), as:

$$W_i = \frac{a_1 * p_i^C + a_2 * p_i^T + a_3 * p_i^V}{\sum_{k \in M} (a_1 * p_k^C + a_2 * p_k^T + a_3 * p_k^V)} \cdot W_m \quad (3)$$

where p_i^C , p_i^T , and p_i^V are local preferences of cell i ($i \in M$) for working population age groups (15-29, 30-49 and 50-64). M denotes the collection of all inhabitable cells in region m . Parameter k refers to the k th cell from M . W_m is the population in region m to be redistributed.

For the dependent population age groups (aged 0-14 and 65+), regression functions were built based on a collection of nine variables $\{R\}$ for population aged 15-29, 30-49, 50-64 in the capital, town and village at the administrative level: $f_d = \sum_{j=0}^R b_j * W_j$, where b is the weight of dependency on a working age population W . Then, the dependent populations (D) were estimated separately for age groups 0-14 and 65+ as:

$$D_i = \frac{\sum_{j=0}^R b_{ji} * W_{ji}}{\sum_{k \in M} (\sum_{j=0}^R b_{jk} * W_{jk})} \cdot D_m \quad (4)$$

where W_{ji} ($W_{ji} \in \mathbb{R}$) is the population of a working age group (15-29, 30-49 or 50-64) of cell i ($i \in M$) as estimated by equation 3. D_m is the dependent population of an age group in region m to be redistributed.

In this study, only positive coefficients were considered and the intercept was excluded from the regressive models. Thus, the models were not necessarily those giving the best predictions, but were constrained in order to avoid counter-intuitive relationships (e.g. negative relationships between local preference and population distribution) and negative predicted populations (Briggs et al., 2007).

3 Workflow

3.1 Data preparation

The age-structured population data in 2011 were acquired from the Hungarian Central Statistical Office. The data were binned in different age groups (i.e., 0-14, 15-29, 30-49, 50-64, and 65+) and organised at two levels to be used in model development: (i) at the fine settlement level (3176 settlements) and (ii) at the district level (175 districts, including the districts of the capital, Budapest, and 174 micro regions). Classification of the urbanisation levels of the settlements was based on their location and legal status: the 23 districts in Budapest were classified as “capital”, 327 settlements were towns, and 2826 were villages.

The land cover information was extracted from open-access datasets. The latest release of the CORINE Land Cover datasets (CLC) version 2012 was acquired from the Hungarian Institute of Geodesy, Cartography and Remote Sensing (FÖMI). The European Environment Agency (EEA) 1 km reference grid for Hungary was used as the base grid for population distribution mapping. To reduce computing requirements, a proportion of cells were selected as inhabitable and used in the modelling. When performing the selection, the OpenStreetMap data which is made available by Geofabrik GmbH (accessed in 2014 at www.geofabrik.de) were used to refine the results based solely on using the CLC data. Thus, a cell was considered inhabitable if urban fabric lands (from the CLC data), or residential roads (from the OpenStreetMap data) were present (Figure 1A). This identified 23,556 inhabitable cells (out of 94,266) and ensured at least one inhabitable cell for each settlement-level unit. Neighbourhood land cover densities for the local preference P for each type of amenity were calculated using equation 2. For the estimation of the neighbourhood density, the residential land density ($\hat{\rho}_S$) was based on the “continuous urban fabric” (CLC code 111) and “discontinuous urban fabric” (CLC code 112); whilst commercial and industrial land density ($\hat{\rho}_C$) was based on the “industrial or commercial units” (CLC code

121), and; green urban and outdoor leisure land density ($\hat{\rho}_G$) was based on the “green urban areas” (CLC code 141) and “sport and leisure facilities” (CLC code 142) (Figure 1B-D).

3.2 Model parameterisation and evaluation

The methods for model parameterisation and evaluation largely followed the approach of Batista e Silva et al. (2013) and Stevens et al. (2015), in which census units at the finest available administration level (the settlement level) were aggregated to the next level up (district level), and these aggregated counts were used for both parameterisation and evaluation. This allows a comparison of the sums of the cell-level predictions with census population data at the original, fine settlement level. All the analytical steps were performed using R version 3.1 (R Core Team, 2012). To ensure positive-only coefficients and intercept-free function developments, the penalised regressive approach was adopted with the “*penalized*” R library (Goeman et al., 2014). Only national-level preferences and dependencies were considered.

For the working age groups (15-29, 30-49 and 50-64), a simple parameterisation approach was adopted for the age-specific weights of the local preferences (w, a in equations 2 and 3). First, for each working age group, values were looped from 0 to 1 for each of the three preference weights (w_1, w_2 , and w_3 , with an interval of 0.1) to calculate cell-level local preferences P using equation 1. This resulted in 1331 different P estimations ($11*11*11$). Second, for each P , a linear regression function was built to fit equation 3 (and estimate a_1, a_2 and a_3) with the district level census data. Third, the cell-level population was predicted dasymetrically and these values were summed at the settlement level and compared with the census data, using root-mean-square deviation (RMSE, for absolute accuracy) and Pearson’s product-moment correlation (r , for relative accuracy and the coefficient of determination r^2). The best 10 candidate models (with the lowest RMSE) were selected and their preference weights (w, a) were averaged as the final values. Lastly, based on the final values of all parameters in equation 4, the final model was evaluated at the district level using RMSE and r^2 .

For the dependent population age groups (0-14 and 65+), the weight of the dependency on each working population group (b in equation 4) was estimated by building regressive functions of the three working population age groups at the district level, using census data. The functions were applied, using the predicted distributions of working age population groups as inputs, to dasymetrically disaggregate the district level populations to cell-level. Finally, following the steps for the working population age groups, the predicted cell-level populations were summed at the settlement level for model evaluation.

3.3 Mapping the population and age structures

The resulting residential preference and population dependency functions were used to redistribute the available census data at the settlement-level onto the 1 km grid. The final outputs were processed and visualised with ArcGIS V 10.2. The main maps were produced by joining predicted population age-groups to the EEA 1 km reference grid. A 3D map of population density was created in ArcScene by extruding the cells’ height by their predicted population sizes. Population density for each age group was largely determined by total population density and was found difficult to distinguish visually with the 1 km² map. Instead, the proportions of each age group were visualised, with contour lines indicating the spatial difference.

4 Results

The estimated local residential preference weights (w and a in equations 2 and 3) for the working population age groups are presented in Table 2. At the national level, all working populations had high preferences for social amenities, with the young age group (15-29) being marginally lower. Middle (30-49) and senior (50-64) working age-groups had similar preference weights. They had relatively lower preferences for economic amenities, in line with a decrease in their purchasing frequency as reported in the literature (Erlandsen & Nymoen, 2008), and a slightly lower preference for urban green amenities. The ratios of a_1 to a_2 to a_3 (weights of the capital, towns and villages) could reflect the relative “attractiveness” of the capital, towns and villages to residents in different age groups. In general, the attractiveness of villages increased with age. The weights reflecting the dependency of children and the elderly on the working population are presented

in Table 3. Children (0-14) depend relatively more on the working population of young adults (15-29) in villages, middle-aged adults (30-49) in villages and towns, and the senior working age group (50-64) in the capital. The elderly depend heavily on young adults (15-29) in the capital, and senior adults (50-64) in the capital, towns and villages. In line with studies on the geography of aging (Harper & Laws, 1995; Shiode et al., 2014), our results suggest that age dependency varies strongly over space and between urbanisation levels. It should be noted that, a zero weight estimated by the regression model does not mean that there is no dependency between the populations, but that the dependency is statistically insignificant. Furthermore, the preferences and dependencies also differ spatially and can be estimated region-specifically for other research purposes.

Six scatter plots (one for total population, Figure 2A, and five for different age groups, Figure 2B-F) are provided in Figure 2 to compare the predicted population densities with the census data at the settlement level. The Pearson's product-moment correlation r for all models was greater than 0.94, indicating all coefficients of determination (r^2) are greater than 0.88. In addition, only marginal differences were found (<0.2%) when the RMSE value of the final models was compared with the lowest RMSE observed in the candidate models. These results suggest a satisfactory predictive power for all the age-specified population redistribution models.

The 3D map of population density generated using ArcScene was processed to include a symbolic representation of age structures in 18 major cities and towns in Hungary (Main Map A). In 2011, the Hungarian population density over the 23,557 inhabitable cells (out of 94,266) from the EEA 1 km reference grid ranged from 3.98 to 38,319 inhabitants per km², with a mean value of 422, a median value of 228, and a standard deviation (std) of 733. Population was greatly concentrated in the Budapest region. The contour maps for the five age groups (Main Map B) give a better sense of how the proportion of an age group differs spatially compared to the other age groups. In the inhabitable cells identified for Hungary, the mean proportion of an age group was 0.152 (std 0.036) for children (aged 0-14); 0.176 (std 0.027), 0.282 (std 0.027) and 0.217 (std 0.028) for young (aged 15-29), middle-aged (aged 30-49), and senior (aged 50-64) working age groups, respectively; and 0.173 (std 0.044) for the elderly (aged 65+). The proportion of children was higher outside the city of Budapest, in the major towns of Pest county around the capital and in northern Hungary. Young adults and middle-aged populations were denser in the centre of all major towns, while the latter also preferred to live near Lake Balaton. The senior working age population preferred not to live in the centre of the capital and towns, but in peri-urban areas and, in particular, the Lake Balaton region. A large proportion of the elderly population was also found in the Lake Balaton area as well as the western part of Budapest. For demonstration, maps of the Budapest region cut out from Maps A and B are presented in the main text as Figure 3 and Figure 4.

5 Discussion and conclusions

A map showing the predicted distribution of Hungarian population and age structure for 2011 on a 1 km grid has been generated. The methodology combines and extends recent theoretical developments of “residential preferences”, including approximating the accessibility of social, economic and urban green/leisure amenities of a residential place and its neighbourhood, based on land cover information extracted from the open-access CORINE and OpenStreetMap datasets. The relative preferences and dependencies of different age structures were distinguished to enable different modelling of their distribution patterns. A commonly used dasymetric modelling approach was adopted to build the model. The finest census population data (3176 settlements) was aggregated to the next administrative level (175 districts) for model parameterisation and evaluation, and the model was applied to disaggregate the census data onto the 1 km grid.

The resolution of the final gridded population map at 1 km² is amongst the finest available spatial population datasets covering the study region (Linard & Tatem, 2012). A more spatially detailed dataset for the European population (100 meter) can be found in Batista e Silva et al. (2013). The GEOSTAT database (www.efgs.info/geostat/) also provides access to population data on 1km² grid, constructed using a bottom-up approach. However, none of these datasets provide spatial information about age structures as presented here. Compared to existing dasymetric approaches, especially those based on land use/cover data (e.g., Briggs et al. (2007), Gallego et al. (2011), and Stevens et al. (2015)), the present model has a simpler

form, is less data-demanding and focuses more on explanatory power. The latter enables the variables of residential preference and population dependency to be scrutinised to help explain observed spatial differences in the pattern of population and age groups. This could be useful for various policy-related purposes, e.g., assessing vulnerability to natural hazard, delivering healthcare services, planning urban/rural developments and educating the general public.

The methodology presented here is flexible and could be further developed or applied to other contexts and other regions. First, for demonstration purposes we focused on the national scale and a spatial resolution of 1 km. The approach could easily be downscaled (e.g. to 100 m resolution) to take account of more localised issues. Second, the age-specific residential preferences may also vary geographically. While this study only looked at nationwide preferences and dependencies, further studies could parameterise the model for different local regions (e.g. counties) to better approximate and understand local residential patterns. Third, fine level census data were available in this study which enabled an empirical parameterisation and led to an excellent predictive power of the model. In data-poor situations, coarser level census data (e.g. county level) could also be used as input. Where age-structure census data are not available, preferences and dependencies could be estimated from qualitative information such as residents' opinion and expert knowledge. Fourth, the functions could be further integrated into urban land use models and applied to project future changes in the spatial patterns of population and age structures, by developing scenarios of changes in preferences and dependency. Exploring future uncertainties in this way could provide valuable decision-support to advance adaptive strategies in changing social and physical environments (Harrison et al., 2015).

This study is subject to several limitations, which can be targeted for future research. A detailed quantitative evaluation of our final model predictions was not possible because such ground truth data does not exist. When such data are not available, which is the case for many population disaggregation studies, future model development may benefit from a qualitative evaluation (Bennett et al., 2013). This would involve developing a participatory approach in which stakeholders from relevant sectors work together to assess the accuracy of the predicted patterns and the general acceptance and performance of the model components.

The use of the CORINE and OpenStreetMap (OSM) data as a proxy for residential preferences for amenities may require further improvements. Class 121 of the CORINE data was used in this study for estimating economic preference, as it covers a wide range of economic infrastructures. However, the class also covers areas which may exert negative effects on residential attractiveness, such as abandoned industrial sites, nuclear power plants and military barracks. Further classification of the urban land cover is required to distinguish these infrastructures and their effects on residential preferences. Moreover, we prioritised flexibility for model integration in our methodology which constrained the factors considered for residential preferences to land cover, the most common output of urban land use/cover models. Thus, point-of-interest (POI) information, such as the location of schools, hospitals, shopping centres, parks and public transport nodes, in the OSM dataset was not utilised. These data could be valuable in further explaining residential preferences and modelling population age structures. Extending our methodology into a more data-intensive approach would also be likely to improve the predictive accuracy of the model as well as understanding of real-world population patterns.

Alternative approaches could be applied for modelling the elderly population distribution. Due to a lack of information on the elderly's residential preference, this study used a dependency approach. More empirical studies are needed to gain a better understanding of the key environmental, economic and social-demographic factors underpinning the elderly population distribution.

Finally, even though regressive approaches are commonly used for dasymetric mapping, they may raise statistical concerns when being applied to predict population distribution over small areas, as the effects of size and shape of the target zones are usually ignored. Future studies looking into fine-level population mapping may need to either take such effects into consideration (Shuttleworth et al., 2011), or take advantages of alternative approaches for population surface modelling such as area-to-point kriging (Kyriakidis, 2004).

Data

The output gridded population and age structure data for Hungary in 2011 is provided in the online supplementary materials.

Acknowledgements

The authors would like to thank the four reviewers for their time and constructive comments on the manuscript. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 603416 (The IMPRESSIONS Project - Impacts and Risks from High-End Scenarios: Strategies for Innovative Solutions; www.impressions-project.eu).

Reference

- Alegana, V. A., Atkinson, P. M., Pezzulo, C., Sorichetta, A., Weiss, D., Bird, T., . . . Tatem, A. J. (2015). Fine resolution mapping of population age-structures for health and development applications. *Journal of The Royal Society Interface*, 12(105). doi: 10.1098/rsif.2015.0073
- Angelini, V., & Laferrère, A. (2012). Residential mobility of the European elderly. *CESifo Economic Studies*, 58(3), 544-569. doi: 10.1093/cesifo/ifr017
- Batista e Silva, F., Gallego, J., & Lavalle, C. (2013). A high-resolution population grid map for Europe. *Journal of Maps*, 9(1), 16-28. doi: 10.1080/17445647.2013.764830
- Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., . . . Andreassian, V. (2013). Characterising performance of environmental models. *Environmental Modelling & Software*, 40, 1-20. doi: 10.1016/j.envsoft.2012.09.011
- Bossard, M., Feranec, J., & Otahel, J. (2000). CORINE land cover technical guide: Addendum 2000.
- 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*, 108(4), 451-466. doi: 10.1016/j.rse.2006.11.020
- Brown, D. L., Kulcsár, L. J., Kulcsár, L., & Obádovics, C. (2005). Post-socialist restructuring and population redistribution in Hungary. *Rural Sociology*, 70(3), 336-359. doi: 10.1526/0036011054831170
- Caruso, G., Peeters, D., Cavailhès, J., & Rounsevell, M. (2007). Spatial configurations in a periurban city. A cellular automata-based microeconomic model. *Regional Science and Urban Economics*, 37(5), 542-567. doi: 10.1016/j.regsciurbeco.2007.01.005
- Caruso, G., Rounsevell, M., & Cojocar, G. (2005). Exploring a spatio - dynamic neighbourhood - based model of residential behaviour in the Brussels periurban area. *International Journal of Geographical Information Science*, 19(2), 103-123. doi: 10.1080/13658810410001713371
- Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., . . . Tatem, A. J. (2014). Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences*, 111(45), 15888-15893. doi: 10.1073/pnas.1408439111
- Erlandsen, S., & Nymoen, R. (2008). Consumption and population age structure. *Journal of Population Economics*, 21(3), 505-520. doi: 10.1007/s00148-006-0088-5
- Fontaine, C. M., & Rounsevell, M. D. A. (2009). An agent-based approach to model future residential pressure on a regional landscape. *Landscape Ecology*, 24(9), 1237-1254. doi: 10.1007/s10980-009-9378-0
- Fontaine, C. M., Rounsevell, M. D. A., & Barbette, A.-C. (2014). Locating household profiles in a polycentric region to refine the inputs to an agent-based model of residential mobility. *Environment and Planning B: Planning and Design*, 41(1), 163-184. doi: 10.1068/b37072
- Gallego, F. J., Batista, F., Rocha, C., & Mubareka, S. (2011). Disaggregating population density of the European Union with CORINE land cover. *International Journal of Geographical Information Science*, 25(12), 2051-2069. doi: 10.1080/13658816.2011.583653
- Gidlöf-Gunnarsson, A., & Öhrström, E. (2007). Noise and well-being in urban residential environments: The potential role of perceived availability to nearby green areas. *Landscape and Urban Planning*, 83(2-3), 115-126. doi: 10.1016/j.landurbplan.2007.03.003
- Gobillon, L., & Wolff, F.-C. (2011). Housing and location choices of retiring households: Evidence from France. *Urban Studies*, 48(2), 331-347. doi: 10.1177/0042098010363493

- Goeman, J., Meijer, R., & Chaturvedi, N. (2014). L1 and L2 penalized regression models.
- Hagan, J., MacMillan, R., & Wheaton, B. (1996). New kid in town: Social capital and the life course effects of family migration on children. *American Sociological Review*, 61(3), 368-385. doi: 10.2307/2096354
- Hansen, E. B., & Gottschalk, G. (2006). What makes older people consider moving house and what makes them move? *Housing, Theory and Society*, 23(1), 34-54. doi: 10.1080/14036090600587521
- Harper, S., & Laws, G. (1995). Rethinking the geography of ageing. *Progress in Human Geography*, 19(2), 199-221. doi: 10.1177/030913259501900203
- Harrison, P. A., Holman, I. P., & Berry, P. M. (2015). Assessing cross-sectoral climate change impacts, vulnerability and adaptation: an introduction to the CLIMSAVE project. *Climatic Change*, 128(3-4), 153-167. doi: 10.1007/s10584-015-1324-3
- Haynes, P., Hill, M., & Banks, L. (2010). Older people's family contacts and long-term care expenditure in OECD countries: A comparative approach using qualitative comparative analysis. *Social Policy & Administration*, 44(1), 67-84. doi: 10.1111/j.1467-9515.2009.00700.x
- Hensher, D., & Reyes, A. (2000). Trip chaining as a barrier to the propensity to use public transport. *Transportation*, 27(4), 341-361. doi: 10.1023/A:1005246916731
- Howley, P., Scott, M., & Redmond, D. (2009). Sustainability versus liveability: an investigation of neighbourhood satisfaction. *Journal of Environmental Planning and Management*, 52(6), 847-864. doi: 10.1080/09640560903083798
- Ioannides, Y. M. (2002). Residential neighborhood effects. *Regional Science and Urban Economics*, 32(2), 145-165. doi: 10.1016/S0166-0462(01)00082-5
- Kok, H. (1999). Migration from the city to the countryside in Hungary and Poland. *GeoJournal*, 49(1), 53-62. doi: 10.1023/A:1007092228633
- Kyriakidis, P. C. (2004). A Geostatistical Framework for Area-to-Point Spatial Interpolation. *Geographical Analysis*, 36(3), 259-289. doi: 10.1111/j.1538-4632.2004.tb01135.x
- Li, S., Colson, V., Lejeune, P., Speybroeck, N., & Vanwambeke, S. O. (2015). Agent-based modelling of the spatial pattern of leisure visitation in forests: A case study in Wallonia, south Belgium. *Environmental Modelling & Software*, 71, 111-125. doi: 10.1016/j.envsoft.2015.06.001
- Li, S., Colson, V., Lejeune, P., & Vanwambeke, S. O. (2016). On the distance travelled for woodland leisure via different transport modes in Wallonia, south Belgium. *Urban Forestry & Urban Greening*, 15, 123-132. doi: 10.1016/j.ufug.2015.12.007
- Linard, C., Gilbert, M., Snow, R. W., Noor, A. M., & Tatem, A. J. (2012). Population distribution, settlement patterns and accessibility across Africa in 2010. *PLoS ONE*, 7(2), e31743. doi: 10.1371/journal.pone.0031743
- Linard, C., & Tatem, A. J. (2012). Large-scale spatial population databases in infectious disease research. *International Journal of Health Geographics*, 11(1), 1-13. doi: 10.1186/1476-072X-11-7
- Linard, C., Tatem, A. J., & Gilbert, M. (2013). Modelling spatial patterns of urban growth in Africa. *Applied Geography*, 44, 23-32. doi: 10.1016/j.apgeog.2013.07.009
- Lyberaki, A., Tinios, P., Mimis, A., & Georgiadis, T. (2013). Mapping population aging in Europe: How are similar needs in different countries met by different family structures? *Journal of Maps*, 9(1), 4-9. doi: 10.1080/17445647.2012.752334
- Martin, D., Lloyd, C., & Shuttleworth, I. (2011). Evaluation of gridded population models using 2001 Northern Ireland census data. *Environment and Planning A*, 43(8), 1965-1980. doi: 10.1068/a43485
- Martin, D., Tate, N. J., & Langford, M. (2000). Refining population surface models: Experiments with Northern Ireland census data. *Transactions in GIS*, 4(4), 343-360. doi: 10.1111/1467-9671.00060
- Mennis, J. (2009). Dasymetric mapping for estimating population in small areas. *Geography Compass*, 3(2), 727-745. doi: 10.1111/j.1749-8198.2009.00220.x
- Moudon, A. V., Lee, C., Cheadle, A. D., Garvin, C., Johnson, D., Schmid, T. L., . . . Lin, L. (2006). Operational definitions of walkable neighborhood: theoretical and empirical insights. *Journal of Physical Activity & Health*, 3, S99.
- R Core Team. (2012). R: A language and environment for statistical computing.
- Schafer, A. (2000). Regularities in travel demand: an international perspective. *Journal of transportation and Statistics*, 3(3), 1-31.
- Schindler, M., & Caruso, G. (2014). Urban compactness and the trade-off between air pollution emission and exposure: Lessons from a spatially explicit theoretical model. *Computers, Environment and Urban Systems*, 45, 13-23. doi: 10.1016/j.compenvurbsys.2014.01.004

- Shiode, N., Morita, M., Shiode, S., & Okunuki, K.-i. (2014). Urban and rural geographies of aging: a local spatial correlation analysis of aging population measures. *Urban Geography*, 35(4), 608-628. doi: 10.1080/02723638.2014.905256
- Shuttleworth, I. G., Lloyd, C. D., & Martin, D. J. (2011). Exploring the implications of changing census output geographies for the measurement of residential segregation: the example of Northern Ireland 1991–2001. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(1), 1-16. doi: 10.1111/j.1467-985X.2010.00647.x
- Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS ONE*, 10(2), e0107042. doi: 10.1371/journal.pone.0107042
- Stockdale, A., & MacLeod, M. (2013). Pre-retirement age migration to remote rural areas. *Journal of Rural Studies*, 32, 80-92. doi: 10.1016/j.jrurstud.2013.04.009
- Strange, W. (1992). Overlapping neighborhoods and housing externalities. *Journal of Urban Economics*, 32(1), 17-39. doi: 10.1016/0094-1190(92)90011-9
- Temelová, J., & Dvořáková, N. (2012). Residential satisfaction of elderly in the city centre: The case of revitalizing neighbourhoods in Prague. *Cities*, 29(5), 310-317. doi: 10.1016/j.cities.2011.11.015
- Yoo, E.-H., Kyriakidis, P. C., & Tobler, W. (2010). Reconstructing population density surfaces from areal data: A comparison of Tobler's pycnophylactic interpolation method and Area-to-Point Kriging. *Geographical Analysis*, 42(1), 78-98. doi: 10.1111/j.1538-4632.2009.00783.x

Tables

Table 1 Classes of commuting distance in Hungary (% of total)

	Home worker	0–5 km	5–15 km	15–25 km	25–50 km	50 km +	Other
Town	5.3	55.7	19.6	7.6	6.4	5	..
Village	5.9	31.8	23.6	18.5	12.7	6.8	..
Budapest	5.5	28.7	40.4	17.1	6.8	1.6	..
All	5.5	42.9	24.8	12.9	8.5	4.9	0.4

Table 2 Weights of local preference for working population age groups

Age group	w1	w2	w3	a1	a2	a3
15-29	0.98	0.94	0.98	5071.77	4390.95	252.14
30-49	1	0.92	0.96	8261.40	6452.23	408.01
50-64	1	0.92	0.96	5362.76	4597.00	307.77

w1 – social preference

w2 – economic preference

w3 – urban green preference

a1 – preference in capital

a2 – preference in town

a3 – preference in village

Table 3 Dependency of children and elderly populations on working populations

Age group	b1	b2	b3	b4	b5	b6	b7	b8	b9
0-14	0	0	0.561	0	0.463	0.154	0.659	0.008	0
65+	0.625	0	0	0	0	0	0.388	0.783	0.724

b1 – dependency on population aged 15-29 in capital

b2 – dependency on population aged 15-29 in town

b3 – dependency on population aged 15-29 in village

b4 – dependency on population aged 30-49 in capital

b5 – dependency on population aged 30-49 in town

b6 – dependency on population aged 30-49 in village

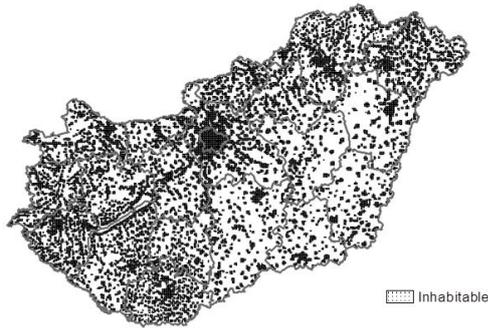
b7 – dependency on population aged 50-64 in capital

b8 – dependency on population aged 50-64 in town

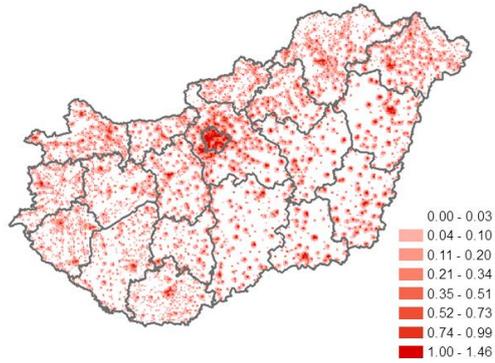
b9 – dependency on population aged 50-64 in village

Figures

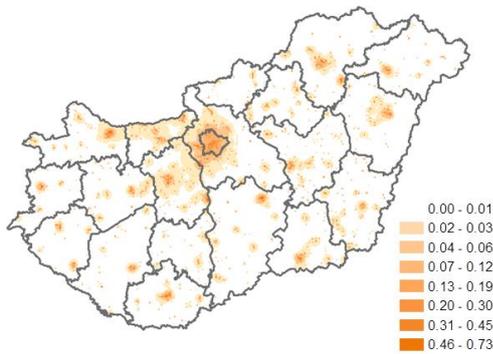
A – Inhabitable cells



B – Land cover density for social amenity



C – Land cover density for economic amenity



D – Land cover density for urban greenery amenity

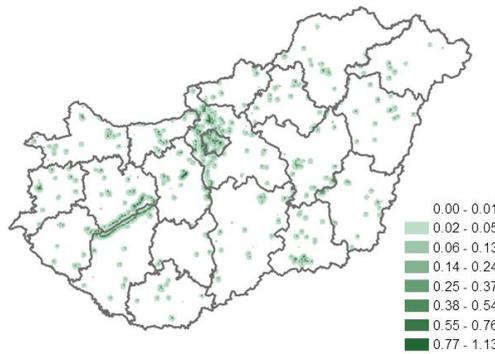


Figure 1 Geographical layers for estimating residential preferences

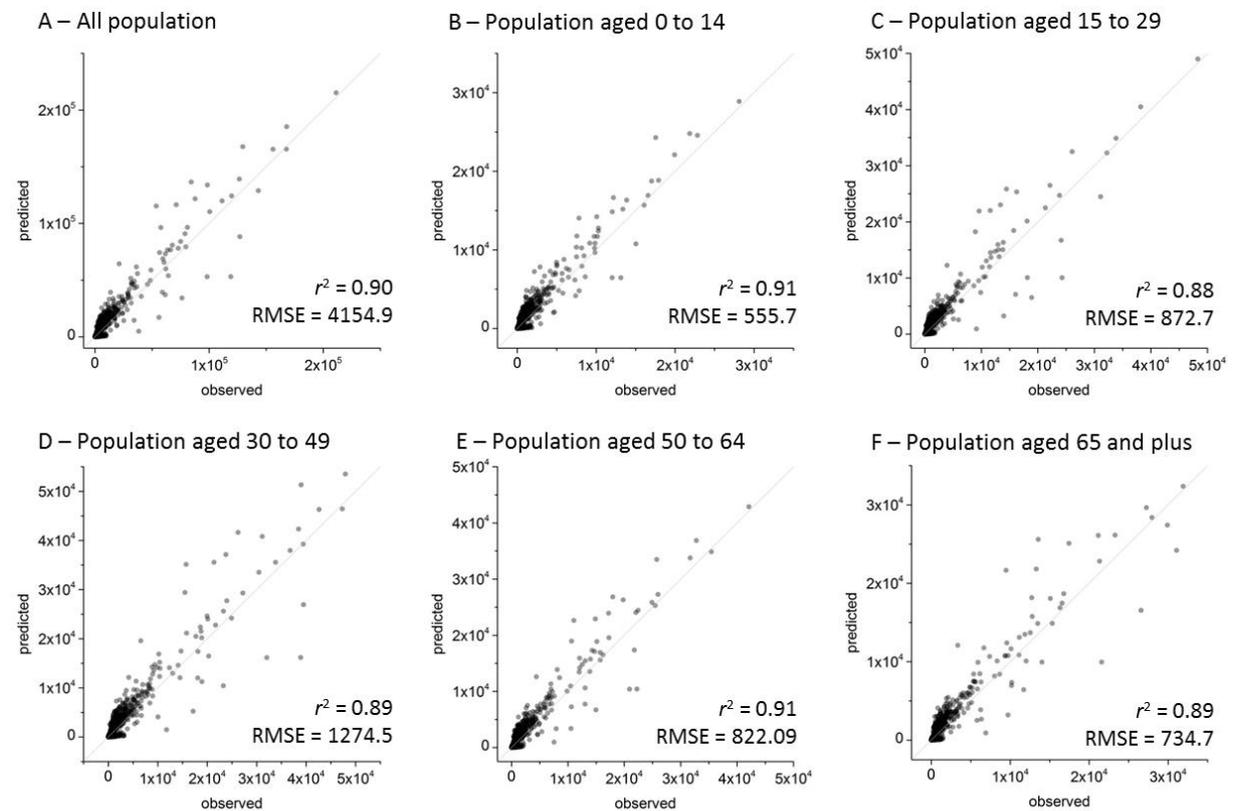


Figure 2 Scatter plots: observed (census) vs predicted populations at the settlement level

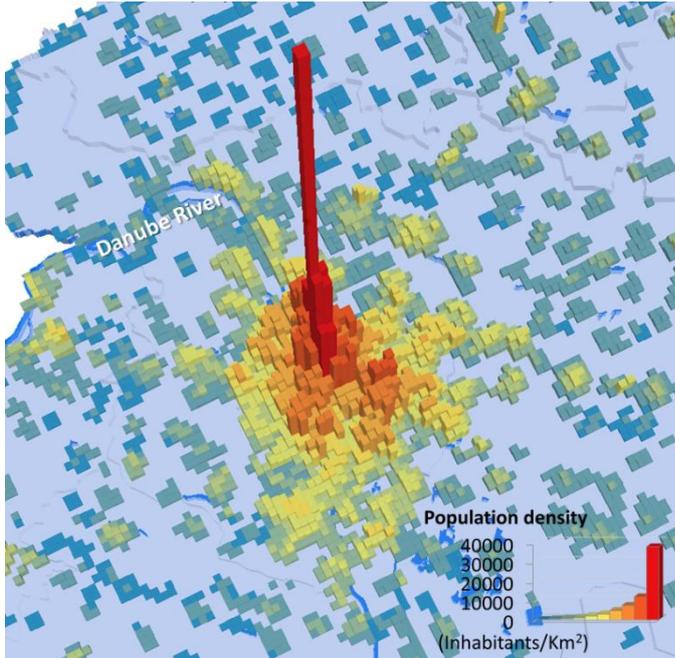
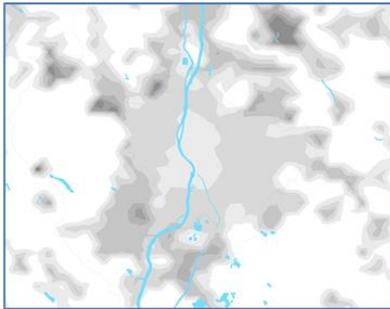
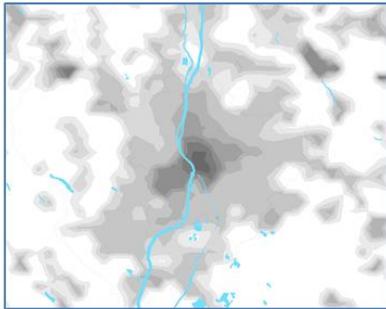


Figure 3 Budapest region on the main map (map A): population density

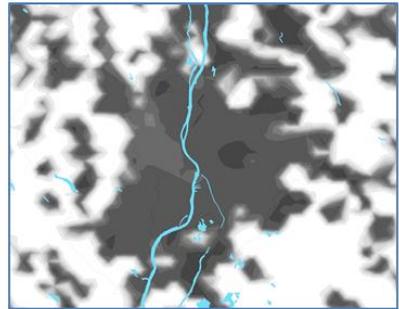
A – Population aged 0-14



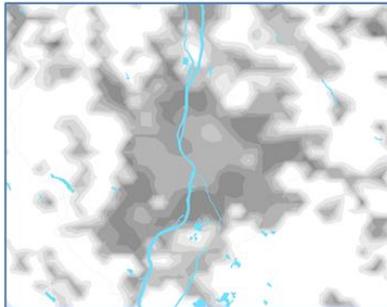
B – Population aged 15-29



C – Population aged 30-49



D – Population aged 50-64



E – Population aged 65+

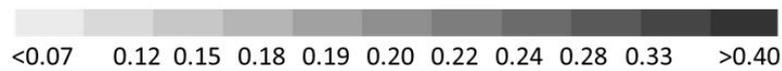
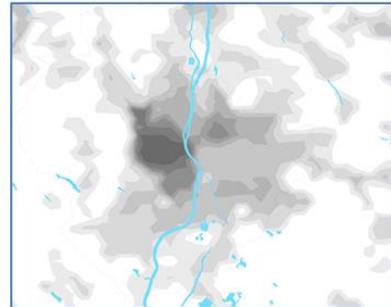


Figure 4 Budapest region on the main map (map B): proportions of age groups in total population.