Spatially Explicit Population Projections: The case of Copenhagen, Denmark

Marina Georgati[a](corresponding author), Carsten Keßler[a]

marinag@plan.aau.dk, kessler@plan.aau.dk,

[a]Department of Planning, Aalborg University, Copenhagen, Denmark

Abstract. Cities expand rapidly with international migration significantly contributing to urban growth and urban population change. However, cities miss out on a great opportunity of reclaiming valuable knowledge on future population distribution due to the lack of established tools and methodologies to project where it is more likely for people of specific socio-demographic groups to set up home. The present work suggests that spatially explicit projections can play a significant role as a tool for urban planning and for managing diversity creatively, especially when a combination of social, demographic and topographic data is utilized. Machine learning techniques have demonstrated capabilities to capture relationships among this plethora of urban features to estimate future population distribution. We present a flexible, ML-based methodology for high-resolution gridded population projections by demographic characteristics, and specifically by region of origin, for the capital region of Copenhagen, Denmark, by combining various socio-demographic and topographic input layers.

Keywords. Migration, Spatial Projections, Population Change, Machine Learning, Convolution Neural Network

1 Introduction

Cities have been expanding rapidly for the last decades, with migration playing a determining role in their development and thus shaping their image. Multiculturalism and internationalism are now parts of most modern cities' identities. However, cultural diversity is not homogeneously dispersed in urban centers. Various local features determine population distribution and these variations in density of cultural expressions implies that some areas are more attractive for immigrants. This fact generates substantial questions around the residential choices of immigrants, the discernible spatial characteristics that affect their choices, and most importantly about their future distribution in the urban fabric. Nowadays, increased interest around these research issues is expressed especially due to the rising inequalities, socio-spatial segregation and the continuous risk of ghettoization that cities suffer. Issues also arise due to the lack of established methods and tools that could utilize the existing knowledge on the determinants of migrant settlement in order to help cities assess where future migrants are likely to set up home.

This paper suggests that local population projections bear dynamics to become a useful tool for urban planning and for managing diversity and socio-economic segregation in urban centers, especially when local social and topographic variables are included. The ever-growing amount of both urban and demographic data, combined with novel geocomputation methods, constitute the breeding ground for answering the aforementioned questions. Therefore, we examine how migration dynamics and the relationships between numerous variables that affect the urban distribution of migrants can be captured and analysed simultaneously.

In contrast to the already conducted work, we examine the incorporation of both demographic and locally explicit characteristics in population projections at high spatial resolution. So far, most of the studies have relied on population gravity models projecting movements between two distinct areas/zones (Jones and O'Neill, 2013; Grübler et al., 2007), but have not estimated the local distribution of the population in the destination areas, let alone the demographic features of the examined population. Furthermore, the geospatial components that determine the residential choices

of specific demographic groups have been mostly disregarded due to their complex or abstract nature.

These factors are not universal; they may affect particular demographic groups in various ways. However, there is evidence that the local variations of migration systems can be captured through the effective use of machine learning methods that can *learn* and relate the plethora of urban variables. Among these factors, data on the housing market, ethnic composition and diasporas, income and education levels, development plans, nearby services such as schools, religion, unemployment, and land use are the most prominent. According to the relatively limited implemented research in the field of spatial population projections using machine learning techniques (Zoraghein and O'Neill, 2020; Robinson and Dilkina, 2018; McKee et al., 2015), the outcomes are promising and offer great potential for future development.

More specifically, ML-based models promise to benefit spatial projections, extending their possibilities through dealing with multiple variables from large datasets simultaneously and at various scales. They are also more flexible and easily customized compared to the traditional human mobility models (Robinson and Dilkina, 2018). Furthermore, they are able to capture non-linear relationships and more complicated migration dynamics (Robinson and Dilkina, 2018) that may not be visible through traditional means. Lastly, they provide manifold opportunities advancing the temporal and spatial resolution of the analysis and offering higher accuracy and precision.

Our main hypothesis is based on downscaling national population projections to local units by developing novel methodologies for high-resolution projections. A ML-based model trained with gridded sociodemographic data and locally explicit spatial data on infrastructure, networks, social services and development plans could substantially contribute to the configuration of reliable projections of the future distribution of various demographic groups recognizing historical patterns that otherwise would not be visible. This paper shows a ML-based approach for high-resolution gridded population projections by region of origin for Copenhagen, Denmark, by combining various sociodemographic and topographic input layers.

1.1 Migration in Copenhagen

Immigration to Denmark has significantly changed over the years, particularly affecting the city of Copenhagen and the surrounding Capital Region. The growing number of immigrants has raised complexities around socio-spatial seclusion and a long-lasting discussion on ghettos, making Denmark the only country with an official ghetto definition (Freiesleben, 2016) that has recently been changed to *parallel societies*. This section describes the present migration status of the Capital Region of Copenhagen and discusses its spatial footprint.

Denmark has experienced several immigration waves starting in mid and late 1960s and deriving from the Nordic countries, Germany, the United Kingdom, the United States, Yugoslavia, Turkey and Pakistan. The 1990s streams from Vietnam, Chile, Poland, Iran, Palestine, and some African countries (IOM, 2011; Roseveare and Jorgensen, 2004) were followed by vast EU migration in the 2000s. In 2019, immigrants and their descendants numbered 793.601, making up 13.7% of the total population in Denmark (Ministry of Immigration and Integration, 2019). A share of 20% of them resided the Capital Region with the highest concentration of migrants living in Ishøj (40%; Denmark Statistics, 2019).

In particular, according to Statistics Denmark's data, in 2018 more than 305.000 immigrants resided in the Capital Region of Copenhagen, with the biggest group originating from other European, EU countries – 79.312 persons, or 26% of all foreigners in the city. These are mostly foreigners from Germany (9.553), the Nordic countries: Swedes (8.174), Norwegians (7.110), and the United Kingdom (7.824). The second largest category is comprised of Western Asia immigrants (61.041), mostly from Turkey (31.333) and Iraq (13.613), while the Southern Asians consist the third largest category. Figure 1 illustrates the average population density per gridded cell inhabited at parish level, where the central part of the city in deep blue and purple shows high population density and medium and high concentration of migrants respectively. Average migration share exceeds the 75% of the total population in Tingbjerg, one of the named Danish parallel societies. Additionally, a peripheral less densely populated (light purple) zone is observed with lower concentrations of migrants. In this case, the average density of persons per grid cell ranges from 100-200 persons and the average share of migrants limits to 7-21%.
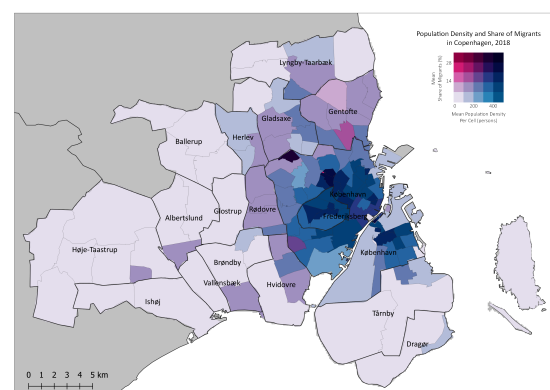

Figure 1. Population Density and Share of Migrants in Copenhagen in 2018

A subject of great interest is presented in the distribution of the following 4 migrant groups in Copenhagen: the European EU, South-Eastern, Western and Southern Asia. The first group, in the best economic situation, chooses the central part of the city with almost half of the migrants in Frederiksberg being born in another EU country. High concentration of South-Eastern Asians is shown in the northern part of the city, which may be connected to the working force deriving from the Philippines in the well-off zones of the city who live with the families where they work. Furthermore, a pattern relating the Western and Southern Asians becomes apparent in the western parishes of the city. These areas – especially the municipality of Ishøj and Høje Taastrup – received a great bulk of migrants with poorer labour market qualifications in the 1990s (Hansen et al., 2015) and in 2000s practiced a distribution policy, whereby housing applicants were assessed according to their income (Freiesleben, 2016).
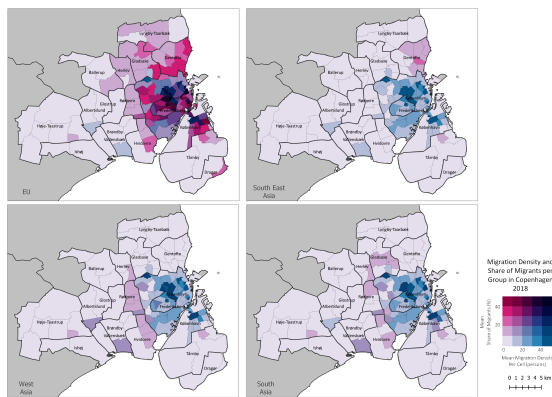

Figure 2. Migration Density and Share of Migrants per group of interest in Copenhagen in 2018

2   Methods

This section presents the available data used for training the ML model and the model architecture. The research is based on previous work by Skaarup Larsen et al. (2018) and a Convolution Neural Network (CNN) is trained with high-resolution demographic and geospatial data. The architecture of the model consists of three consecutive convolutional layers with a 3x3 filter size and 256 filters each and two dense layers with the first obtaining 512 neurons and the second connecting all the neurons into one output; the projection of the migrants' distribution for each year in question. The Leaky Rectified Linear Unit (Leaky ReLU) is used as activation function to receive back only the positive values and the Mean Absolute Error (MAE) as the cost function for the evaluation of the projected data by comparing the model's spatial projections to the projected regional population of the demographic group in question. The Adaptive Moment

Estimation (ADAM) optimizer is used for minimizing the cost function, taking advantage of its progressive adaptations along with the reduction of the learning rate (Skaarup Larsen et al., 2018).

2.1   Software and Data Availability

Extending the capabilities of the PopNet prototype, which produces country-level spatial projections on a 250m resolution grid based on the identification of spatial patterns from historical data on total population, built-up and natural environments, land cover, infrastructure and slope (Skaarup Larsen et al., 2018), we use gridded demographic data, including information about the region of origin, the age, the income, the educational attainments, the rates of natural growth, and topographic data on land use and infrastructure.

The demographic input derives from Statistics Denmark's 1990 to 2018 records for migrants from around the globe divided into 13 groups of interest. The demographic data is protected by Statistics Denmark's data confidentiality policy and is not openly accessible. The data is provided in EXCEL files for each year and category with a unique identifier indicating the coordinates of the centroid of the grid cell it refers to. Even though the data is anonymous and aggregated to 100x100m grid cells, its character remains sensitive and cannot be shared publicly.

In addition, open data on land cover and infrastructure in the capital region of Copenhagen is used. The land cover derives from Copernicus Land Monitoring Service (European Environment Agency (EEA)) as rasters at 100m resolution for all of Europe including 44 classes for 5 time periods. Movia Trafik provides a complete dataset of bus stops including their opening year starting in 2000 in a vector file. The street network and the train stations are provided from Kortforsyningen as vector files. The NUTS3 Administrative Units for 2021 are accessed by Eurostat.

The preparation of the data is done partially by python scripting, tools such as a PostgreSQL/PostGIS database, GDAL and OGR2OGR, in an Anaconda virtual environment. The first processing step includes the reprojection of the datasets of the street and railway network, and the bus stops to the Lambert Azimuthal Equal Area (LAEA) projection and the extraction of the data to the extent for the case study area. All the rest analyses are executed using the LAEA projection. The case study area includes 2 NUTS3 areas in the metropolitan region of Copenhagen with unique ids DK011 and DK012. The train stations file is further processed assigning an additional attribute for the opening year of each station. A layer showing the total number of accessible railway and metro stations in a biking distance of 15' with average biking speed of 15km/h is produced as raster for each year of interest. Similar layers showing the total number of acces-

sible bus stops in a walking distance of 5' with average walking speed of 5km/h are produced for the same period. The land cover data is divided in 5 classes based on the CORINE level 1 categories: artificial surfaces, agricultural areas, forests and semi natural areas, wetlands, water bodies. All the demographic layers are converted to vector and then to raster format. Each group of migrants consists a separate input layer combined with the corresponding open data based on its reference year. This processing results in multiband georeferenced images of 100m resolution grid for each year of study. The implementation is available at: https://github.com/mgeorgati/demo_popnet.

Correlations per inhabited grid cell among the selected layers are examined firstly to avoid data redundancy. The layers of age, income, education, births and deaths refer to total population. Based on the matrix (Figure 3), we observe plausibly negative correlations on the distribution of non-mobile adults and elderly people to almost all the layers. This can partially be explained by the fact that the majority of migrants are at working age choosing central locations, while older Danes tend to move to the suburbs of the city. Another interesting aspect revealed by the matrix is the high positive correlation between high education and income of the total population. Additionally, migrants and rich people, that are defined as people with 10% highest equivalent dispensable income at the age of 25-64 are negatively correlated with an intensifying tendency after 2010. Considering the relations among the migrant groups, the European EU migrants seem to select locations similar to the local population, while Africans, both from Northern and Sub-Saharan Africa show high correlations to Western Asians.
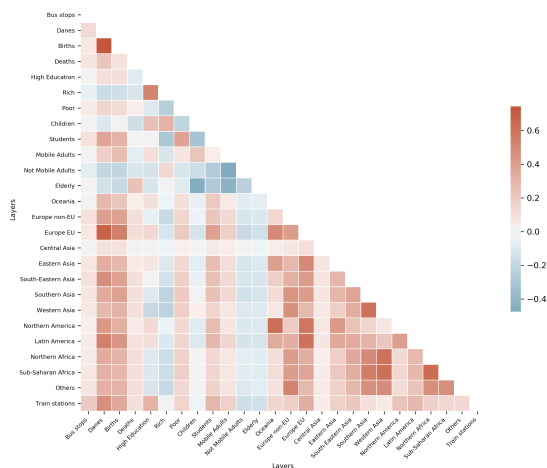

Figure 3. Correlation matrix among demographic layers (2018)

3 Results and Discussion

We are currently performing experiments by using different combinations of input data for the various migrant groups keeping the model architecture unchanged. The input layers cover a 20-year period in 2-year intervals from 1990 to 2010 (only even years included) and projections cover the following four periods until 2018. We tested the model keeping a small number of epochs (e=60, t=80') and the same basic modelling parameters. The batch size that defines the updating frequency of the neural network's weights, is set to 16. The size of the chunk, which is a subdivision of the grid, is of great significance because the model evaluates and predicts each chunk individually and is unable to redistribute people into neighbouring chunks, is set to 32.

The training of the model is performed though utilizing different input sets. The first set of experiments included 25 bands/layers in total containing information about the distribution of the 13migrants' groups, the local population and another 11 demographic characteristics of the total population (age, education, income, births, deaths, marriages). In the second set of experiments the demographic characteristics were replaced by the topographic features, while the third set included both of them.

Among the experiments, we select here to display the produced output of the migrants distribution of the 3rd largest group in Copenhagen, coming from Southern Asia, in comparison to the ground truth data for the corresponding years. Figure 4 shows the historical distribution of the selected group in combination to the total population density in 2012 for reference. The preliminary results show consistency particularly in the case of the 2nd experiment, but reveal a few issues both in the numerical outcomes of the total population of interest and the spatial distribution of this population. Each row in Figure 5 represents the estimated error per grid cell for each one of the performed experiments. The maps display zero error in the transparent layer, while the red and blue palette show over- and underestimation of inhabited grid cells respectively. The light pink and blue, which are dominant in most maps represent the lowest error ranging from 0.01 to 1%, while the dark red (A3,A4) and blue show error exceeding 10%.

Apparently, the model successfully recognises non-inhabited cells, in and out of the case study area. The most reliable output seems to be produced by the 2nd input dataset, which includes only the topographic features. The error is limited up to 1% in most areas, with a tendency in underestimating the population in the densely populated central areas and overestimating the peripheral areas. Numerically, the projected total migrant population in 2012 was 33.000 persons, 11% lower than the historical observation of 37.000
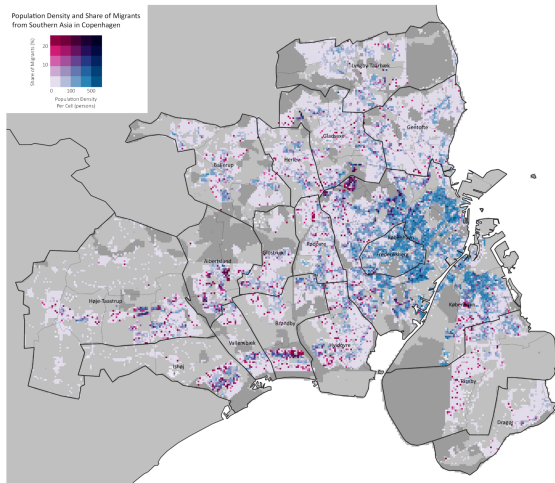
Figure 4. Population Density and Share of Migrants from Southern Asia in Copenhagen in 2012

persons. Although historically the population follows a growth course, the projected population is declining. In the cases of the other two experiments the error is much higher, showing a systematic rise at the last time steps. An interesting pattern is presented in A2, where the error advances (orange/brown) in the peripheral zone surrounding the city center where high concentrations of the selected group are observed. The spreading of underestimation of the projected population in the last experiment (C1-C4), starting from the central part of city and sprawling along the Finger Plan, the urban plan upon which Copenhagen is developed, constitutes another diverting pattern of the produced outputs.
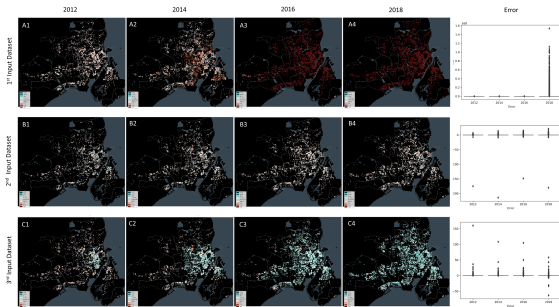

Figure 5. Estimation of the error of the distribution of Southern Asians in Copenhagen

4   Conclusion and Future Work

Summing up, our goal is to develop a flexible, ML-based methodology to project the future distribution of specific socio-demographic groups, and specifically of migrants, in high-resolution grids. We take the advantage of the great data availability that a developed country, such as Denmark, offers, and the special urban

features of Copenhagen with the high water coverage, low elevation, rich history and a unique multicultural character.

The preliminary results have shown some interesting and worth exploring patterns. We believe that with further optimization of the input data and the model architecture, the accuracy and the reliability of the results will improve even more. The next stages of this work will focus on experiments with further input data, such as the housing prices and other building features. The systematic adoption of patterns that creates uniform outputs throughout the case study area should also be explored with aim to enhance the long-term outputs. Regarding the changes in the model architecture, an assessment of the effects of the cost-, activation- and optimization functions is also necessary, along with the advancement of the training time and the estimation of the distribution of the errors, instead of the total loss. Other interesting aspects of future work include an image segmentation approach and the implementation of a CNN combined with time series forecasting for the advancement of the accuracy, especially in the densely populated areas.

References

Denmark Statistics: Immigrants in Denmark 2019 [Indvandrere i Danmark 2019], Tech. rep., https://doi.org/10.1017/CBO9781107415324.004, 2019.

European Environment Agency (EEA): European Union, Copernicus Land Monitoring Service, https://land.copernicus.eu/pan-european/corine-land-cover.

Eurostat: Administrative Units / Statistical Units - GISCO - Eurostat, https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units.

Freiesleben, A. M. v.: Et Danmark af parallelsamfund: Segregering, ghettoisering og social sammenhængskraft: Parallelsamfundet i dansk diskurs 1968-2013–fra utopi til dystopi, 2016.

Grübler, A., O'Neill, B., Riahi, K., Chirkov, V., Goujon, A., Kolp, P., Prommer, I., Scherbov, S., and Slentoe, E.: Regional, national, and spatially explicit scenarios of demographic and economic change based on SRES, Technological Forecasting and Social Change, 74, 980–1029, https://doi.org/10.1016/j.techfore.2006.05.023, 2007.

Hansen, M. F., Schultz-Nielsen, M. L., and Tranæs, T.: The Impact of Immigrants on Public Finances: A Forecast Analysis for Denmark, IZA Discussion Paper 8844, 2015.

IOM, I. O. F. M.: International Organization For Migration, https://www.iom.int/countries/denmark, 2011.

Jones, B. and O'Neill, B. C.: Historically grounded spatial population projections for the continental United States, Environmental Research Letters, 8, https://doi.org/10.1088/1748-9326/8/4/044021, 2013.

Kortforsyningen: Kortforsyningen | Kortforsyningen, https://www.kortforsyningen.dk/.

McKee, J. J., Rose, A. N., Bright, E. A., Huynh, T., and Bhaduri, B. L.: Locally adaptive, spatially explicit projection of US population for 2030 and 2050, Proceedings of the National Academy of Sciences of the United States of America, 112, 1344–1349, https://doi.org/10.1073/pnas.1405713112, 2015.

Ministry of Immigration and Integration: International migration- Denmark, Tech. rep., 2019.

Movia Trafik: Movia, https://www.moviatrafik.dk/.

Robinson, C. and Dilkina, B.: A machine learning approach to modeling human migration, Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies, COMPASS 2018, https://doi.org/10.1145/3209811.3209868, 2018.

Roseveare, D. and Jorgensen, M.: Migration and Integration of Immigrants in Denmark, http://dx.doi.org/10.1787/284832633602, 2004.

Skaarup Larsen, H., Bach-Sørensen, N., and Breilev Lindgreen, T.: Projecting Spatial Population Distribution Using a Convolutional Neural Network, Master's thesis, Aalborg University, 2018.

Zoraghein, H. and O'Neill, B. C.: U.S. state-level projections of the spatial distribution of population consistent with shared socioeconomic pathways, Sustainability (Switzerland), 12, https://doi.org/10.3390/SU12083374, 2020.