Disaggregation of National Level Population Projections to Municipal Level Using a Neural Network Approach

Orlando Olaya Bucaro* (olayabucaro@iiasa.ac.at)
Andrea Tamburini** (Andrea.Tamburini@oeaw.ac.at)
Erich Striessnig*** (erich.striessnig@univie.ac.at)
Claudio Bosco**** (Claudio.bosco@ec.europa.eu)

*International Institute for Applied Systems Analysis, Wittgenstein Centre (IIASA, VID/OeAW, University of Vienna)
**VID/OeAW, Wittgenstein Centre (IIASA, VID/OeAW, University of Vienna)
***University of Vienna, Wittgenstein Centre (IIASA, VID/OeAW, University of Vienna)
****European Commission, Joint Research Centre (JRC)

Abstract

Population projections for small geographical areas are challenging even when data availability is good. Despite the presence of register data in Norway the current municipality level population projections by Statistics Norway are not satisfactory and are in the process of being replaced from a cohort-component framework to microsimulation. We propose a simpler and generalizable approach for downscaling national level population projections into municipality level projections, leveraging Norwegian register data and other data sources using an innovative neural network-based machine learning model. An additional advantage of this downscaling approach is that additional dimensions can easily be added to sub-national projections. We show this by disaggregating the Wittgenstein Centre population projections. The machine learning model is also trained by categorizing municipalities by special economic activities that might affect the population structure in that area. Such activities are the presence of fish farming, oil production, universities, or a high concentration of agricultural production.

Extended abstract

Introduction

Population projections for small geographical areas are challenging even when data availability is good, mostly due to the lack of methods and techniques for subnational projections (Wilson et al., 2022). Such is the case in Norway, despite this the presence of register data in Norway the municipality level population projections by Statistics Norway (SSB) were not considered satisfactory by the statistical agency and SSB are in the process of replacing their municipality level projections from a cohort component framework to microsimulation (Zhiyang et al., 2023).

We propose a simpler and generalizable approach for downscaling national level population projections into municipal level projections, leveraging Norwegian register data and other data sources using an innovative machine learning model. Previous spatial disaggregation methods such as Striessnig et al. (2019) are built under the assumption that there is a lack of data to produce good small area projections, in the case of the US one is often limited to census data. Microsimulation models generally requires large amount of data and big computational capacity (Li & O'Donoghue, 2012). In the subnational microsimulation model for Norway developed by SSB outlined in the technical report by Zhiyang et al. (2023) the input data is every individual living in Norway.

We choose to use Norway as a case study for multiple reasons. The available register data, the range of sizes of municipalities, from big municipalities with millions of inhabitants like Oslo to the small ones like Utsira with around 200 inhabitants. Furthermore, in the case of Norway there is a lack of population projections by education at municipality level. We have therefore disaggregated the Wittgenstein Centre (WIC) population projections to all municipalities in Norway using an innovative machine learning approach that has not been applied before at this geographical scale. The machine learning model is trained using aggregated Norwegian registry data and some municipalities are categorized by special economic activities that might affect the population structure in that area. Such activities are the presence of fish farming, oil production, universities, or a high concentration of agricultural production in the municipality. What makes the WIC population projections attractive is the additional dimension of education, compared to the subnational SSB projections, and the Shared Socioeconomic Pathways (SSP) population scenarios (Lutz et al., 2018).

By comparing the results from our model with the official SSB subnational cohort component projections, the SSB microsimulation results and the SSB national cohort component projections we can assess the validity and usefulness of our disaggregation approach. A spatial disaggregation of this sort could prove useful for applications in need of subnational projections with additional dimensions or SSP coherent scenarios.

Method

The method for the population projection disaggregation is based on the machine learning approach introduced in the paper “Empirically based spatial projections of US population age structure consistent with the shared socioeconomic pathways” by Striessnig et al. (2019). Their approach was based on regression trees to project changes in county age structure based on current and past county demographic characteristics while keeping the projection consistent with the SSPs (Striessnig et al., 2019). The method has later been refined and applied to the disaggregation of the WIC population projections for all the European countries down to NUTS-2 level. The machine learning method was then changed to use a set of gradient boosted regression trees (XGBTree). In this work we refine the approach further by introducing a artificial neural network as well as adding additional information to the model to improve the disaggregation. Additional spatial, discrete, and continuous variables are used as input in the neural network.

Given the distinct nature of our research objectives and the modular approach adopted, we devised unique methodologies for the age disaggregation of the total population and the educational attainment segregation within age groups. This paper will solely detail the methodology concerning age disaggregation, as the educational attainment approach remains under development and is beyond the scope of this abstract.

For spatial disaggregation of age groups, we utilized a modeling framework grounded in artificial neural networks (ANNs) tailored for each age bracket. ANNs, inspired by biological neural networks, comprise interconnected nodes or ”neurons” arranged in layers to execute tasks such as classification and regression. In this research, the primary application of ANNs is for regression, targeting the prediction of continuous outputs based on input datasets. Our available data integrates population structure details, like age distribution at both national and local scales from previous years, and spatially explicit metrics. Notable spatial metrics include regions’ centroid latitude and longitude, the urbanization percentage, and proximity to major Norwegian universities. Given the limited dataset—specific age population data is only accessible from 1990 from the national statistical office—ANNs were chosen for their adaptability and simultaneous handling of diverse variables.

Furthermore, due to the abundance of covariates relative to data points, a meticulous covariate selection was undertaken. Initially, a collinearity study was conducted, followed by a ”jackknife” covariate selection technique. This involved gauging the tangible explanatory power of each covariate and eliminating those that negligibly impacted the model’s R^2 performance, as outlined in Bosco et al., 2018.

Data spanning 2000-2004, 2005-2009, and 2010-2014 were utilized for model training and validation. Through repeated random sub-sampling, we fine-tuned various architecture components, including hidden neuron count, activation functions, and learning rates. Upon identifying a promising model, we assessed its efficacy on 2015-2019 data. Additionally, we incorporated a modified version of the Selective Improvement by Evolutionary Variance Extinction (SIEVE) technique (Castelletti et al., 2005), similar to the method in Bosco et al., 2019. This method aimed to optimize data utility, minimizing the random weight assignment's impact at the ANN estimation onset. Crucially, this method introduces a layer of uncertainty into the model, which would otherwise be overlooked.

Data

In Norway there is register data available with a high level of detail. By using the register data in an aggregated form we can improve the disaggregation and are not required to extract register data for every individual in the Norwegian population.

The projections that are disaggregated are the latest WIC population projections (WIC 2023) for Norway (KC et al., 2023). The WIC population projections are by age, sex and education: five-year age groups and six levels of education. In the disaggregation the age and education levels are aggregated into wider age and education groups due to the limitations of the disaggregation method. The five population scenarios in the WIC projections corresponding to the five SSP scenario narratives are disaggregated to maintain the projections consistent with the SSPs. In figure 1 the five SSPs are compared to to the SSB population projections and the United Nations projections. The WIC SSP projections estimate a much bigger population in the future that both SSB and the UN. This is partiality due to the larger starting population estimate in 2020. However, note that for the SSB and UN projections only the medium scenarios are presented in the figure.

Results

During the testing phase, the model exhibited satisfactory performance, with R^2 values as follows: Under 15 years: 0.87, 15-24 years: 0.6, 25-44 years: 0.78, 45-64 years: 0.85, and 65 years and above: 0.93. Given these results, the models were subsequently utilized for projection exercises.

The results of the desegregation for three large cities in Norway is shown in figure 2 and compared to the regional SSB projections for the years 2025 to 2050. In general the SSP disaggregation estimates follows the same trend as the SSB projections. However, a higher proportion of under 15 year olds and a lower proprtion of 45 to 64 year olds for all three cities is estimated compared to the SSB projection. When doing the comparison we must keep in mind that the WIC projection and SSB differ in several ways. Below are also two maps showing the relative change in the age groups for the yougest and oldes age group. The WIC projections include education, are part of a global population projection model and are aligned to the SSP scenarios. These differences need to be accounted for in the comparison in order to show the performance of our disaggregation model.

Figure 1: Projected total population Norway 2020 - 2050

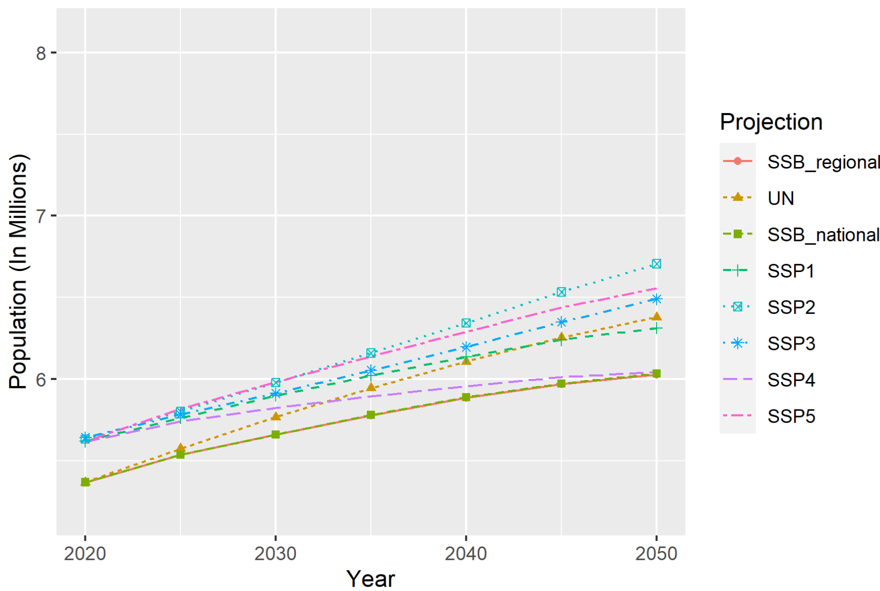
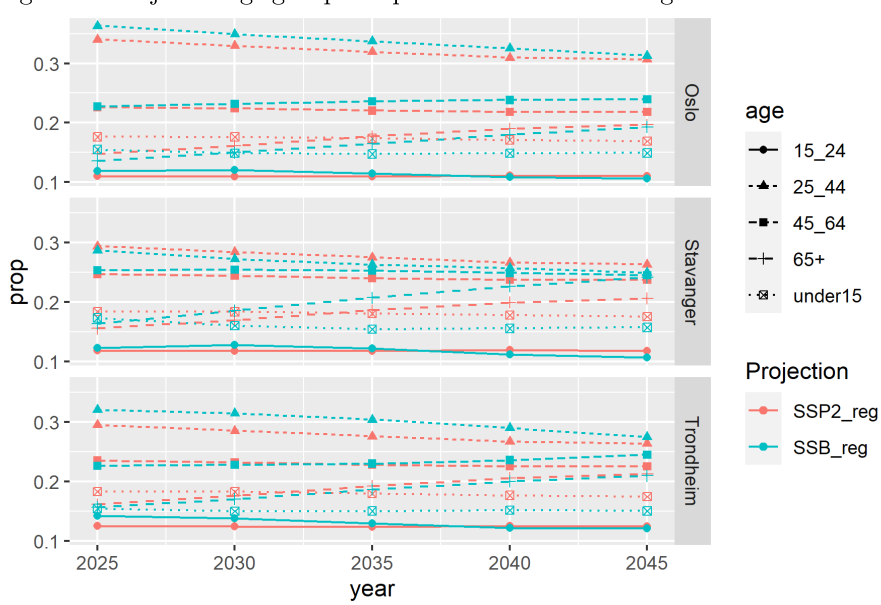
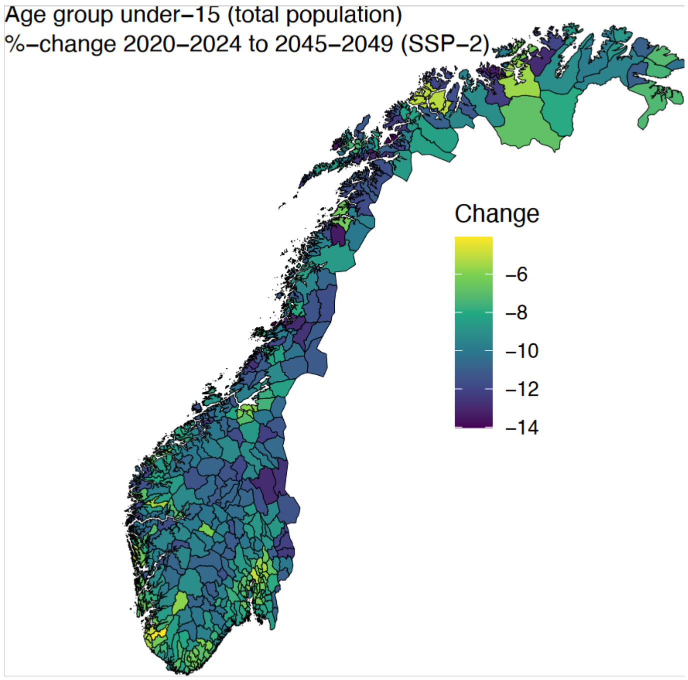
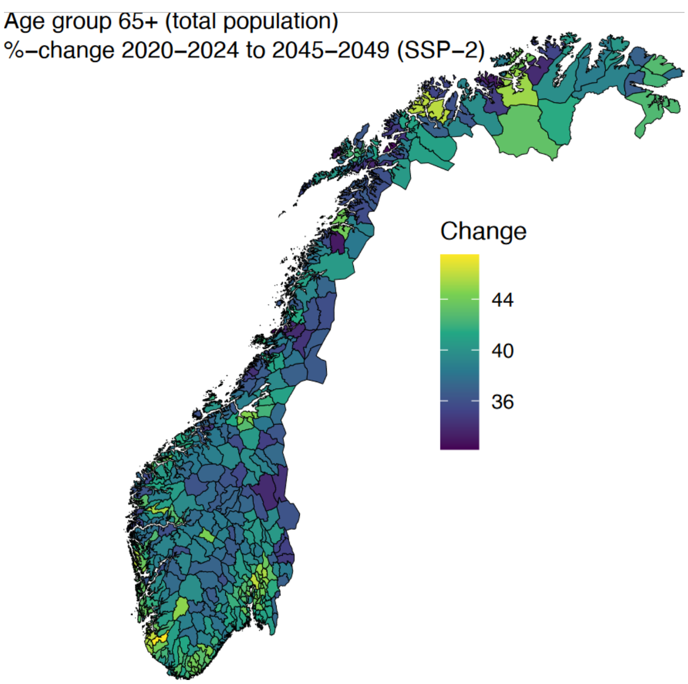


Figure 2: Projected age group composition in three Norwegian cities 2025 - 2050




Age group under−15 (total population)
%−change 2020−2024 to 2045−2049 (SSP−2)
Change
−6
−8
−10
−12
−14


Age group 65+ (total population)
%−change 2020−2024 to 2045−2049 (SSP−2)
Change
44
40
36

References

Bosco, C., Tejedor-Garavito, N., de Rigo, D., Tatem, A. J., Pezzulo, C., Wood, R., Chamberlain, H., & Bird, T. (2018). Geostatistical Tools to Map the Interaction between development Aid and Indices of Need. *AidData Working Paper #49*.

Bosco, C., Watson, S., Game, A., Brooks, C., Rigo, D. D., Greenhalgh, J., Nilsen, K., Ninneman, A., Wood, R., Sarchil Qader, & Bengtsson, L. (2019). Towards high-resolution sex-disaggregated dynamic mapping. Retrieved October 31, 2023, from http://rgdoi.net/10.13140/RG.2.2.12800.79360

Castelletti, A., de Rigo, D., Rizzoli, A. E., Soncini-Sessa, R., & Weber, E. (2005). A selective improvement technique for fastening Neuro-Dynamic Programming in Water Resources Network Management. *16th IFAC World Congress*.

KC, S., Moradhvaj, Potancokova, M., Adhikari, S., Yildiz, D., Mamolo, M., Sobotka, T., Zeman, K., Abel, G., Goujon, A., & Lutz, W. (2023). Wittgenstein Center (WIC) Population and Human Capital Projections - 2023.

Li, J., & O'Donoghue, C. (2012). A survey of dynamic microsimulation models: Uses, model structure and methodology. *International Journal of Microsimulation*, *6*(2), 3–55.

Lutz, W., Goujon, A., KC, S., Stonawski, M., & Stilianakis, N. (2018). *Demographic and human capital scenarios for the 21st century :2018 assessment for 201 countries.* Publications Office.

Striessnig, E., Gao, J., O'Neill, B. C., & Jiang, L. (2019). Empirically based spatial projections of US population age structure consistent with the shared socioeconomic pathways. *Environmental Research Letters*, *14*(11), 114038.

Wilson, T., Grossman, I., Alexander, M., Rees, P., & Temple, J. (2022). Methods for Small Area Population Forecasts: State-of-the-Art and Research Needs. *Population Research and Policy Review*, *41*(3), 865–898.

Zhiyang, J., Leknes, S., & Løkken, S. A. (2023). *Moving beyond expectations. From cohort-component to microsimulation projections* (tech. rep. No. 999). Statistics Norway, Research Department.