Spatial Projections of Age-Structured Populations

Erich Striessnig1, Jing Gao2, Brian O’Neill3, Leiwen Jiang4

1Wittgenstein Centre (IIASA, VDI/OAW, WU) 2University of Delaware 3University of Denver 4ADRI

Background & Motivation
➢ Importance: Spatially-explicit population age structure plays an essential role in population-environment interactions:
- Children and the elderly are particularly susceptible to health stress caused by extreme heat events, climate-related infectious disease, and natural disasters.
- Age structure has important implications for consumption behavior and labor supply affecting societies’ environmental impact.
- Understanding how population age structure may vary over space and time is therefore imperative for better adaptation planning and mitigation in the face of global environmental changes.

➢ Challenge: Spatially-explicit population age structure is highly variable across space and time. As shown in Figure 1, from 1940 to 2010, age structure in the U.S. changed from being spatially highly autocorrelated to highly heterogeneous.

Methodology
➢ An extensive exploratory data analysis informs the development of a county-level model of decadal changes in the fractions of 5 age groups (i.e. 0-19, 20-29, 30-54, 55-69, 70+) that is capable of producing global, long-term population projections.
➢ The best fit to our data was achieved with a model employing a set of Regression Trees, with one tree for each age group. The combined predicted fractions from the five trees are then rescaled to sum up to 1 using iterative proportional fitting (IPF).
➢ Figure 2 shows the structure of the tree for the 0-19 age group. Repeatedly splitting the sample by the most relevant variable, counties end up in different end-nodes determining positive or negative change in this proportion relative to national and changes in the 0-19 share for that node.

➢ Being a non-parametric model, regression trees are capable of integrating complex non-linear relationships between response and predictor variables.
➢ Cluster analysis suggests the existence of a continuum of age profiles among U.S. counties that can be summarized into 4-5 distinctive clusters at any historical time point. The tree models allow us to handle all profiles within one framework.
➢ Over time, counties’ age profiles change, e.g. due to the aging of the baby boomers. The tree model allows for such changes by using different branches to make projections for the same county at different times depending on its prior conditions.
➢ The tree model is robust to extreme values, both in learning from the experience of the past and in predicting forward, as the predicted decadal changes cannot exceed what has been observed at most in the past. This is particularly useful for long-term modeling, as what are viewed as extreme values currently may happen much more often under certain future scenarios.

Results
➢ Figure 4 shows observed (1940-2010) and projected (2020-2100) proportions of total county population in five different age groups for SSP2.
➢ In accordance with the general trend of population aging predicted for the US under all five SSPs, the proportions of children under the age of 20, as well as the population of working age (30-54), decrease over time.
➢ In contrast, there is a steady increase in the population share above age 70.

Future Steps
➢ Possible application to other contexts where sub-national variation in age-structure would matter a lot (India, China).
➢ Possible extension to include education sub-nationally which is already included in the national level SSPs.
➢ Include results from ongoing work on state-level population projections further downscaled to the grid level (Jiang, Zaroghin, et al., forthcoming)

Contact: Erich.Striessnig@oeaw.ac.at and Brian.O’Neill@du.edu

Figure 1. Clusters of U.S. counties with similar age structures. Note that clusters are not comparable across years.

Figure 2. Regression tree for 0-19 age group showing most relevant splits. Numbers in bubbles show percentage of 3103 counties in each end-node.

Figure 3. Validation fit for the 1990-2000 period based on the regression trees developed based on the 2000-2010 experience. County level residuals for the proportion aged 0-19 projected for the year 2000 (left), as well as R^2 in five different age groups (right).

Figure 4. Observed (1940-2010) and projected (2020-2100) proportions of population at the US county level in five different age groups under SSP2.

The substantial aging of the population even in the Middle of the Road scenario is reflected spatially throughout the entire conterminous US (Figure 5, top row).
- Counties around large cities, which attract large amounts of working age population and their kids, maintain a relatively high proportion of youth.
- The thinly populated counties in the Midwest and the Rocky Mountains experience aging most drastically.
- SSP2 and SSP3 have similar shares of national population in the 70+ age group late in the century, but most counties have a larger fraction of population in this age group in SSP3 than in SSP2 (bottom left).
- Both SSP3 and SSP5 (bottom row) produce a more intense concentration of the younger population in urban areas than in SSP2.
- The older age structure in SSP5 is concentrated in Western and Midwestern states compared to SSP2 (bottom right).

Figure 5. Observed (2010) and projected (2100) proportion 70+ under SSP3 (top), Difference in proportion 70+ between SSP3 and SSP2 (bottom left) and SSP5 and SSP2 (bottom right), U.S. in 2100.