SPARCCLE

Socioeconomic Pathways, Adaptation and Resilience to Changing Climate in Europe

Deliverable number	**3.2**
Deliverable name:	Multidimensional socioeconomic projections report
WP / WP number:	3
Delivery due date:	31.08.2025
Actual date of submission:	29.08.2025
Deliverable description:	Report on the multidimensional (MD) socioeconomic (soc) projections, such as SSP populations by age/sex/urban/rural downscaled to NUTS3 level using a spatial population downscaling gravity-model and further disaggregated by education & poverty levels using advanced methods. The output data will include both historical observations and projections to 2100 according to several scenarios of societal change, which will be discussed in the report, drawing on the analysis on Task 3.1. The deliverable will be submitted by default as a Report (SEN) on the condition that public version of the work will be published without significant delay as an open-access pre-print. If work already published as an open-access article, Deliverable leader can choose whether SEN or PU. Otherwise the deliverable should be submitted as Report (PU). Related task: Task 3.1
Dissemination level:	SEN
Lead author(s):	Andrea Tamburini, Anne Goujon
Contributors:	

Version	Date	Issued by	Description	Summary of changes
0.1	25.07.2025	Andrea Tamburini, Anne Goujon	First Draft	Initial draft
0.2	30.07.2025	Greta Shum, Giovanni Forzieri	Review Version	Reviewed draft
0.3	27.08.2025	Andrea Tamburini, Anne Goujon	Final Version	Incorporated feedback from internal reviewers

1. Changes with respect to the DoA
None

2. Dissemination and uptake
Disseminated among the consortium members.

3. Short Summary of results (<250 words)
This report delivers a new multidimensional dataset providing population projections at the NUTS-2 regional level across Europe. The projections capture five age groups (under 15, 15—24, 25—44, 45—64, and 65+), disaggregated by sex and, from age 25 onward, further differentiated by educational attainment (low, medium, high, based on ISCED categories).
The modelling framework applied combines different approaches to reflect demographic complexity. Age structures were estimated using a neural network architecture, educational attainment distributions were generated with a Bayesian time series model, and sex ratios were derived by scaling national-level patterns to the regional level. Together, these methods ensure consistency with established demographic projections while adding crucial sub-national granularity.

The resulting dataset provides a new evidence base for examining the demographic drivers of exposure, vulnerability, and adaptation capacity to climate change in Europe. It supports gender-sensitive and education-specific analyses of risk and offers a robust foundation for forward-looking adaptation and resilience strategies.

4. Evidence of accomplishment
Document submitted.

Multidimensional socioeconomic projections report

Author(s):

Andrea Tamburini, Anne Goujon

Funded by the European Union

Funded by the European Union under grant agreement No 101081369 (SPARCCLE). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or HORIZON-RIA — HORIZON Research and Innovation Actions. Neither the European Union nor the granting authority can be held responsible for them.

Executive Summary

Understanding the future risks posed by climate change requires realistic projections of population dynamics, not only in terms of overall size but also in their detailed demographic and socioeconomic composition. Exposure to hazards depends both on where people live and, on their characteristics, such as age, sex, and educational attainment, which shape vulnerability and adaptive capacity. Despite their importance, projections at the sub-national level that integrate these dimensions remain scarce.

This report presents a multidimensional dataset that disaggregates the Wittgenstein Centre's population projections to the NUTS-2 regional level across Europe. The dataset provides projections for five age groups (under 15, 15—24, 25—44, 45—64, and 65+), disaggregated by sex, and—starting from age 25—further differentiated by three levels of educational attainment based on ISCED categories: low (0—2), medium (3—4), and high (5—8). To achieve this, different modelling approaches were employed: age distributions were derived using a neural network architecture, educational attainment through a Bayesian time series framework, and sex ratios by applying national-level patterns scaled to the regional context.

This work forms part of WP3, which examines the socioeconomic factors that shape exposure, vulnerability, and adaptive capacity to climate change. By producing granular socioeconomic scenarios at national and sub-national levels, the outputs strengthen the evidence base for assessing vulnerabilities with Earth observation and AI methods, analysing gender-specific risks, and supporting inclusive, long-term adaptation strategies.

Key Words

Disaggregation, multidimensional, machine learning, Bayesian time series

Table of Contents

List of Figures

1 Summary

Improving our understanding of future risk from climate change requires realistic projections of future populations, both in their size and distribution. Distribution refers not only to geographic breakdowns but also to the breakdown by important characteristics, such as age, sex, and educational attainment. While the location where people will live may determine future exposure to hazards, population characteristics also co-determine the degree of vulnerability and the capacity to adapt to changing environmental conditions. Despite the importance of these factors, there remains a paucity of population projections (or disaggregation thereof) at the sub-national level.

This report details the modelling procedures employed to generate a multidimensional dataset that disaggregates the Wittgenstein Centre's population projections[1] at the NUTS-2[2] regional level. The disaggregation covers five age groups (under 15, 15—24, 25—44, 45—64, and 65+), two sexes (female and male), and—starting from age 25—three levels of educational attainment, corresponding to ISCED[3] 0—2 (low), ISCED 3—4 (medium), and ISCED 5—8 (high).

The age was derived using a neural network architecture; the educational attainment dimension was derived a Bayesian time series modelling approach, and the sex dimension was obtained by applying national-level sex ratios, adjusted and scaled to the regional level.

The report is organized into three chapters, each addressing one of the dimensions—age, sex, and educational attainment.

This work is part of WP3, which addresses the socioeconomic factors shaping exposure, vulnerability, and adaptive capacity to climate change. Our contribution focuses on developing granular socioeconomic scenarios for the EU at national and sub-national levels. These scenarios provide critical inputs for assessing vulnerabilities with Earth observation and AI methods (Task 3.2) and for analysing gender-specific vulnerabilities and societal transformations (Task 3.3). By integrating detailed demographic, social, and economic data, the work strengthens the evidence base for more inclusive and effective adaptation strategies.

---

[1] https://pure.iiasa.ac.at/id/eprint/19487/ and https://dataexplorer.wittgensteincentre.org/wcde-v3/

[2] We limited our disaggregation to the NUTS-2 level due to severe data limitations regarding age structure and educational attainment at the NUTS-3 level; extending the analysis to this finer spatial scale remains a planned direction for future development.

[3] https://ec.europa.eu/eurostat/statistics-explained/index.php?title=International_Standard_Classification_of_Education_(ISCED)

2 The age dimension

Understanding the age structure of future populations is essential for assessing vulnerability and adaptive capacity in the face of climate change (Yao et al. 2025; Sestito et al. 2025; Chen et al. 2024; Achebak et al. 2019). While most projections focus on national-level totals, sub-national age distributions are equally critical, particularly in regions like Europe where both demographic and climatic conditions vary substantially across space. To address this gap, we developed a machine learning-based model to downscale age-specific population projections—based on the Shared Socioeconomic Pathways (SSPs)—to the NUTS-2 level across 35 European countries[4]. This section describes the methodology and outcomes of this disaggregation, which supports more realistic and spatially-explicit assessments of future climate-related vulnerability.

2.1 The data sources

1- The Eurostat Database[5]: provides the NUTS-2 breakdown of the age-specific baseline population across 35 countries. The information is reported yearly starting from 1990 until 2023 and for five-yearly age groups from "under 5" up to "75+". For a detailed overview of data coverage, see Table 1 in the Appendix.

2- The Wittgenstein Centre Human Capital Data Explorer[6] (WCDE): provides SSP-coherent age-specific population information at the country level for 2020-2100 (forthcoming, also for the historical period 1950-2015) (KC et al. 2024).

The different SSP scenario assumptions for the population age structure in Europe can be summarized as follows:

[4] AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, EL, ES, FI, FR, HR, HU, IE, IS, IT, LT, LU, LV, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, UK.

[5] https://ec.europa.eu/eurostat/databrowser/explore/all/all_themes?lang=en&display=list&sort=category

[6] https://dataexplorer.wittgensteincentre.org/wcde-v3/

Scenario	Fertility	Mortality	Migration	Comments
SSP1 – Sustainability / Rapid Social Development	Low (Low10)[7]	Low	Medium	Very low fertility and longer lifespans, even with moderate immigration, lead to strong population aging across most European countries.
SSP2 – Middle-of-the-Road / Continuation	Medium	Medium	Medium	Fertility remains below replacement but declines more slowly than in SSP1. Combined with medium mortality and migration, this produces population aging at a more gradual pace.
SSP3 – Fragmentation / Stalled Development	Low	High	Low	Low birth rates, shorter lifespans, and limited immigration drive both population decline and significant aging.
SSP4 – Inequality	Low	Medium	High	Extremely low fertility strongly accelerates aging. High migration introduces younger cohorts, slightly moderating but not reversing the trend toward an older population.
SSP5 – Conventional Development	Low (Low10)	Low	Medium	Similar to SSP1: very low fertility and longer lifespans drive significant aging, despite some immigration.

[7] The Low10 fertility scenario is an additional low-fertility scenario where education-specific Total Fertility Rates (TFRs) are 10% lower than the medium fertility assumptions up to the year 2040, with the difference subsequently increasing to 12.5% lower than the medium fertility assumptions by 2060 and remaining at that level until 2100. This scenario is specifically used in SSP1 and SSP5 for low-fertility countries.

3- Harvard Dataverse [8]: provides the "Global 1- km Downscaled Urban Land Fraction Grids, SSP-Consistent Projections and Base Year, v1 (2000 - 2100)" (Gao and Pesaresi 2021) and the "Global 1-km Downscaled Population Grids, SSP-Consistent Projections and Base Year, v1.01 (2000 - 2100)" (Gao 2017; 2020). This downscaled information can be grouped at the NUTS-2 level to generate total population counts and urban land fractions to be used in the construction of the independent variables.

2.2 Modelling Approach

To produce SSP-consistent age disaggregation at the NUTS-2 level, we followed a strategy inspired by previous work (Striessnig et al. 2019) that develops separate models for each age group. This modular approach increases accuracy by reducing model complexity and allowing for targeted tuning. We focused on five age groups: under 15, 15—24, 25—44, 45—64, and 65+, chosen for their relevance in terms of labour market dynamics (Loichinger 2015), education (Lutz et al. 2008), and climate vulnerability (Otto et al. 2017). Based on this categorization, we developed a disaggregation model tailored to each age group.

We employed feedforward artificial neural networks (ANNs) (Carvalho et al. 2011) to model the age-group shares at the regional level. These models are well-suited to capture complex, non-linear relationships and allow for flexible integration of multiple scenario pathways. Key covariates included regional-level demographic indicators derived from Eurostat and urbanization and accessibility measures from Harvard Dataverse. SSP-based national projections from the Wittgenstein Centre served as the top-down constraint.

Given the limited training data (for each age group, we obtain a data set of 1245 observations) and the risk of overfitting, we used a simplified version of the Selective Improvement Evolutionary Variance Extinction (SIEVE, De Rigo et al. (2005)) framework to stabilize the training process and improve robustness. This approach progressively selects and refines the best-performing feedforward neural networks—based on mean absolute error—to efficiently optimize the choice of initial weights across age-specific models.

For full methodological details and validation results, please refer to Tamburini, Bosco and Striessnig (2025, forthcoming)[9].

To generate subnational age-group projections aligned with the SSPs, we modelled relative age-group proportions rather than absolute population counts. Specifically, the dependent variables were defined as the ratio of a region's age-specific population share to the corresponding national share—allowing

[8] https://dataverse.harvard.edu/

[9] Tamburini, Andrea, Claudio Bosco and Erich Striessnig. 2025 (forthcoming). A neural network architecture for disaggregating age-specific population projections to the sub-national level. IIASA Working Paper. Laxenburg, Austria: WP-25-xx

for intuitive interpretation and consistent scaling with external population totals. These ratios were modelled separately for each of five age groups using a tailored feedforward neural network (Simp-SIEVE-NN), trained on a set of 53 covariates that included historical age structure trends, urbanization, national-level features, and regional clustering (for a detailed list of the covariates see Table 3 in the Appendix.). Extensive preprocessing steps—such as correlation filtering, standardization and covariates selection—were applied to reduce redundancy and ensure robust training.

The simplified SIEVE approach, adapted for this task, enabled both improved performance and uncertainty quantification. Compared to two baseline models—one that predicts by repeating the most recent observation and another that predicts using the average of the previous five ratios—our method achieved, in testing phase, substantially higher explained variance (Expl. Var. > 0.856 for all age groups), with consistent uncertainty bounds across predictions. Even in more challenging groups like 15—24, where spatial mobility is high, the model remained unbeaten by the selected competitors. By incorporating multiple neural networks, the Simp-SIEVE-NN framework captures the variability in model behaviour, represented by prediction intervals. For a complete technical description of the model design, training process and validation, please refer to the accompanying working paper (Tamburini et al. (2025), forthcoming).

2.3 Results

Across all Shared Socioeconomic Pathway (SSP) scenarios, Europe is projected to undergo pronounced population aging, driven by persistently low or declining fertility rates and increasing life expectancy. The extent and pace of this demographic shift vary by scenario, with the most severe aging observed under low-fertility, low-migration assumptions. Building on this understanding, we used national-level population structures, SSP-specific projections for total populations at the NUTS-2 level, and regional urbanization rates to apply our five age-specific models recursively. This enabled us to generate detailed subnational disaggregations of the national-level results.

As reported in Figure 1, the findings confirm that aging will affect all European regions, but its progression is uneven. Socio-economic characteristics strongly influence the trajectory, with regions that include or coincide with national urban centres typically experiencing a slower rate of aging. These spatial disparities underline the importance of considering regional heterogeneity in future planning and adaptation efforts.

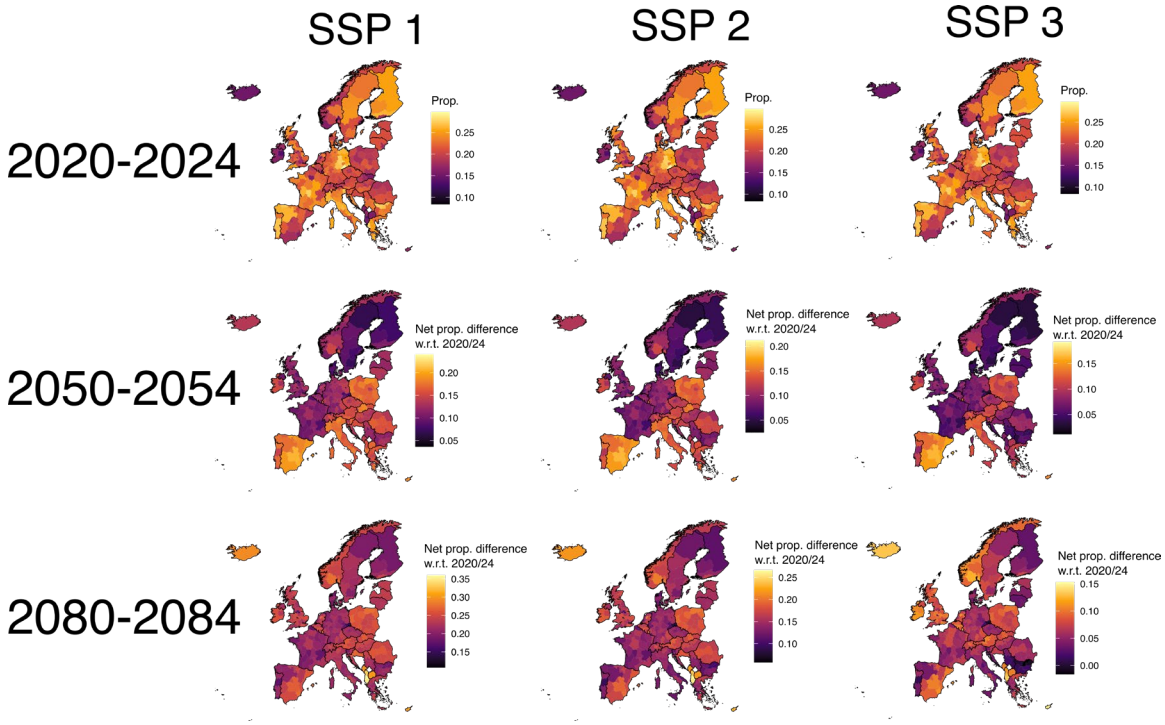

Figure 1: Population projections, age group 65+ for different SSPs and time period. For interval 2020-2024 the plot reports the proportion, for the projected ones, the difference between the 2020-2024 one and its future proportion in the different SSPs.

3 The sex dimension

The sex dimension, together with age, plays a fundamental role in determining sensitivity to climate change, particularly in the context of demographic aging. Women's higher observed life expectancy contributes to a feminization of older age cohorts, which has direct implications for health care needs, dependency ratios, and overall vulnerability. During pregnancy, exposure to extreme heat or air pollution can increase the risk of complications such as preterm birth, low birth weight, and developmental issues (Pinho-Gomes et al. 2024; Achebak et al. 2019; Vésier and Urban 2023). Recognizing this, we extended our projection framework by incorporating sex-specific information consistently across time, space, and scenario.

3.1 The data sources

For this exercise, the data are obtained from two sources:
1- The Eurostat Database: provides the NUTS-2 breakdown of the age-SEX-specific baseline population across 35 countries. The information is reported yearly starting from 1990 until 2023 and for five-yearly age groups from "under 5" up to "75+". For a detailed overview of data coverage, see Table 1 in the Appendix.
2- The Wittgenstein Centre Human Capital Data Explorer (WCDE): provides SSP-coherent age-Sex-specific population information at the country level for 2020-2100.

3.2 Modelling Approach

Instead of developing a new model for the sex disaggregation, we adopted a pragmatic approach that leverages existing national-level projections from the Wittgenstein Centre, which include age-specific sex ratios under the five SSP scenarios. These national trajectories were then aligned with regional-level historical data on sex ratios by age group, sourced from the Eurostat database, ensuring consistency with observed patterns.
Using this alignment, we applied SSP-specific national trends to adjust the most recent observed regional sex ratios and projected their evolution forward in time. These adjusted ratios were then used to disaggregate the total population projections by age group (produced in the previous modelling step) into male and female components for each NUTS-2 region, scenario, and time interval.
In other words, for each age group, the national level future development of the sex ratios was applied to the regional level sex-ratios under the assumption that they move, independently of their final observed value according to the national level schedule.
The result is a fully harmonized dataset capturing population projections by sex and age group at the subnational level, consistent with both observed regional structures and the overarching scenario narratives. This method offers a coherent and efficient way to reflect differential demographic

dynamics while maintaining internal consistency with the SSP framework and ensuring compatibility across modelled dimensions.

We do not explicitly model the evolution of sex ratios at the subnational level for each age group. Instead, we rely on the national-level trends provided by the Wittgenstein Centre projections and apply these uniformly across regions. This simplification means that potential region-specific deviations in male and female dynamics—such as sex-selective migration patterns—are not directly captured. Nonetheless, the method still robustly reflects differences in life expectancy between men and women and ensures consistency with the SSP-specific assumptions, making it a reasonable and scenario-coherent compromise given the available data.

3.3 The results

The results of the sex ratio modelling for European regions show the expected age-related pattern. At younger ages, the sex ratio is close to parity, with values around 0.5 reflecting the near balance between men and women at birth and in early adulthood. In contrast, for the age group 65+, the ratio shifts clearly in favour of women, a direct consequence of their higher life expectancy compared to men. This imbalance is illustrated in Figure 2 for the NUTS-2 regions of Slovakia, where the predominance of women in older ages is evident but gradually flattens toward parity over the course of the century, in line with the projected improvements in male life expectancy.

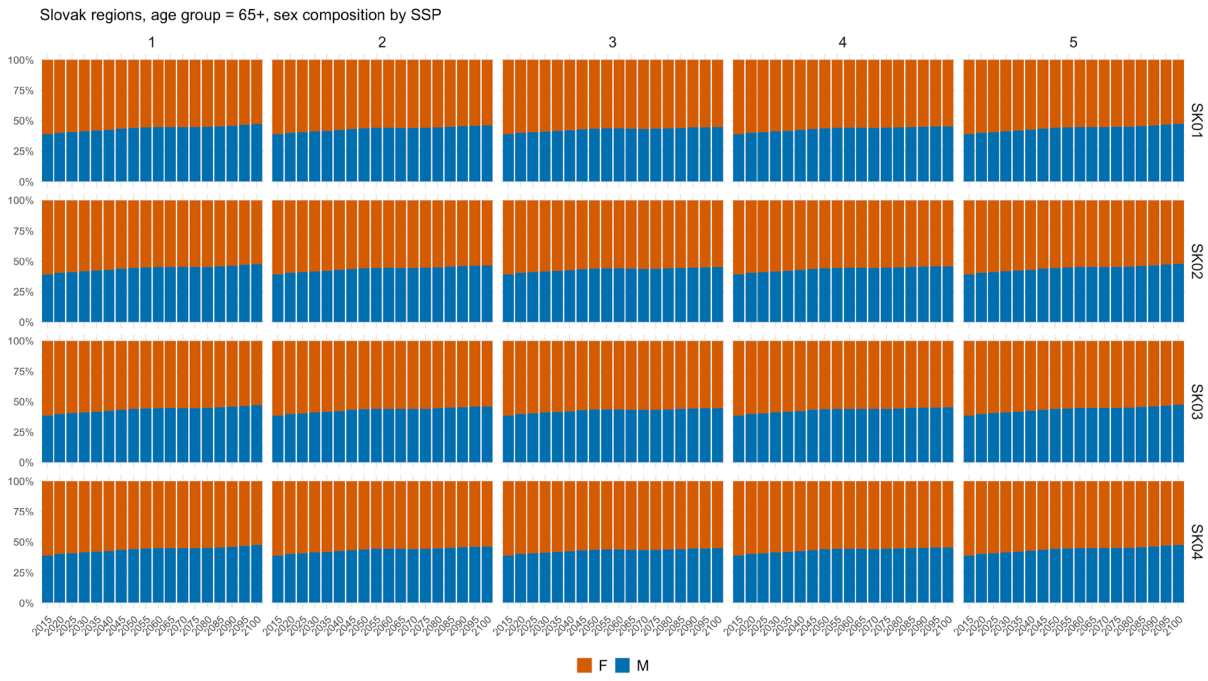

Figure 2: Sex ratio for age group 65+ in the Slovak regions according to SSP scenarios

4 Educational attainment dimension

Education plays a decisive role in shaping adaptive capacity to climate change and climate extremes, not only by strengthening access to information, knowledge, and skills necessary for adaptation, but also indirectly through its association with higher income, better employment opportunities, and enhanced access to resources; higher education is frequently associated with greater awareness and knowledge of risk prevention, whereas low educational attainment has been directly linked to increased mortality risk during extreme events such as heat hazards, thereby making the level of education a key factor in reducing vulnerability and improving resilience (Yang et al. 2021; Sestito et al. 2025; Reid et al. 2009)

4.1 The Data Source

The dataset *lfst_r_lfsd2pop* from Eurostat provides regional population estimates (in thousands) disaggregated by sex, age group, and educational attainment level at the NUTS-2 level, based on the EU Labour Force Survey. Educational attainment is reported using the ISCED classification (low, medium, high), enabling a multidimensional analysis of regional population structures across Europe. For our analysis, we were able to extract consistent data for 21 countries, covering all the NUTS-2 regions included in the age dimension analysis—with the exception of FI20 (Åland) and the overseas departments of France, due to missing data (For a detailed overview of data coverage, see Table 2 in the Appendix). Despite some gaps and sampling variability, particularly in smaller regions and earlier years, this dataset remains the most harmonized and comprehensive source for regional educational attainment available at the European level and serves as the historical input for our modelling.

4.2 Modelling Approach

Modelling educational attainment distributions over time requires a flexible and robust statistical approach, especially when accounting for heterogeneity across countries, sexes, and demographic scenarios. Traditional models often work with level-based latent processes; however, log-returns, defined as first differences of log-transformed shares, offer several advantages in dynamic compositional modelling:

- Variance stabilization: Log-returns reduce heteroskedasticity common in share-based data.
- Additivity: They allow modeling changes additively in log-space, aligning well with autoregressive structures.
- Compositional compatibility: When exponentiated and normalized, they map naturally back to the unit simplex.

Building on this data transformation, we employed a Bayesian time series modeling framework. This class of models was chosen because of its ability to both handle and explicitly quantify uncertainty in the input data, which in our case are derived from survey sources and therefore subject to sampling variability and measurement error. Beyond uncertainty quantification, Bayesian time series models also allow the incorporation of prior information, hierarchical structures, and flexible parameter estimation, making them particularly suitable for cross-country and multi-dimensional applications. Our specification extends a latent-trend multinomial framework into the log-return space, capitalizing on the benefits of log-returns for capturing relative changes while introducing stratification by country (N), sex (S), and scenario (P). After the usual MCMC posterior sampling of predictive counts from the multinomial model, we applied a correction procedure to align the results with two key sets of external marginal constraints—regional population projections (by country, region, sex, age, and scenario) and national education category projections (by country, sex, category, age, and scenario). This adjustment was implemented using a two-dimensional Iterative Proportional Fitting (IPF) algorithm enhanced with entropy regularization, ensuring internal consistency while preserving the statistical structure of the sampled distributions. The resulting structure is robust to heterogeneous regional dynamics and varying starting years across countries, ensuring greater comparability and reliability of trend estimates across diverse demographic and geographic contexts.

4.2.1 Notation

Symbol	Description
N	Number of countries
R_n	Number of regions in country n
S	Number of sexes
T	Number of time points
A	Number of age groups
C	Number of education categories (fixed at 3)
P	Number of projection scenarios
$t_{init}[n]$	Initial modeled year for country n
$log_{p,n,r,s,c,t,a}$	Log-probability for country n, region r, ...
$logreturn_{p,n,r,s,c,t,a}$	First-order difference of log-probabilities
$N_{n,r,s,t,a}$	Region's population by age

4.2.2 Time Rescaling

Each country's time series is rescaled to $[0,1]$ to support cross-national trend pooling:

$$time\_scaled_{n,t} = \frac{t - t_{init}[n]}{T - t_{init}[n]}$$

4.2.3 Priors

We place country- and sex-specific priors on both latent initial states and return dynamics:

$$\tau\_init_{n,s,a} \sim \Gamma(1,0.1)$$
$$\tau\_return_{n,s,a} \sim \Gamma(1,0.1)$$
$$\mu\_\beta_{n,s,c,a} \sim N(0,1)$$
$$\tau\_\beta_{n,s,c,a} \sim \Gamma(1,0.1)$$
$$\beta_{n,r,s,c,a} \sim N\left(\mu\_\beta_{n,s,c,a}, \tau\_\beta^{-1}{}_{n,s,c,a}\right)$$

4.2.4 Latent Process in Log-Return Space

4.2.4.1 Initialization

At the initial year $t = t_{init}[n]$, log-probabilities are initialized as:

$$\log p_{n,r,s,c,t_{\text{init}}[n],a} \sim \mathcal{N}\left(0, \tau^{-1}_{\text{init},n,s,a}\right)$$

4.2.4.2 Dynamic Evolution

From $t = t_{init}[n] + 1$, onward, log-returns are drawn and accumulated additively:

$$\text{log-return}_{n,r,s,c,t,a} \sim \mathcal{N}\left(\beta_{n,r,s,c,a}, \tau^{-1}_{\text{return},n,s,a}\right)$$
$$\log p_{n,r,s,c,t,a} = \log p_{n,r,s,c,t-1,a} + \text{log-return}_{n,r,s,c,t,a}$$

4.2.5 Probability Transformation and Likelihood

After computing log-probabilities for c = 1, 2, the third category is derived via normalization:

$$\alpha_{n,r,s,1,t,a} = 1$$
$$\alpha_{n,r,s,2,t,a} = \exp\left(\log p_{n,r,s,1,t,a}\right)$$
$$\alpha_{n,r,s,3,t,a} = \exp\left(\log p_{n,r,s,2,t,a}\right)$$
$$Z_{n,r,s,t,a} = \sum_{c=1}^{3} \alpha_{n,r,s,c,t,a}$$
$$p_{n,r,s,c,t,a} = \frac{\alpha_{n,r,s,c,t,a}}{Z_{n,r,s,t,a}}$$
$$y_{n,r,s,t,a} \sim \text{Multinomial}\left(p_{n,r,s,1:3,t,a}, N_{n,r,s,t,a}\right)$$

In forward projections where total population varies by scenario:

$$pp_{n,r,s,c,t,a,p} = p_{n,r,s,c,t,a} \cdot N_{n,r,s,t,a,p}$$

This formulation preserves model uncertainty while enabling scenario-based benchmarking.

4.2.6 Model Highlights and Advantages

- **Log-returns** provide a more stable and interpretable structure for temporal dynamics in compositional data.
- **Sex** and **country** dimensions enable nuanced trend estimation across diverse demographic structures.
- Handles varying region counts via per-country regional indexing R_n.
- Capable of integrating alternative future scenarios (P) without retraining.

4.2.7 Post-Processing: Entropy-Regularized IPF Adjustment

After sampling posterior predictive counts $pp_{n,r,s,c,t,a,p}$ from the multinomial model, we apply a correction procedure to ensure **coherence** with two types of external marginal constraints:

1. Regional population projections $N_{n,r,s,t,a,p}$ (by country n, region r, sex s, age a, scenario p)
2. National education category projections $\mathrm{PoPnat}_{n,s,c,t,a,p}$(by country n, sex s, category c, age a, scenario p)

This is achieved through a **two-dimensional Iterative Proportional Fitting (IPF)** algorithm enhanced with **entropy regularization**.

4.2.7.1 Algorithm Overview

Let $X^{(0)} = \mathrm{pp}^{\mathrm{sample}}_{n,r,s,c,t,a,p}$denote the initial $[R_n \times C]$ matrix for a given iteration. At each iteration z, we perform:

Row normalization:

$$X_{r,c}^{(z+1/2)} = X_{r,c}^{(z)} \cdot \frac{N_{n,r,s,t,a,p}}{\sum_c X_{r,c}^{(z)}}$$

Column normalization

$$X_{r,c}^{(z+1)} = X_{r,c}^{(z+1/2)} \cdot \frac{\mathrm{PoPnat}_{n,s,c,t,a,p}}{\sum_r X_{r,c}^{(z+1/2)}}$$

Entropy − based regularization:

$$X_{r,c}^{(z+1)} \leftarrow X_{r,c}^{(z+1)} \cdot \exp\left(\lambda \cdot \log\left(\frac{X_{r,c}^{(z+1)}}{X_{r,c}^{(0)}}\right)\right)$$

Where λ is a regularization parameter.

4.2.7.2 Entropy Regularization

The entropy penalty acts as a **stability constraint** to prevent over-fitting of marginal targets. It preserves relative proportions from the initial posterior while still enforcing the desired marginal structure. Formally, it is inspired by minimizing the *Kullback-Leibler divergence* (Kullback and Leibler 1951) between the updated matrix and the original posterior draw.

This leads to solutions that are:

- Numerically stable
- Less prone to over-adjustment
- Consistent with both data-driven and projection-derived marginals

4.2.7.3 Application to Scenario Coherence

Since the marginal totals $N_{n,r,s,t,a,p}$ and $\mathrm{PoPnat}_{n,s,c,t,a,p}$ are SSP scenario (p) coherent, the IPF procedure inherently enforces **scenario alignment** for the posterior estimates.

Each posterior sample is thus reweighted per-scenario, ensuring internal consistency of subnational projections with national-level assumptions across all dimensions: country, sex, age, and scenario.

4.2.7.4 Convergence and validation

To ensure the reliability of the Bayesian time series model, we assessed convergence using standard diagnostics, including the Gelman-Rubin statistic ($\hat{R}$) and visual inspection of trace plots across chains. All key parameters showed satisfactory convergence, with $\hat{R}$ values close to 1 and stable, well-mixed traces. In terms of validation, due to the lack of suitable gold-standard data for direct comparison, we employed an indirect validation strategy. Specifically, we verified the internal coherence of the results by comparing aggregated projections to national-level distributions by educational attainment, as well as to regional distributions disaggregated by age and sex. This consistency across hierarchical levels and dimensions supports the plausibility of the model outputs and increases confidence in their use for scenario-based population analyses.

4.3 Results

We produce disaggregated projections by age group, sex, and educational attainment for 21 countries, focusing on three key adult age groups: 25—44, 45—64, and 65 and older. For a detailed overview of data coverage, see Table 2 in the Appendix.

The projection results indicate a consistent improvement in educational attainment across Europe, with a steady decline in the share of the low-education group and corresponding growth in the mid- and high-education categories. These gains, however, evolve at different paces by sex, with educational expansion progressing more slowly among men and more rapidly among women. Despite these improvements, the overall composition of the population begins to stagnate after 2040, particularly among younger cohorts, as their absolute size declines due to sustained low fertility and accelerated population ageing. As a result, the relative growth of the elderly population becomes more pronounced, reshaping the demographic balance and tempering the impact of continued educational expansion among the youth.

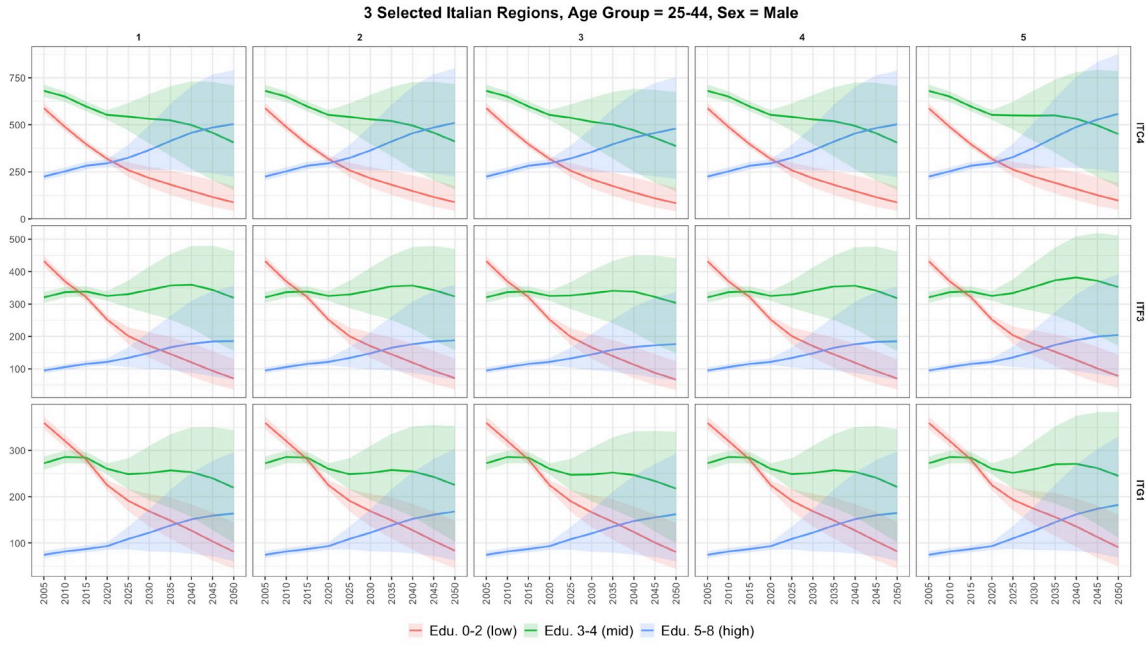

Figure 3: Educational attainment composition, in thousands of units, for three selected Italian regions in the 5 SSP scenarios. Age group: 25-44, sex: male.

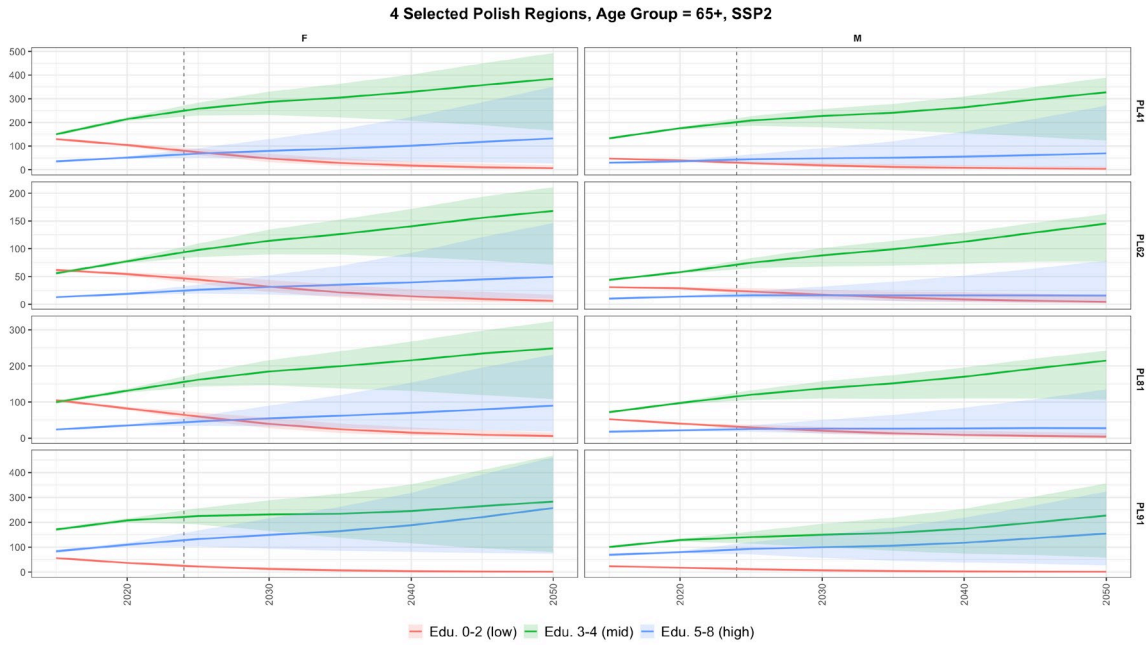

Figure 4: Educational attainment composition, in thousands of units, for scenario SSP-2 for four selected Polish regions according to the sex. Age group: 65+

5 References

Achebak, Hicham, Daniel Devolder, and Joan Ballester. 2019. 'Trends in Temperature-Related Age-Specific and Sex-Specific Mortality from Cardiovascular Diseases in Spain: A National Time-Series Analysis'. The Lancet Planetary Health 3 (7): e297–306. https://doi.org/10.1016/S2542-5196(19)30090-7.

Carvalho, Adenilson R., Fernando M. Ramos, and Antonio A. Chaves. 2011. 'Metaheuristics for the Feedforward Artificial Neural Network (ANN) Architecture Optimization Problem'. Neural Computing and Applications 20 (8): 1273–84. https://doi.org/10.1007/s00521-010-0504-3.

Chen, Kai, Evan De Schrijver, Sidharth Sivaraj, et al. 2024. 'Impact of Population Aging on Future Temperature-Related Mortality at Different Global Warming Levels'. Nature Communications 15 (1): 1796. https://doi.org/10.1038/s41467-024-45901-z.

De Rigo, D., A. Castelletti, A.E. Rizzoli, R. Soncini-Sessa, and E. Weber. 2005. 'A SELECTIVE IMPROVEMENT TECHNIQUE FOR FASTENING NEURO-DYNAMIC PROGRAMMING IN WATER RESOURCE NETWORK MANAGEMENT'. IFAC Proceedings Volumes 38 (1): 7–12. https://doi.org/10.3182/20050703-6-CZ-1902.02172.

Gao, Jing. 2017. Downscaling Global Spatial Population Projections from 1/8-Degree to 1-Km Grid Cells. With University Corporation For Atmospheric Research (UCAR):National Center For Atmospheric Research (NCAR):NCAR Library (NCARLIB). UCAR/NCAR. Application/pdf, 2082 KB. https://doi.org/10.5065/D60Z721H.

Gao, Jing. 2020. 'Global 1-Km Downscaled Population Grids, SSP-Consistent Projections and Base Year, v1.01 (2000 - 2100)'. Version 1.0. With Jing Gao. Harvard Dataverse.. https://doi.org/10.7910/DVN/TLJ99B.

Gao, Jing, and Martino Pesaresi. 2021. 'Downscaling SSP-Consistent Global Spatial Urban Land Projections from 1/8-Degree to 1-Km Resolution 2000–2100'. *Scientific Data* 8 (1): 281. https://doi.org/10.1038/s41597-021-01052-0.

Kullback, S., and R. A. Leibler. 1951. 'On Information and Sufficiency'. The Annals of Mathematical Statistics 22 (1): 79–86. https://doi.org/10.1214/aoms/1177729694.

Loichinger, Elke. 2015. 'Labor Force Projections up to 2053 for 26 EU Countries, by Age, Sex, and Highest Level of Educational Attainment'. Demographic Research 32 (15): 443–86. https://doi.org/10.4054/DemRes.2015.32.15.

Lutz, Wolfgang, Jesus Crespo Cuaresma, and Warren Sanderson. 2008. 'The Demography of Educational Attainment and Economic Growth'. Science 319 (5866): 1047–48. https://doi.org/10.1126/science.1151753.

Otto, Ilona M., Diana Reckien, Christopher P. O. Reyer, et al. 2017. 'Social Vulnerability to Climate Change: A Review of Concepts and Evidence'. Regional Environmental Change 17 (6): 1651–62. https://doi.org/10.1007/s10113-017-1105-9.

Pinho-Gomes, A, A Mcintosh, and M Woodward. 2024. 'Sex Differences in Mortality Associated with Heatwaves: A Systematic Review and Meta-Analysis'. European Journal of Public Health 34 (Supplement_3): ckae144.283. https://doi.org/10.1093/eurpub/ckae144.283.

Reid, Colleen E., Marie S. O'Neill, Carina J. Gronlund, et al. 2009. 'Mapping Community Determinants of Heat Vulnerability'. Environmental Health Perspectives 117 (11): 1730–36. https://doi.org/10.1289/ehp.0900683.

Sestito, B, L Reimann, M Mazzoleni, W J W Botzen, and J C J H Aerts. 2025. 'Identifying Vulnerability Factors Associated with Heatwave Mortality: A Spatial Statistical Analysis across Europe'. Environmental Research Letters 20 (4): 044025. https://doi.org/10.1088/1748-9326/adbcc8.

Striessnig, Erich, Jing Gao, Brian C O'Neill, and Leiwen Jiang. 2019. 'Empirically Based Spatial Projections of US Population Age Structure Consistent with the Shared Socioeconomic Pathways'. Environmental Research Letters 14 (11): 114038. https://doi.org/10.1088/1748-9326/ab4a3a.

Vésier, Chloé, and Aleš Urban. 2023. 'Gender Inequalities in Heat-Related Mortality in the Czech Republic'. International Journal of Biometeorology 67 (8): 1373–85. https://doi.org/10.1007/s00484-023-02507-2.

Yang, Jun, Maigeng Zhou, Zhoupeng Ren, et al. 2021. 'Projecting Heat-Related Excess Mortality under Climate Change Scenarios in China'. Nature Communications 12 (1): 1039. https://doi.org/10.1038/s41467-021-21305-1.

Yao, Xin, Ying Qu, Ashok K. Mishra, et al. 2025. 'Elderly Vulnerability to Temperature-Related Mortality Risks in China'. Science Advances 11 (6): eado5499. https://doi.org/10.1126/sciadv.ado5499.

6 Appendix

Table 1: Age-sex dimension geographical scope.

Country	Number of Regions	NUTS-2 Regions
AL	3	AL01, AL02, AL03
AT	9	AT11, AT12, AT13, AT21, AT22, AT31, AT32, AT33, AT34
BE	11	BE10, BE21, BE22, BE23, BE24, BE25, BE31, BE32, BE33, BE34, BE35
BG	6	BG31, BG32, BG33, BG34, BG41, BG42
CH	7	CH01, CH02, CH03, CH04, CH05, CH06, CH07
CY	1	CY00
CZ	8	CZ01, CZ02, CZ03, CZ04, CZ05, CZ06, CZ07, CZ08
DE	38	DE11, DE12, DE13, DE14, DE21, DE22, DE23, DE24, DE25, DE26, DE27, DE30, DE40, DE50, DE60, DE71, DE72, DE73, DE80, DE91, DE92, DE93, DE94, DEA1, DEA2, DEA3, DEA4, DEA5, DEB1, DEB2, DEB3, DEC0, DED2, DED4, DED5, DEE0, DEF0, DEG0
DK	5	DK01, DK02, DK03, DK04, DK05
EE	1	EE00
EL	13	EL30, EL41, EL42, EL43, EL51, EL52, EL53, EL54, EL61, EL62, EL63, EL64, EL65
ES	19	ES11, ES12, ES13, ES21, ES22, ES23, ES24, ES30, ES41, ES42, ES43, ES51, ES52, ES53, ES61, ES62, ES63, ES64, ES70
FI	5	FI19, FI1B, FI1C, FI1D, FI20
FR	27	FR10, FRB0, FRC1, FRC2, FRD1, FRD2, FRE1, FRE2, FRF1, FRF2, FRF3, FRG0, FRH0, FRI1, FRI2, FRI3, FRJ1, FRJ2, FRK1, FRK2, FRL0, FRM0, FRY1, FRY2, FRY3, FRY4, FRY5
HR	2	HR03, HR04
HU	8	HU11, HU12, HU21, HU22, HU23, HU31, HU32, HU33
IE	3	IE04, IE05, IE06
IS	1	IS00
IT	21	ITC1, ITC2, ITC3, ITC4, ITF1, ITF2, ITF3, ITF4, ITF5, ITF6, ITG1, ITG2, ITH1, ITH2, ITH3, ITH4, ITH5, ITI1, ITI2, ITI3, ITI4
LT	2	LT01, LT02
LU	1	LU00
LV	1	LV00
ME	1	ME00
MK	1	MK00
MT	1	MT00
NL	12	NL11, NL12, NL13, NL21, NL22, NL23, NL31, NL32, NL33, NL34, NL41, NL42
NO	7	NO01, NO02, NO03, NO04, NO05, NO06, NO07
PL	17	PL21, PL22, PL41, PL42, PL43, PL51, PL52, PL61, PL62, PL63, PL71, PL72, PL81, PL82, PL84, PL91, PL92

PT	7	PT11, PT15, PT16, PT17, PT18, PT20, PT30
RO	8	RO11, RO12, RO21, RO22, RO31, RO32, RO41, RO42
RS	4	RS11, RS12, RS21, RS22
SE	8	SE11, SE12, SE21, SE22, SE23, SE31, SE32, SE33
SI	2	SI03, SI04
SK	4	SK01, SK02, SK03, SK04
UK	41	UKC1, UKC2, UKD1, UKD3, UKD4, UKD6, UKD7, UKE1, UKE2, UKE3, UKE4, UKF1, UKF2, UKF3, UKG1, UKG2, UKG3, UKH1, UKH2, UKH3, UKI3, UKI4, UKI5, UKI6, UKI7, UKJ1, UKJ2, UKJ3, UKJ4, UKK1, UKK2, UKK3, UKK4, UKL1, UKL2, UKM5, UKM6, UKM7, UKM8, UKM9, UKN0

Table 2: Age-sex-education dimension geographical scope

Country	N_Regions	Region_Codes
AT	9	AT11, AT12, AT13, AT21, AT22, AT31, AT32, AT33, AT34
BE	11	BE10, BE21, BE22, BE23, BE24, BE25, BE31, BE32, BE33, BE34, BE35
BG	6	BG31, BG32, BG33, BG34, BG41, BG42
CZ	8	CZ01, CZ02, CZ03, CZ04, CZ05, CZ06, CZ07, CZ08
DE	38	DE11, DE12, DE13, DE14, DE21, DE22, DE23, DE24, DE25, DE26, DE27, DE30, DE40, DE50, DE60, DE71, DE72, DE73, DE80, DE91, DE92, DE93, DE94, DEA1, DEA2, DEA3, DEA4, DEA5, DEB1, DEB2, DEB3, DEC0, DED2, DED4, DED5, DEE0, DEF0, DEG0
DK	5	DK01, DK02, DK03, DK04, DK05
ES	19	ES11, ES12, ES13, ES21, ES22, ES23, ES24, ES30, ES41, ES42, ES43, ES51, ES52, ES53, ES61, ES62, ES63, ES64, ES70
FI	4	FI19, FI1B, FI1C, FI1D
FR	22	FR10, FRB0, FRC1, FRC2, FRD1, FRD2, FRE1, FRE2, FRF1, FRF2, FRF3, FRG0, FRH0, FRI1, FRI2, FRI3, FRJ1, FRJ2, FRK1, FRK2, FRL0, FRM0
HR	2	HR03, HR04
HU	8	HU11, HU12, HU21, HU22, HU23, HU31, HU32, HU33
IE	3	IE04, IE05, IE06
IT	21	ITC1, ITC2, ITC3, ITC4, ITF1, ITF2, ITF3, ITF4, ITF5, ITF6, ITG1, ITG2, ITH1, ITH2, ITH3, ITH4, ITH5, ITI1, ITI2, ITI3, ITI4
LT	2	LT01, LT02
NL	12	NL11, NL12, NL13, NL21, NL22, NL23, NL31, NL32, NL33, NL34, NL41, NL42
PL	17	PL21, PL22, PL41, PL42, PL43, PL51, PL52, PL61, PL62, PL63, PL71, PL72, PL81, PL82, PL84, PL91, PL92
PT	7	PT11, PT15, PT16, PT17, PT18, PT20, PT30
RO	8	RO11, RO12, RO21, RO22, RO31, RO32, RO41, RO42
SE	8	SE11, SE12, SE21, SE22, SE23, SE31, SE32, SE33
SI	2	SI03, SI04
SK	4	SK01, SK02, SK03, SK04

Table 3: Covariates list for the age disaggregation model

Variable Code	Variable Meaning
g.pop.NUTS2	Growth of the population in NUTS-2 region
lag1.tot_pop_NUTS2	5-years lag tot. pop. In NUTS-2 region
lag2.tot_pop_NUTS2	10-years lag tot. pop. In NUTS-2 region
lag1ratio.tot_pop_NUTS2	Ratio of 5- and 10-years lag tot. pop. in NUTS-2 region
lag1_under15	5-years lag under 15 pop
lag1_15_24	5-years lag under 15-24 pop
lag1_25_44	5-years lag under 25-44 pop
lag1_45_64	5-years lag under 45-64 pop
lag1_65	5-years lag under 65+ pop
lag2_under15	10-years lag under 15 pop
lag2_15_24	10-years lag 15-24 pop
lag2_25_44	10-years lag 25-44 pop
lag2_45_64	10-years lag 45-64 pop
lag2_65	10-years lag 65+ pop
lag1_g_under15	5-years lag growth under 15 pop
lag1_g_15_24	5-years lag growth 15-24 pop
lag1_g_25_44	5-years lag growth 25-44 pop
lag1_g_45_64	5-years lag growth 45-64 pop
lag1_g_65	5-years lag growth 65+ pop
lag1_quot_under15	Ratio between regional and national proportion, age under 15, 5-years lag.
lag1_quot_15_24	Ratio between regional and national proportion, age 15-24, 5-years lag.
lag1_quot_25_44	Ratio between regional and national proportion, age 25-44, 5-years lag.
lag1_quot_45_64	Ratio between regional and national proportion, age 45-64, 5-years lag.
lag1_quot_65	Ratio between regional and national proportion, age 65+, 5-years lag.
lag2_quot_under15	Ratio between regional and national proportion, age under 15, 10-years lag.
lag2_quot_15_24	Ratio between regional and national proportion, age 15-24, 10-years lag.
lag2_quot_25_44	Ratio between regional and national proportion, age 25-44, 10-years lag.
lag2_quot_45_64	Ratio between regional and national proportion, age 45-64, 10-years lag.
lag2_quot_65	Ratio between regional and national proportion, age 65+, 10-years lag.
ctry_prop_under15	Proportion at the country level
ctry_prop_15_24	Proportion at the country level
ctry_prop_25_44	Proportion at the country level
ctry_prop_45_64	Proportion at the country level
ctry_prop_65	Proportion at the country level
lag1_ctry_prop_under15	Proportion at the country level, 5-years lag
lag1_ctry_prop_15_24	Proportion at the country level, 5-years lag
lag1_ctry_prop_25_44	Proportion at the country level, 5-years lag
lag1_ctry_prop_45_64	Proportion at the country level, 5-years lag
lag1_ctry_prop_65	Proportion at the country level, 5-years lag
lag2_ctry_prop_under15	Proportion at the country level, 10-years lag
lag2_ctry_prop_15_24	Proportion at the country level, 10-years lag
lag2_ctry_prop_25_44	Proportion at the country level, 10-years lag
lag2_ctry_prop_45_64	Proportion at the country level, 10-years lag
lag2_ctry_prop_65	Proportion at the country level, 10-years lag
lag1_ctry_pop_g_under15	Proportional growth of the country population, age under 15, 5 years lag.
lag1_ctry_pop_g_15_24	Proportional growth of the country population, 15-24, 5 years lag.
lag1_ctry_pop_g_25_44	Proportional growth of the country population, 25-44, 5 years lag.
lag1_ctry_pop_g_45_64	Proportional growth of the country population, 45-64 5 years lag.
lag1_ctry_pop_g_65	Proportional growth of the country population, 65+, 5 years lag.
cluster_1	Belongs to cluster 1 (dichotomous variable)

cluster_2	Belongs to cluster 2 (dichotomous variable)
cluster_3	Belongs to cluster 3 (dichotomous variable)
urb_frac	Percentage of urbanised surface in the NUTS-2 region.