Supplementary Information: Subnational disaggregation of SSP-coherent population projections for Europe

Andrea Tamburini[1], Claudio Bosco[2], Erich Striessnig[3]

1. International Institute for Applied Systems Analysis, Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/OeAW, University of Vienna), Laxenburg, Austria.
2. European Commission - Joint Research Centre, Ispra (VA), Italy.
3. University of Vienna, Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/OeAW, University of Vienna), Vienna, Austria.
Corresponding author(s): Andrea Tamburini (tamburini@iiasa.ac.at)

This document provides supplementary information to accompany the manuscript. Specifically, we present a table describing the contents of the replication datasets required to reproduce the analyses, which are available in a Zenodo repository (https://doi.org/10.5281/zenodo.20178934). The table comprises four columns: File, Subfolder, Column Names, and Comment. The first column reports the name of each data file; the second indicates the subfolder within the ReplicationData directory in which it is contained, thereby also conveying the overall folder structure; the third lists the column names present in each file, shortened in flagged cases where entries are evidently repetitive; and the fourth provides any relevant additional comments.

File	Subfolder	Columns names	Comment
cluster_df_hierarchical_PCA_wardD.RData	/data	geo, cluster	
data_for_clustering.RData	/data	sex, geo, interval, TOTAL, under15, 15_24, 25_44, 45_64, 65+, UNK	
demo_r_pjangroup__with2024.RData	/data	freq, unit, sex, age, geo, TIME_PERIOD, values	
demo_r_pjangroup_ageGroups_timeIntervals_dep1_lessAgeGroups_with2024.RData	/data/dep_variables	sex, geo, interval, TOTAL, under15, 15_24, 25_44, 45_64, 65+, UNK	
dep_vrbl.RData	/data/dep_variables	sex, geo, interval, age.group, nuts2_prop, ctry, ctry_prop, quot, diff, log.quot, diff_1, diff_diffs	
Gao_adjusted_baseYear_NUTS2_pop_sizes_SSPS.RData	/data/GAO_wcde_adjusted_2024	Gao_adjusted_baseYear_NUTS2_pop_sizes_SSPS: [geo, ssp1_pop.tot.NUTS_ID.Gao_adj_2010-2014, ssp1_pop.tot.NUTS_ID.Gao_adj_2015-2019, ssp1_pop.tot.NUTS_ID.Gao_adj_2020-2024, ssp1_pop.tot.NUTS_ID.Gao_adj_2025-2029 […]	for all periods and SSPs
NUTS2_projection.RData	/data/GAO_wcde_adjusted_2024	character	
urban_land_fraction_KOMMUNE_SSPs12345_period.RData	/data/GAO_wcde_adjusted_2024	geo, scenario, period, urb_frac	
wcde_raw_ageGrouped_adjusted_baseYear_Estat.RData	/data/GAO_wcde_adjusted_2024	age.group, geo, scenario, 2010-2014, 2015-2019, 2020-2024, 2025-2029, [...]	for all periods
train_2000.2004_1995.1999_ctry.RData	/data/train&test	geo, age.group, g.pop.NUTS2, lag1.tot_pop_NUTS2, lag2.tot_pop_NUTS2, lag1ratio.tot_pop_NUTS2, lag1_under15, lag1_15_24, lag1_25_44, lag1_45_64, lag1_65+, lag2_under15, lag2_15_24, lag2_25_44, lag2_45_64, lag2_65+, lag1_g_under15, lag1_g_15_24, lag1_g_25_44, lag1_g_45_64, lag1_g_65+, sex, interval, diff, log.quot, diff_1, diff_diffs, quot_under15, quot_15_24, quot_25_44, quot_45_64, quot_65+, lag1_quot_under15, lag1_quot_15_24, lag1_quot_25_44, lag1_quot_45_64, lag1_quot_65+, lag2_quot_under15,	

		lag2_quot_15_24, lag2_quot_25_44, lag2_quot_45_64, lag2_quot_65+, ctry_prop_under15, ctry_prop_15_24, ctry_prop_25_44, ctry_prop_45_64, ctry_prop_65+, lag1_ctry_prop_under15, lag1_ctry_prop_15_24, lag1_ctry_prop_25_44, lag1_ctry_prop_45_64, lag1_ctry_prop_65+, lag2_ctry_prop_under15, lag2_ctry_prop_15_24, lag2_ctry_prop_25_44, lag2_ctry_prop_45_64, lag2_ctry_prop_65+, lag1_ctry_pop_g_under15, lag1_ctry_pop_g_15_24, lag1_ctry_pop_g_25_44, lag1_ctry_pop_g_45_64, lag1_ctry_pop_g_65+, y	
train_2005.2009_2000.2004_ctry.RData	/data/train&test	Same as train_2000.2004_1995.1999_ctry.Rdata	
train_2010.2014_2005.2009_ctry.RData	/data/train&test	Same as train_2000.2004_1995.1999_ctry.Rdata	
train_2015.2019_2010.2014_ctry.RData	/data/train&test	Same as train_2000.2004_1995.1999_ctry.Rdata	
train_2020.2024_2015.2019_ctry.RData	/data/train&test	Same as train_2000.2004_1995.1999_ctry.Rdata	
UK_reconstructed20202024.RData	/data	TOTAL, geo, 15_24, 25_44, 45_64, 65+, under15, interval	
urban_land_fraction_NUTS2_SSP2_period.RData	/dataGao/cropped	geo, period, urb_frac	
urban_land_fraction_NUTS2_SSP2.RData	/dataGao/cropped	geo, urb_frac_2000, ssp2_2010, ssp2_2020, ssp2_2030, ssp2_2040, ssp2_2050, ssp2_2060, ssp2_2070, ssp2_2080, ssp2_2090, ssp2_2100	
cluster_df_hierarchical_PCA_wardD.RData	/data	geo, cluster	
data_for_clustering.RData	/data	sex, geo, interval, TOTAL, under15, 15_24, 25_44, 45_64, 65+, UNK	
demo_r_pjangroup__with2024.RData	/data	freq, unit, sex, age, geo, TIME_PERIOD, values	
demo_r_pjangroup_ageGroups_timeIntervals_dep1_lessAgeGroups_with2024.RData	/data/dep_variables	sex, geo, interval, TOTAL, under15, 15_24, 25_44, 45_64, 65+, UNK	
dep_vrbl.RData	/data/dep_variables	sex, geo, interval, age.group, nuts2_prop, ctry, ctry_prop, quot, diff, log.quot, diff_1, diff_diffs	
Gao_adjusted_baseYear_NUTS2_pop_sizes_SSPS.RData	/data/GAO_wcde_adjusted_2024	Gao_adjusted_baseYear_NUTS2_pop_sizes_SSPS: [geo, ssp1_pop.tot.NUTS_ID.Gao_adj_2010-2014, ssp1_pop.tot.NUTS_ID.Gao_adj_2015-2019, ssp1_pop.tot.NUTS_ID.Gao_adj_2020-2024, ssp1_pop.tot.NUTS_ID.Gao_adj_2025-2029 […]	for all periods and SSPs
NUTS2_projection.RData	/data/GAO_wcde_adjusted_2024	character	
urban_land_fraction_KOMMUNE_SSPs12345_period.RData	/data/GAO_wcde_adjusted_2024	geo, scenario, period, urb_frac	
wcde_raw_ageGrouped_adjusted_baseYear_Estat.RData	/data/GAO_wcde_adjusted_2024	age.group, geo, scenario, 2010-2014, 2015-2019, 2020-2024, 2025-2029, [...]	for all periods
train_2000.2004_1995.1999_ctry.RData	/data/train&test	geo, age.group, g.pop.NUTS2, lag1.tot_pop_NUTS2, lag2.tot_pop_NUTS2, lag1ratio.tot_pop_NUTS2, lag1_under15, lag1_15_24, lag1_25_44, lag1_45_64, lag1_65+, lag2_under15, lag2_15_24, lag2_25_44, lag2_45_64, lag2_65+, lag1_g_under15, lag1_g_15_24, lag1_g_25_44, lag1_g_45_64, lag1_g_65+, sex, interval, diff, log.quot, diff_1, diff_diffs, quot_under15, quot_15_24, quot_25_44, quot_45_64, quot_65+, lag1_quot_under15, lag1_quot_15_24, lag1_quot_25_44, lag1_quot_45_64, lag1_quot_65+, lag2_quot_under15, lag2_quot_15_24, lag2_quot_25_44,	

		lag2_quot_45_64, lag2_quot_65+, ctry_prop_under15, ctry_prop_15_24, ctry_prop_25_44, ctry_prop_45_64, ctry_prop_65+, lag1_ctry_prop_under15, lag1_ctry_prop_15_24, lag1_ctry_prop_25_44, lag1_ctry_prop_45_64, lag1_ctry_prop_65+, lag2_ctry_prop_under15, lag2_ctry_prop_15_24, lag2_ctry_prop_25_44, lag2_ctry_prop_45_64, lag2_ctry_prop_65+, lag1_ctry_pop_g_under15, lag1_ctry_pop_g_15_24, lag1_ctry_pop_g_25_44, lag1_ctry_pop_g_45_64, lag1_ctry_pop_g_65+, y	
train_2005.2009_2000.2004_ctry.RData	/data/train&test	Same as train_2000.2004_1995.1999_ctry.Rdata	
train_2010.2014_2005.2009_ctry.RData	/data/train&test	Same as train_2000.2004_1995.1999_ctry.Rdata	
train_2015.2019_2010.2014_ctry.RData	/data/train&test	Same as train_2000.2004_1995.1999_ctry.Rdata	
train_2020.2024_2015.2019_ctry.RData	/data/train&test	Same as train_2000.2004_1995.1999_ctry.Rdata	
UK_reconstructed20202024.RData	/data	TOTAL, geo, 15_24, 25_44, 45_64, 65+, under15, interval	
urban_land_fraction_NUTS2_SSP2_period.RData	/dataGao/cropped	geo, period, urb_frac	
urban_land_fraction_NUTS2_SSP2.RData	/dataGao/cropped	geo, urb_frac_2000, ssp2_2010, ssp2_2020, ssp2_2030, ssp2_2040, ssp2_2050, ssp2_2060, ssp2_2070, ssp2_2080, ssp2_2090, ssp2_2100	
cluster_df_hierarchical_PCA_wardD.RData	/data	geo, cluster	
data_for_clustering.RData	/data	sex, geo, interval, TOTAL, under15, 15_24, 25_44, 45_64, 65+, UNK	
demo_r_pjangroup__with2024.RData	/data	freq, unit, sex, age, geo, TIME_PERIOD, values	
demo_r_pjangroup_ageGroups_timeIntervals_dep1_lessAgeGroups_with2024.RData	/data/dep_variables	sex, geo, interval, TOTAL, under15, 15_24, 25_44, 45_64, 65+, UNK	
dep_vrbl.RData	/data/dep_variables	sex, geo, interval, age.group, nuts2_prop, ctry, ctry_prop, quot, diff, log.quot, diff_1, diff_diffs	
Gao_adjusted_baseYear_NUTS2_pop_sizes_SSPS.RData	/data/GAO_wcde_adjusted_2024	Gao_adjusted_baseYear_NUTS2_pop_sizes_SSPS: [geo, ssp1_pop.tot.NUTS_ID.Gao_adj_2010-2014, ssp1_pop.tot.NUTS_ID.Gao_adj_2015-2019, ssp1_pop.tot.NUTS_ID.Gao_adj_2020-2024, ssp1_pop.tot.NUTS_ID.Gao_adj_2025-2029 […]	for all periods and SSPs
NUTS2_projection.RData	/data/GAO_wcde_adjusted_2024	character	
urban_land_fraction_KOMMUNE_SSPs12345_period.RData	/data/GAO_wcde_adjusted_2024	geo, scenario, period, urb_frac	
wcde_raw_ageGrouped_adjusted_baseYear_Estat.RData	/data/GAO_wcde_adjusted_2024	age.group, geo, scenario, 2010-2014, 2015-2019, 2020-2024, 2025-2029, [...]	for all periods
train_2000.2004_1995.1999_ctry.RData	/data/train&test	geo, age.group, g.pop.NUTS2, lag1.tot_pop_NUTS2, lag2.tot_pop_NUTS2, lag1ratio.tot_pop_NUTS2, lag1_under15, lag1_15_24, lag1_25_44, lag1_45_64, lag1_65+, lag2_under15, lag2_15_24, lag2_25_44, lag2_45_64, lag2_65+, lag1_g_under15, lag1_g_15_24, lag1_g_25_44, lag1_g_45_64, lag1_g_65+, sex, interval, diff, log.quot, diff_1, diff_diffs, quot_under15, quot_15_24, quot_25_44, quot_45_64, quot_65+, lag1_quot_under15, lag1_quot_15_24, lag1_quot_25_44, lag1_quot_45_64, lag1_quot_65+, lag2_quot_under15, lag2_quot_15_24, lag2_quot_25_44, lag2_quot_45_64, lag2_quot_65+,	

		ctry_prop_under15, ctry_prop_15_24, ctry_prop_25_44, ctry_prop_45_64, ctry_prop_65+, lag1_ctry_prop_under15, lag1_ctry_prop_15_24, lag1_ctry_prop_25_44, lag1_ctry_prop_45_64, lag1_ctry_prop_65+, lag2_ctry_prop_under15, lag2_ctry_prop_15_24, lag2_ctry_prop_25_44, lag2_ctry_prop_45_64, lag2_ctry_prop_65+, lag1_ctry_pop_g_under15, lag1_ctry_pop_g_15_24, lag1_ctry_pop_g_25_44, lag1_ctry_pop_g_45_64, lag1_ctry_pop_g_65+, y	
train_2005.2009_2000.2004_ctry.RData	/data/train&test	Same as train_2000.2004_1995.1999_ctry.Rdata	
train_2010.2014_2005.2009_ctry.RData	/data/train&test	Same as train_2000.2004_1995.1999_ctry.Rdata	
train_2015.2019_2010.2014_ctry.RData	/data/train&test	Same as train_2000.2004_1995.1999_ctry.Rdata	
train_2020.2024_2015.2019_ctry.RData	/data/train&test	Same as train_2000.2004_1995.1999_ctry.Rdata	
UK_reconstructed20202024.RData	/data	TOTAL, geo, 15_24, 25_44, 45_64, 65+, under15, interval	
urban_land_fraction_NUTS2_SSP2_period.RData	/dataGao/cropped	geo, period, urb_frac	
urban_land_fraction_NUTS2_SSP2.RData	/dataGao/cropped	geo, urb_frac_2000, ssp2_2010, ssp2_2020, ssp2_2030, ssp2_2040, ssp2_2050, ssp2_2060, ssp2_2070, ssp2_2080, ssp2_2090, ssp2_2100	

Table S1: contents of the replication dataset. For each file included in the Zenodo repository (https://doi.org/10.5281/zenodo.20178934), the table reports the subfolder within the ReplicationData directory, the column names contained in the file, and relevant comments. Column names are shortened where entries are evidently repetitive. The subfolder structure reflects the organisation of the replication archive.