

ACCEPTED MANUSCRIPT • OPEN ACCESS

Empirically-based spatial projections of U.S. population age structure consistent with the shared socioeconomic pathways

To cite this article before publication: Erich Striessnig *et al* 2019 *Environ. Res. Lett.* in press <https://doi.org/10.1088/1748-9326/ab4a3a>

Manuscript version: Accepted Manuscript

Accepted Manuscript is “the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an ‘Accepted Manuscript’ watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors”

This Accepted Manuscript is © 2019 The Author(s). Published by IOP Publishing Ltd.

As the Version of Record of this article is going to be / has been published on a gold open access basis under a CC BY 3.0 licence, this Accepted Manuscript is available for reuse under a CC BY 3.0 licence immediately.

Everyone is permitted to use all or part of the original content in this article, provided that they adhere to all the terms of the licence <https://creativecommons.org/licenses/by/3.0>

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected and is not published on a gold open access basis under a CC BY licence, unless that is specifically stated in the figure caption in the Version of Record.

View the [article online](#) for updates and enhancements.

Empirically-based Spatial Projections of U.S. Population Age Structure Consistent with the Shared Socioeconomic Pathways

Erich Striessnig¹, Jing Gao², Brian C. O'Neill³, Leiwen Jiang^{4,5}

¹ Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU), Austria

² University of Delaware, USA

³ University of Denver, USA

⁴ Population Council, USA

⁵ Asian Demographic Research Institute (ADRI), China

Abstract

Spatially-explicit population projections by age are increasingly needed for understanding bilateral human-environment interactions. Conventional demographic methods for projecting age structure experience substantial challenges at small spatial scales. In search of a potentially better performing alternative, we develop an empirically-based spatial model of population age structure and test its application in projecting U.S. population age structure over the 21st century under various socioeconomic scenarios (SSPs). The model draws on 40 years of historical data explaining changes in spatial age distribution at the county level. It demonstrates that a very good model fit is achievable even with parsimonious data input, and distinguishes itself from existing methods as a promising approach to spatial age structure modelling at the global level where data availability is often limited. Results suggest that wide variations in the spatial pattern of county-level age structure are plausible, with the possibility of substantial aging clustered in particular parts of the country. Aging is experienced most prominently in thinly-populated counties in the Midwest and the Rocky Mountains, while cities and surrounding counties, particularly in California, as well as the southern parts of New England and the Mid-Atlantic region, maintain a younger population age structure with a lower proportion in the most vulnerable 70+ age group. The urban concentration of younger people, as well as the absolute number of vulnerable elderly people can vary strongly by SSP.

Keywords: Spatial population, age structure, population projections, Shared Socioeconomic Pathways

Introduction

The spatial pattern of a population's distribution is a key driver of its vulnerability to many social and environmental stresses. Existing literature has found its influence to be larger than or at least equal to the influence of physical environmental factors, such as heat waves and sea-level rise (Hallegatte et al. 2013; Neumann et al. 2015; Bryan Jones et al. 2015; Hauer, Evans, and Mishra 2016; Lehner, Deser, and Sanderson 2018). Age is another important demographic factor determining people's vulnerability to all sorts of hazards, however, only few studies have focused on it so far, in particular in conjunction with its distribution across space (Marsha et al. 2018). Large-scale (continental to global) long-term (e.g. over the entire 21st century) spatial projections of age-structured populations are scarce. Most such projections are local in scale and short-term in range (Swanson, Schlottmann, and Schmidt 2010; Salvo, Lobo, and Maurer 2013). Many future impact assessments, therefore, take place at the national level (Dong et al. 2015).

A key obstacle to the production of large-scale, long-term spatial projections of age structure is the difficulty of applying conventional demographic approaches to such spatial and temporal scales. The dominant method for population projections is the cohort-component approach (Burch 2018), deriving directly from the demographic equation balancing births, deaths, migration, and population growth, with each component differentiated by age. Popularized in the 1920s and 1930s (Whelpton 1936), the most widely used national-level population projections today all rely on this method (United Nations 2019; Lutz et al. 2019). Its application to smaller spatial scales is challenging, because it requires both sufficient data on and future projections of age-specific demographic rates for each spatial unit. In data-poor environments, such information on current conditions is typically not available, and the sheer number of assumptions that must be made for future projections for small regions is an obstacle, especially for migration which can vary greatly over space and time.

Nonetheless, county-level population projections that include age structure have been carried out for the U.S. based on the cohort-component method (Bierwagen et al. 2010; U.S. Environmental Protection Agency 2017). However, some input assumptions (fertility and mortality) were not varied at the county level, therefore leaving out some degrees of heterogeneity. To reduce data requirements, Hauer (2019) projected U.S. county-level population and age structure by extrapolating rates of change in each age group, and then scaling the size of all population subgroups up or down to match a target aggregate national projection. This method does not allow for spatial variation in outcomes between national scenarios, making it perhaps better suited to shorter time horizons. A simpler scaling method has been used for Europe, where Terama et al. (2017) downscaled national level age structure projections from a cohort component model to sub-national regions, proportionally to their age structures in the base year, a method that does not allow for changes over time in the relative age distribution across regions.

Seeking for a potentially better performing alternative, we explored empirically based model development methods. We conducted thorough and systematic exploratory data analysis of best

1
2
3 available datasets with global coverage. Summary results for U.S. counties presented in SI section
4 1 revealed the following spatial and temporal patterns:
5

- 6 1. In any historical census, U.S. counties can be distinguished by roughly four distinctive age
7 profiles. Similar age profiles appear spatially clustered during earlier census years (1940-
8 1960) but evolve to show dispersed spatial patterns in recent censuses (1970-2010).
9
- 10 2. Over time, most counties switch among different age profiles for various reasons, e.g. aging
11 of the baby boomers or internal migration. Meanwhile, counties with extreme age profiles
12 (e.g. those that are home to large college-aged populations or retirement communities) tend
13 to be more temporally static.
14
- 15 3. The relationships between changes in county-level age structure and its potential predictors
16 are complex and non-linear.
17
18

19 We developed a new model using regression trees to project changes in county age structure based
20 on current and past county demographic characteristics, reflecting these empirical patterns.
21 Regression trees can model multiple non-linear relationships simultaneously. During training, the
22 model empirically identifies the best way to model counties with different age profiles and other
23 demographic characteristics as separate cases using different “tree branches”. When making
24 projections, at each future decadal time step the model projects county age structure change using
25 whichever branch of the regression tree corresponds to its demographic characteristics at that point
26 in time. This is analogous to the observed pattern that counties switch among different age profiles
27 over time. In addition, because regression trees treat different parts of the data range separately,
28 they are robust to extreme values, which may occur more frequently under certain future scenarios.
29 Altogether, the characteristics inherent to regression trees make them a particularly suitable
30 method for modeling the highly temporally variant phenomenon of age structure change.
31
32

33 We used this model for adding five age groups to existing projections of the spatial distribution of
34 population that contain age structure information only at the national aggregate level. The five age
35 groups satisfy most needs for age structure information by providing outcomes for age groups such
36 as the elderly, children, and the working age population (see Methodology). The existing national
37 and spatial population projections were developed for the Shared Socioeconomic Pathways
38 (SSPs), a set of five alternative scenarios of future societal development. The SSPs have been
39 widely used in climate and other global change research (O’Neill et al. 2017; Striessnig and
40 Loichinger 2016; Jiang and O’Neill 2017; Riahi et al. 2017). They include national-level
41 projections of population by age and sex (KC and Lutz 2017), and gridded spatial projections of
42 population counts (without age detail) at 1/8-degree and 1-km resolutions (Bryan Jones and
43 O’Neill 2016; Gao 2017). Differences between SSPs are described in detail in the Methodology,
44 as well as SI section 2.
45
46

47 In spite of its parsimonious requirements for input variables, the model achieved very good fit
48 when validated over historical data (see SI section 3), a prime prerequisite for its application to
49 future projections in data-poor contexts where the cohort-component approach is not an option.
50
51
52
53
54
55
56
57
58
59
60

1
2
3 The model's primary covariates capture the counties' demographic histories inscribed in their own
4 age structures, with some metrics accounting for potential spatial autocorrelation among
5 neighboring counties. The model's estimates are stable over long time horizons even without
6 constraints at the national level (SI section 4). Moreover, unrealistically high change rates
7 (implying unreasonable birth, death, or migration rates) cannot occur using the methodology
8 presented here, as the range of possible outcomes of our regression trees are by definition within
9 the range observed in the past.
10
11
12

13 Hence, our empirically based approach is a promising candidate for a well-performing global
14 model under data-sparse conditions. Though the work presented here is for the U.S. only, we are
15 currently exploring the usability of our approach for other countries, and plan to integrate the
16 lessons learnt to develop a global model eventually. The U.S. is a good first test case because here
17 different SSPs lead to very different future trajectories of population growth, population aging,
18 and spatial distributions of population (SI section 2).
19
20
21
22
23

24 Methodology

25
26 **Exploratory Analysis:** Before setting up our model, we used a range of exploratory data analysis
27 and visualization techniques to understand the patterns in the U.S. decadal census data since 1940,
28 which is the first U.S. census available with sufficient age and spatial detail (Manson et al. 2018).
29 Most notably, we applied hierarchical cluster analysis to county-level age profiles from every
30 census, respectively. Our findings suggested that while in the 1940s and 50s, age structure across
31 the U.S. was characterized by marked spatial autocorrelation at the county level, the pattern
32 observed since the 1970s followed a different regime in which spatial autocorrelation of age
33 structure is far less pronounced (SI section 1). This is consistent with existing literature confirming
34 the profound sociodemographic changes that the U.S. underwent up until the 1970s that
35 substantially altered the spatial patterns of age structure (Jackson 1987): With the onset of the baby
36 boom, average family sizes increased markedly, leading to an increased demand for larger homes
37 outside America's urban cores. The mobility revolution facilitated longer commutes and the move
38 to the suburbs, while the gradual westward shift of the population was still ongoing.
39
40
41
42

43 Considering these patterns, we trained our empirical model on the census data from the more recent
44 decades only (1980-2010) during which the observed spatial patterns in age structure are consistent
45 and to keep the 1970 census wave for validation purposes. For future projections, it is unlikely that
46 the data generating process will return to the pre-1970s regime, given that it has already been stable
47 for more than four decades. Note, however, that although spatial autocorrelation of age structure
48 has declined, variation in age structure across counties remains high (see Figure S1), driven mainly
49 by persistent urban-rural differences (for details see Figure S3 of the SI Appendix).
50
51
52

53 **Model Design:** To avoid small numbers in thinly populated rural counties and to obtain more
54 robust modeling results over long time horizons, we use five aggregate age groups (0-19, 20-29,
55
56
57
58
59
60

30-54, 55-69, 70+). This choice of age ranges reflects earlier findings from studies of age-specific mobility patterns observed in the U.S. in the second half of the 20th century (Johnson et al. 2005; Johnson and Winkler 2015) while trying to account for differences in household consumption behavior over the lifecycle (Ando and Modigliani 1963; Mason and Lee 2013; Lee and Mason 2011), as well as age-specific vulnerability patterns (Jonkman et al. 2009; Zagheni, Muttarak, and Striessnig 2016). Modeling each age group separately helps keeping the structure of our trees simple and interpretable. The regression trees are fitted using the *rpart* package in *R Statistical Software* (Therneau et al. 2015). The process of pruning the trees, as well as the detailed structures of all five trees are described in SI section 2.

Our dependent variable, is the change in the relative shares (Δrs), i.e. the difference in the population share of each age group in the total county population over a decade (i.e., between two census waves) relative to the same difference observed at the national level:

$$\Delta rs = \frac{share_{t, county} - share_{t-1, county}}{share_{t, national} - share_{t-1, national}} \quad (1)$$

Thus, we are modelling county-level deviations in the change in the share over time from the same change at the national level. This approach allows us to automatically account for the national level change that is occurring in the aggregate scenario for which we are producing a spatial age structure outcome, and to focus the model on the differences across counties relative to the national trend. We tested various alternative specifications of the dependent variable, including the raw age-specific shares of population or differences in those shares compared to the national shares. The best fit to the observed shares was achieved by the change in the relative share variable in equation (1).

In choosing independent variables, care was taken to use variables that would be available in the context of projections (see full list in Table S3). Lacking information on the drivers of population change at the county level, temporally lagged shares of all five age groups in the total county population up to three census waves in the past allow us to study the influence of historic demographic shifts on future developments. The use of lagged shares of all age groups to project the future changes in a single age group provides a cohort perspective that mimics the behavior of a cohort-component model. If, for example, a baby boom produces an especially large share of population in the 0-19 age group in one time period, older age groups should expect to grow in future time periods as the baby boomers age. The lagged structure of the full set of age groups allows for this dynamic to be represented in the model. To rule out the possibility of linear dependence among our predictors, we conducted a thorough collinearity investigation (see SI section 3.1 for details).

1
2
3 **Training and Validation:** The trees use lagged information on age structure from the previous
4 three census waves to project changes over a given decade. We trained the model by closely
5 reproducing observed change over the 2000-2010 decade using patterns and trends observed in the
6 census waves of 1980, 1990 and 2000. While training the model on one decade of change may
7 seem limiting for supporting century-scale projections, additional confidence in the applicability
8 of the model comes from (1) the use of three preceding decades of data in estimating the model,
9 (2) a validation test run on a different decade, and (3) the fact that the 2000-2010 decade is not
10 unusual with respect to trends since at least 1970 (see Figure S1).
11
12
13

14 As shown in section 3.2 of the SI, the goodness of fit for all five regression trees is confirmed by
15 values of R^2 close to and above 0.90. To evaluate the model's generalizability, we applied it to
16 estimate changes over the 1990-2000 period, using data from the census waves of 1970, 1980, and
17 1990. The performance continues to be high with values of R^2 well above 0.80 for all five trees.
18 In addition, no spatial pattern could be identified in the residuals, supporting the quality of the
19 model for short-term modeling.
20
21
22

23 Longer-term performance of the model would be supported by successfully capturing demographic
24 processes related to fertility, mortality or migration. To examine whether this is the case, we tested
25 the trees' sensitivity to county level migration and crude birth rates, even though information on
26 those would not actually be available for use in future projections. Neither variable added to the
27 trees' explanatory power, as the trees did not pick them and were able to find appropriate surrogate
28 variables, implying that the informational content of these omitted variables is already captured by
29 lagged age structures and neighborhood characteristics, and that our projection model is reasonable
30 without them.
31
32
33

34 **National Population Projections:** The main input to our future projections are the SSPs. At the
35 national aggregate level, population aging is the dominant trend in all five of them, but the speed
36 of aging varies between them (Figure S5) and so do the drivers of population change. While under
37 SSP5 ("Fossil-fueled development") rich OECD countries, including the U.S., are assumed to
38 experience high fertility and migration combined with low mortality, which limits the rate of aging,
39 the opposite is true under SSP3 ("Regional rivalry"), leading to faster aging of the U.S. population.
40 Under SSP2 ("Middle of the road"), all three of these drivers are assumed to be at medium levels,
41 yet the resulting national share of elderly population in 2100 is relatively large, almost the same
42 as in SSP3, albeit for very different reasons. While under SSP3 population aging is driven by low
43 fertility and low migration, under SSP2 it is the higher net in-migration in the first half of the
44 century that leads to increased aging once migration starts to decline in the second half of the
45 century, as is the assumption with all SSPs. Despite all SSPs sharing the same assumptions for the
46 U.S. with regard to urbanization (Jiang and O'Neill 2017), they lead to substantially different
47 spatial population outcomes at the sub-national scale (Bryan Jones and O'Neill 2016).
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Our statistical model is capable of distinguishing between those different national level pathways,
4 providing different sub-national distributions of aging depending on which demographic forces
5 drive them.
6

7
8 **Projection:** Given the SSPs as input, we then used the five regression trees to project county-level
9 population shares of the five age groups independently from each other in ten-year steps covering
10 the 2010-2100 period. At each ten-year step, the regression trees determine which branches are
11 used to make projections for any given county. Depending on the evolution of each county's age
12 related characteristics, different tree branches may be used for the same county at different times.
13 To ensure that (1) the five age-group shares of each county sum to one, and (2) the adjusted
14 outcomes are consistent with the national aggregate age structure given by the SSPs, we
15 renormalized the counties' age-group shares estimated by the trees using iterative proportional
16 fitting (IPF). Running this renormalization at every time step avoids potentially larger errors
17 caused by the accumulation of small mismatches over time. It is worth noting that the regression
18 trees performed highly, and the sum of their unconstrained estimates within each county is already
19 close to one. As shown in SI section 4, IPF leads only to minor changes to the age structures
20 predicted for individual counties.
21
22
23
24

25
26 The outcomes of our county age structure model depend on the existing national and spatial
27 population models in two main ways. First, national population size acts as a constraint on the
28 spatial population model, and that model produces population size and density by county that
29 enters the county age structure model as independent variables. Second, national age structure is
30 used as a constraint on the county age structure model as part of the IPF. Therefore, national-level
31 population outcomes, such as high growth rates and a young age structure from a high fertility
32 scenario, and spatial population outcomes, such as concentrated growth in urban areas, will both
33 influence the results from the county age structure model. These factors interact with the current
34 and lagged demographic conditions across counties to produce future outcomes. What would not
35 be reflected in the model are changes in spatial patterns of demographic change relative to
36 historical patterns. For example, if domestic or international migration patterns strongly shift their
37 spatial pattern, so that new cities and states become preferred destinations (or origins in the case
38 of domestic migration), this will not be reflected in the spatial population model and therefore also
39 not in the county age structure model. If the age structure of rural-urban migration changes
40 substantially relative to historical trends, this will change the nature of rural-urban differences in
41 age structure away from historical patterns and not be reflected in the model.
42
43
44
45
46
47
48
49

50 **Results**

51 We applied the regression tree methodology described above to project county-level age structure
52 for the conterminous U.S. consistent with all five of the SSPs. We show results for the Middle of
53 the Road scenario (SSP2) as well as for SSPs 3 and 5, which bound future assumptions on the
54
55
56
57
58
59
60

drivers of population change in the U.S.; results for the remaining SSPs are reported in the SI Appendix section 5.

Figure 1 shows observed (1940-2010) and projected (2020-2100) proportions of total county population in five different age groups. In accordance with the general trend of population aging predicted for the U.S. 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, an age group particularly vulnerable to the consequences of climate change, such as increased frequency and intensity of heat waves. While at the national level this proportion is predicted to increase from around 9 % in 2010 to between 26 % (SSP5) and 34 % (SSP1) in 2100, results show that in 25 % of counties, that proportion can rise to between 33 % and 40 % by the end of the century. In places like Charlotte County, FL, where the proportion 70+ is already high in 2010 (24 %), it is projected to reach between 40-50 % in 2100. This would be twice as high as observed in any single county in 2010 (27.6 % in Sumter County, FL). The spatial distribution of these sub-national differences in aging are explored below.

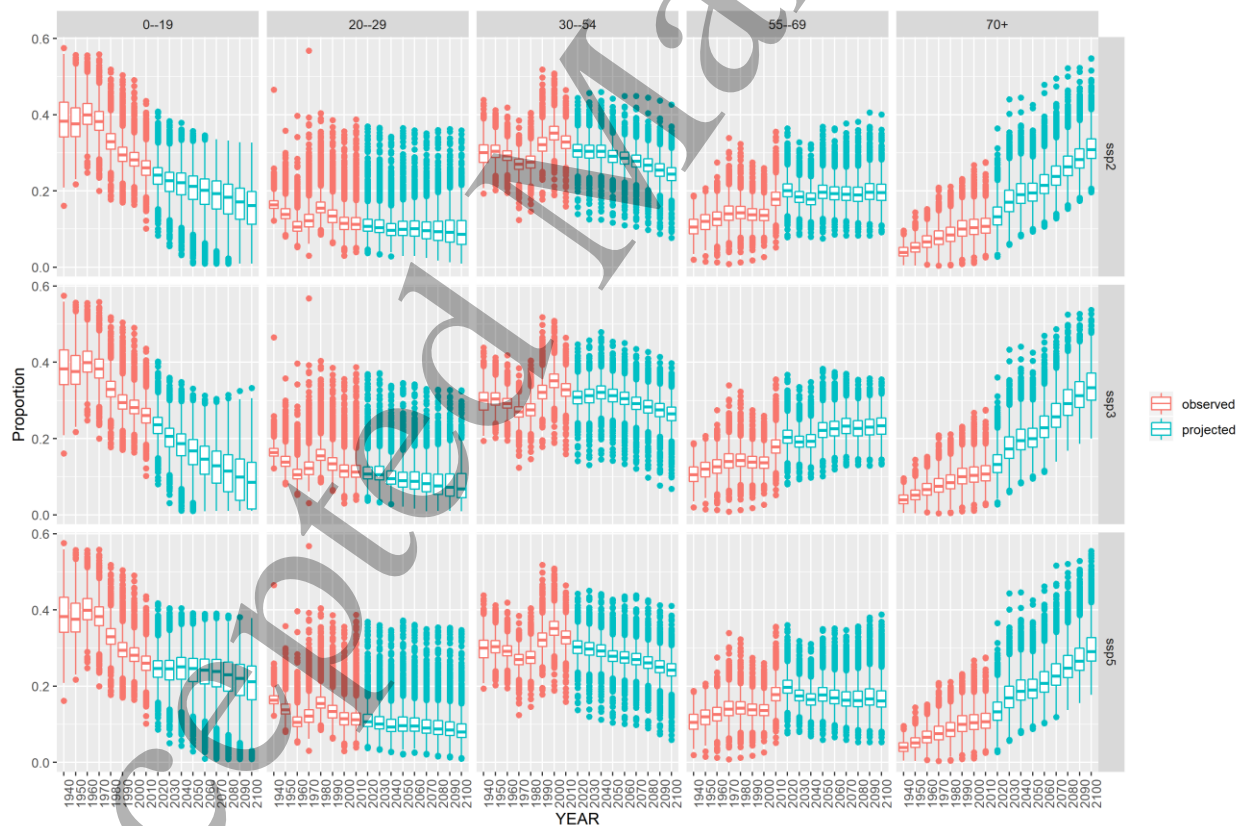


Figure 1. Observed (1940-2010) and projected (2020-2100) proportions of population at the U.S. county level in five different age groups under SSP2 (top), SSP3 (middle), and SSP5 (bottom). Boxes around the median indicate inter-quartile range. Outliers which are more than 1.5 times inter-quartile range from the median are indicated individually as dots.

1
2
3 We also find a large increase in the number of counties with very little population under the age
4 of 20. While currently there are no counties with less than 5% of their population under age 20, in
5 SSP2 and SSP5 this figure grows to 10 % of all counties, and in SSP3 it reaches 35 % of all
6 counties by the end of the century. This result is driven by (1) fertility decline at the national level,
7 most strongly in the population decline future described by SSP3 where the total fertility rate
8 (TFR) for the U.S. in 2095-2100 is assumed to reach 1.47 (compared to 1.87 today; CIA 2017),
9 corresponding to a decline in the proportion aged 0-19 at the national level from 27 % in 2010 to
10 15.3 % in 2100; (2) negative population momentum, as the proportion of potential parents is also
11 going down steadily; (3) the increase in the proportion of elderly population above the age of 70;
12 and (4) a spatial model that keeps the variance in proportions of total national population across
13 counties approximately constant over time (for details see Bryan Jones and O'Neill 2016). Only
14 under SSP5 is national level TFR expected to rise to above replacement level (2.29) by the end of
15 the century, limiting the number of counties with very small proportions of children.
16
17
18
19
20

21 The substantial aging of the population even in the Middle of the Road scenario is reflected
22 spatially throughout the entire conterminous U.S. (Figure 2). Yet counties with large cities, as well
23 as their high-density neighboring counties, which attract large amounts of working age population
24 and their children, maintain a relatively high proportion of youth and conversely, a lower
25 proportion of elderly people. This is most visible in California, the East Coast and in the Chicago
26 area. The thinly populated counties in the Midwest and the Rocky Mountains experience aging
27 most drastically.
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

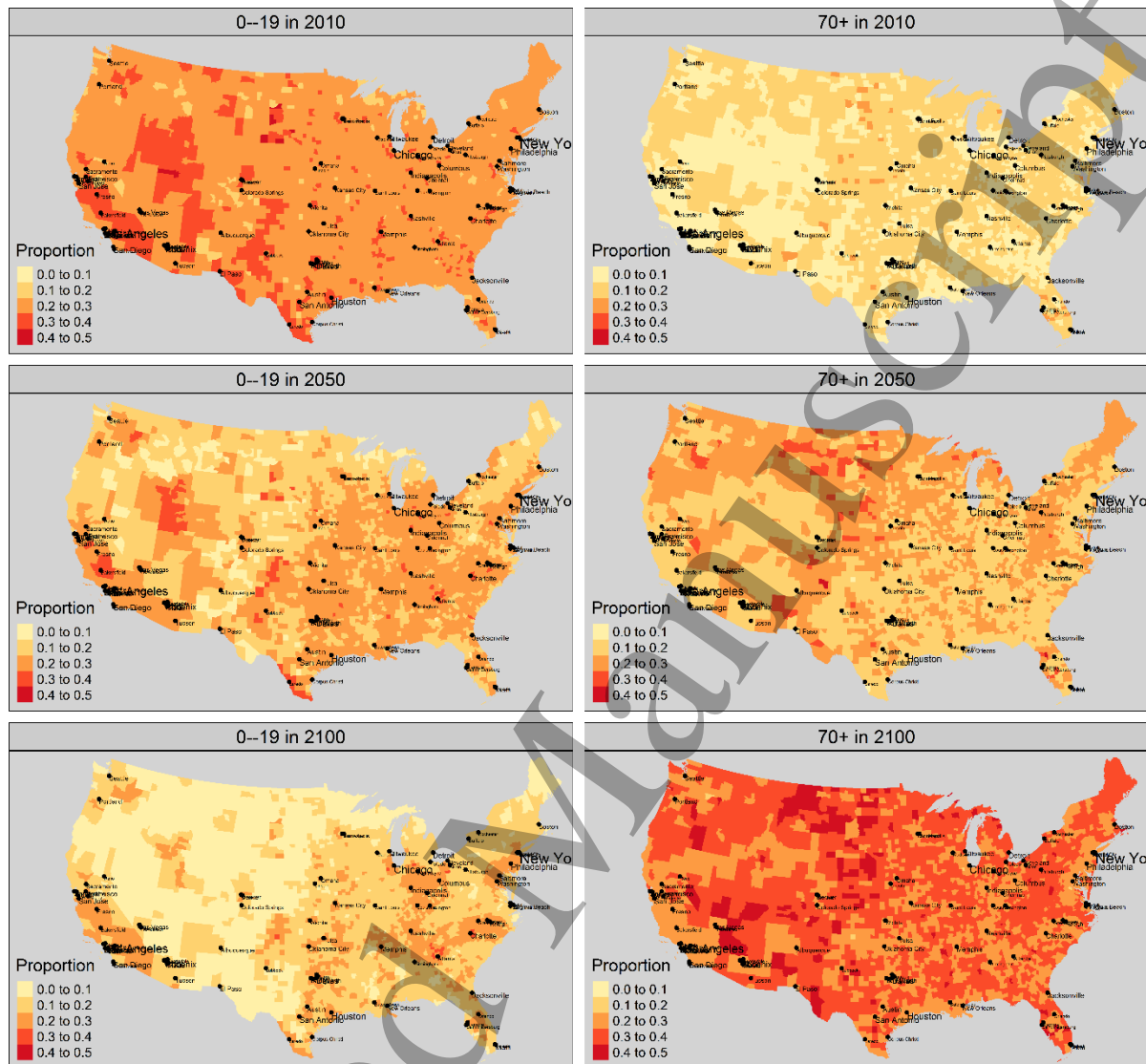


Figure 2. Observed (2010) and projected (2050, 2100) proportion 0-19 (left) and 70+ (right) under SSP2.

Although similar at the national level, patterns of aging by the end of the century differ substantially between SSP2 and SSP3 at the sub-national level. These two scenarios 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 (Figure 3), while urban counties, where most of the population resides, are affected by aging to a lesser extent in SSP3. The more intense urban concentration of younger people in SSP3 is due to the differences in the drivers of population change at the national level, as well as their manifestation at the county level. As the total U.S. population starts to shrink in the second half of the century under SSP3, a large proportion of counties are left with almost no population under the age of 20 due to sustained low fertility. Yet the empirical model allocates more youth to places with both high population density and growth that are surrounded by counties with similar characteristics. Positive neighborhood

1
2
3 effects, resulting from the gravitational force of large agglomerations, diminish with distance from
4 the urban cores and lead to the pattern that urban areas maintain higher fractions of children than
5 rural areas despite population decline at the national level.
6
7

8 A similar effect is at play in SSP5, which also produces a more intense concentration of the
9 younger population in urban areas than in SSP2 (Figure 3). In this case, however, it is the more
10 rapid population growth (rather than a shrinking population) that favors concentration of younger
11 age groups in and around cities. In addition, the higher population density plays a role in creating
12 more extensive urban and suburban areas of relatively young populations. Aggregated results for
13 all five SSPs in rural and urban counties across four macro-regions of the conterminous U.S. can
14 be found in Figure S15 of the SI Appendix. They support the finding of more rapid aging in rural
15 America described here, particularly in the Western parts of the country.
16
17
18
19
20
21

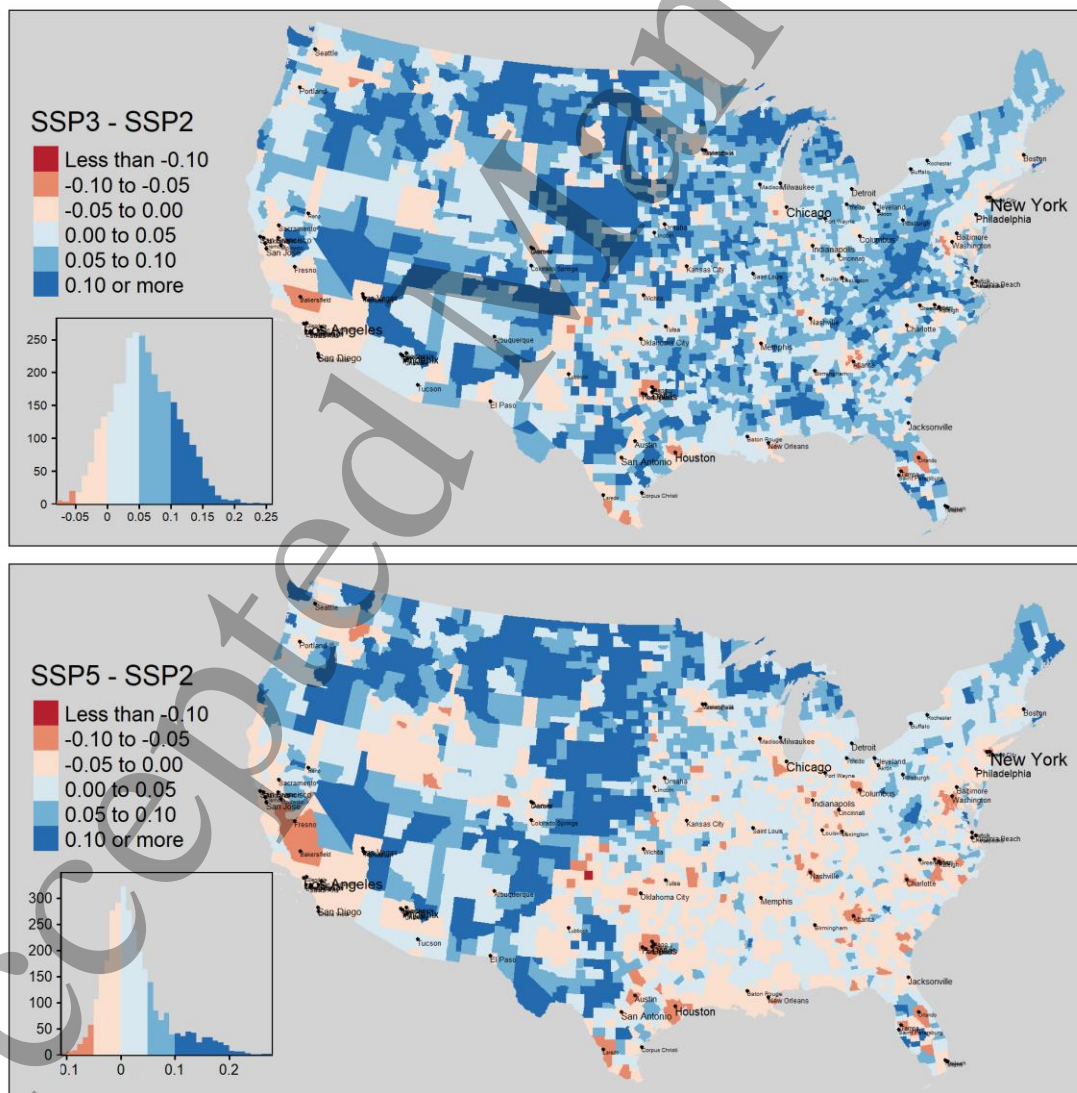


Figure 3. Difference in proportion 70+ between SSP3 and SSP2 (top) and SSP5 and SSP2 (bottom), U.S. in 2100.

While results in terms of proportions of population by age give a clearer indication of age structure changes separate from overall population growth, total numbers of people in each age group are also an important indicator of vulnerable populations. For example, the spatial distribution of the total numbers of people above the age of 70 by county (see Figure S16) shows that SSP5, which has a younger population overall than SSP2, actually has a larger number of people in the 70+ category in most counties due to the strong population growth in that scenario. Conversely, SSP3 has an older age structure but fewer people in the oldest age group due to slower overall population growth. Across SSPs, the spatial distribution of the number of people above age 70 does not shift in dramatic ways, in contrast to the substantial differences in the spatial distribution of population shares above age 70.

Discussion

Our projections provide plausible county level age distributions that are consistent with existing national and county-level population projections based on each of the SSPs.

Spatial projections of age-structured populations are critical to understanding risks posed by hazards from climate and other environmental changes, as well as to consumption behavior that may drive those changes. Our model provides a new category of demographic information to existing projections of U.S. population distribution that makes those projections much more relevant to risk analysis. We find the possibility of wide variation in age structure outcomes at the county level, as well as substantial aging clustered in particular parts of the country: Across all SSPs, cities and surrounding counties maintain a younger population age structure with a lower proportion in the most vulnerable 70+ age group. The largest of these clusters can be found in California, as well in the southern parts of New England and the Mid-Atlantic region, ranging roughly from Albany to Richmond. Remote rural counties, on the other hand, tend to age more rapidly. According to the U.S. National Climate Assessment's recent Climate Science Special Report (2017), climate change impacts will also vary regionally. Northern regions of the U.S. are expected to see the largest increases in the intensity of heatwaves, while chronic, long-duration drought will become increasingly likely in the Southwest. Combining societal projections with projections of environmental hazards such as heat waves, droughts, or floods can yield improved estimates of potential impacts on the most vulnerable segments of society with the potential for improving the rigor of intervention efforts and lowering their cost.

The model also represents a novel method in demography for projecting future age structure that is especially well suited to regions with data limitations. Although it differs from traditional cohort-component methods that treat components of population change directly, it takes these into consideration indirectly through proxy variables whose ability to implicitly capture the effects of

1
2
3 demographic rates has been successfully tested. We anticipate that the model will be useful for
4 projecting spatial age structure in places where the data requirements of conventional models
5 cannot easily be satisfied.
6

7
8 Data limitations for supporting empirically-based models used to make long-term (almost century
9 scale) projections pose a common challenge across many fields, and the work presented here is no
10 exception. The use of changes observed over one decade to project changes over a century can be
11 problematic, as the changes in age structure shares observed over that decade can be anomalous.
12 However, our analysis draws on 40 years of data (1970s through 2000s), substantially more than
13 many projection models in common use in the global change field, and our exploratory data
14 analysis (see SI section 1 for details) confirms that this period is in fact characterized by a data
15 generating process that is both stable and markedly different from the one observed in the earlier
16 (1940-1960) period.
17
18
19

20
21 Future work is planned to improve the model in the following aspects: (1) Investigate possibilities
22 to generate future scenarios beyond historically observed patterns, e.g. using empirical models
23 trained for different countries as representations of different sociodemographic trends. (2)
24 Incorporate a larger number of age groups. This requires dealing with the trade-off between model
25 performance and the number of age groups tracked simultaneously. (3) Improve validation to
26 better assess the model's performance in long-term projections. (4) Allow city size and distance to
27 city to vary dynamically across time and SSP. (5) Integrate further important dimensions of
28 population heterogeneity besides age, e.g. gender, race or educational attainment.
29
30
31

32 33 **Data Availability Statement**

34
35 The data that support the findings of this study are openly available at DOI. >>link to be included
36 before formal publication<<.
37
38
39

40 41 **Code Availability Statement**

42
43 Codes and materials used in this work are available upon requests. Correspondence should be
44 addressed to striess@iiasa.ac.at.
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

- Ando, Albert, and Franco Modigliani. 1963. "The 'Life Cycle' Hypothesis of Saving: Aggregate Implications and Tests." *The American Economic Review* 53 (1): 55–84.
- Barthelemy, Johan, and Thomas Suesse. 2016. "Package Mipfp: Multidimensional Iterative Proportional Fitting and Alternative Models v3.0." *SMART Infrastructure Facility - Papers*, January. <https://ro.uow.edu.au/smartpapers/175>.
- Bierwagen, Britta G., David M. Theobald, Christopher R. Pyke, Anne Choate, Philip Groth, John V. Thomas, and Philip Morefield. 2010. "National Housing and Impervious Surface Scenarios for Integrated Climate Impact Assessments." *Proceedings of the National Academy of Sciences* 107 (49): 20887–92. <https://doi.org/10.1073/pnas.1002096107>.
- Burch, Thomas K. 2018. "The Cohort-Component Population Projection: A Strange Attractor for Demographers." In *Model-Based Demography: Essays on Integrating Data, Technique and Theory*, edited by Thomas K. Burch, 135–51. Demographic Research Monographs. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-65433-1_10.
- CIA. 2017. "The World Factbook." 2017. <https://www.cia.gov/library/publications/resources/the-world-factbook/index.html>.
- Dong, Weihua, Zhao Liu, Hua Liao, Qihong Tang, and Xian'en Li. 2015. "New Climate and Socio-Economic Scenarios for Assessing Global Human Health Challenges Due to Heat Risk." *Climatic Change* 130 (4): 505–18. <https://doi.org/10.1007/s10584-015-1372-8>.
- Fienberg, Stephen E. 1970. "An Iterative Procedure for Estimation in Contingency Tables." *The Annals of Mathematical Statistics* 41 (3): 907–917.
- Gao, Jing. 2017. "Downscaling Global Spatial Population Projections from 1/8-Degree to 1-Km Grid Cells." NCAR Library. <https://doi.org/10.5065/D60Z721H>.
- Hallegatte, Stephane, Colin Green, Robert J. Nicholls, and Jan Corfee-Morlot. 2013. "Future Flood Losses in Major Coastal Cities." *Nature Climate Change* 3 (9): 802–6. <https://doi.org/10.1038/nclimate1979>.
- Hauer, Mathew E. 2019. "Population Projections for US Counties by Age, Sex, and Race Controlled to Shared Socioeconomic Pathway." *Scientific Data* 6: 190005.
- Hauer, Mathew E., Jason M. Evans, and Deepak R. Mishra. 2016. "Millions Projected to Be at Risk from Sea-Level Rise in the Continental United States." *Nature Climate Change* 6 (7): 691–95. <https://doi.org/10.1038/nclimate2961>.
- Jackson, Kenneth T. 1987. *Crabgrass Frontier: The Suburbanization of the United States*. Oxford University Press.
- Jiang, Leiwen, and Brian C. O'Neill. 2017. "Global Urbanization Projections for the Shared Socioeconomic Pathways." *Global Environmental Change* 42 (January): 193–99. <https://doi.org/10.1016/j.gloenvcha.2015.03.008>.
- Johnson, Kenneth M., Paul R. Voss, Roger B. Hammer, Glenn V. Fuguitt, and Scott Mcniven. 2005. "Temporal and Spatial Variation in Age-Specific Net Migration in the United States." *Demography* 42 (4): 791–812. <https://doi.org/10.1353/dem.2005.0033>.
- Johnson, Kenneth M., and Richelle L. Winkler. 2015. "Migration Signatures across the Decades: Net Migration by Age in U.S. Counties, 1950-2010." *Demographic Research* 32 (38): 1065–80. <https://doi.org/10.4054/DemRes.2015.32.38>.
- Jones, B., and B. C. O'Neill. 2016. "Spatially Explicit Global Population Scenarios Consistent with the Shared Socioeconomic Pathways." *Environmental Research Letters* 11 (8): 084003. <https://doi.org/10.1088/1748-9326/11/8/084003>.

- 1
2
3 Jones, Bryan, and Brian C O'Neill. 2016. "Spatially Explicit Global Population Scenarios Consistent with
4 the Shared Socioeconomic Pathways." *Environmental Research Letters* 11 (8): 084003.
5 <https://doi.org/10.1088/1748-9326/11/8/084003>.
6
7 Jones, Bryan, Brian C. O'Neill, Larry McDaniel, Seth McGinnis, Linda O. Mearns, and Claudia Tebaldi.
8 2015. "Future Population Exposure to US Heat Extremes." *Nature Climate Change* 5 (7): 652–55.
9 <https://doi.org/10.1038/nclimate2631>.
10
11 Jonkman, Sebastiaan N., Bob Maaskant, Ezra Boyd, and Marc L. Levitan. 2009. "Loss of Life Caused by
12 the Flooding of New Orleans After Hurricane Katrina: Analysis of the Relationship Between Flood
13 Characteristics and Mortality." *Risk Analysis* 29 (5): 676–98. <https://doi.org/10.1111/j.1539-6924.2008.01190.x>.
14
15 KC, Samir, and Wolfgang Lutz. 2017. "The Human Core of the Shared Socioeconomic Pathways:
16 Population Scenarios by Age, Sex and Level of Education for All Countries to 2100." *Global
17 Environmental Change* 42 (January): 181–92. <https://doi.org/10.1016/j.gloenvcha.2014.06.004>.
18
19 Lee, Ronald D., and Andrew Mason. 2011. "Generational Economics in a Changing World." *Population
20 and Development Review* 37 (s1): 115–42. <https://doi.org/10.1111/j.1728-4457.2011.00380.x>.
21
22 Lehner, Flavio, Clara Deser, and Benjamin M. Sanderson. 2018. "Future Risk of Record-Breaking Summer
23 Temperatures and Its Mitigation." *Climatic Change* 146 (3–4): 363–75.
24 <https://doi.org/10.1007/s10584-016-1616-2>.
25
26 Lutz, W., G. Amran, A. Belanger, A. Conte, N. Gailey, D. Ghio, E. Grapsa, et al. 2019. "Demographic
27 Scenarios for the EU: Migration, Population and Education." Luxembourg: Publications Office of
28 the European Union. <http://dx.doi.org/10.2760/590301>.
29
30 Manson, Steven, Jonathan Schroeder, David Van Riper, and Steven Ruggles. 2018. "National Historical
31 Geographic Information System: Version 13.0." Minneapolis, MN: IPUMS.
32 <https://doi.org/10.18128/D050.V13.0>.
33
34 Marsha, A., S. R. Sain, M. J. Heaton, A. J. Monaghan, and O. V. Wilhelmi. 2018. "Influences of Climatic
35 and Population Changes on Heat-Related Mortality in Houston, Texas, USA." *Climatic Change*
36 146 (3–4): 471–85. <https://doi.org/10.1007/s10584-016-1775-1>.
37
38 Mason, Andrew, and Ronald Lee. 2013. "Labor and Consumption across the Lifecycle." *The Journal of the
39 Economics of Ageing* 1–2 (November): 16–27. <https://doi.org/10.1016/j.jeoa.2013.06.002>.
40
41 Neumann, Barbara, Athanasios T. Vafeidis, Juliane Zimmermann, and Robert J. Nicholls. 2015. "Future
42 Coastal Population Growth and Exposure to Sea-Level Rise and Coastal Flooding - A Global
43 Assessment." *PLOS ONE* 10 (3): e0118571. <https://doi.org/10.1371/journal.pone.0118571>.
44
45 O'Neill, Brian C., Elmar Kriegler, Kristie L. Ebi, Eric Kemp-Benedict, Keywan Riahi, Dale S. Rothman, Bas J.
46 van Ruijven, et al. 2017. "The Roads Ahead: Narratives for Shared Socioeconomic Pathways
47 Describing World Futures in the 21st Century." *Global Environmental Change* 42 (January): 169–
48 80. <https://doi.org/10.1016/j.gloenvcha.2015.01.004>.
49
50 Ratcliffe, Michael, Charlynn Burd, Kelly Holder, and Alison Fields. 2016. "Defining Rural at the US Census
51 Bureau." *American Community Survey and Geography Brief*, 1–8.
52
53 Riahi, Keywan, Detlef P. van Vuuren, Elmar Kriegler, Jae Edmonds, Brian C. O'Neill, Shinichiro Fujimori,
54 Nico Bauer, et al. 2017. "The Shared Socioeconomic Pathways and Their Energy, Land Use, and
55 Greenhouse Gas Emissions Implications: An Overview." *Global Environmental Change* 42
56 (10.1016/j.gloenvcha.2016.05.009): 153–68. <https://doi.org/10.1016/j.gloenvcha.2016.05.009>.
57
58 Salvo, Joseph J., Arun Peter Lobo, and Erica Maurer. 2013. "New York City Population Projections by
59 Age/Sex & Borough, 2010-2040." New York: Department of City Planning.
60 [https://www1.nyc.gov/assets/planning/download/pdf/data-maps/nyc-
population/projections_report_2010_2040.pdf](https://www1.nyc.gov/assets/planning/download/pdf/data-maps/nyc-population/projections_report_2010_2040.pdf).

- 1
2
3 Striessnig, Erich, and Elke Loichinger. 2016. "Future differential vulnerability to natural disasters by level
4 of education." *Vienna Yearbook of Population Research* 13 (2015): 221–40.
5 <https://doi.org/10.1553/populationyearbook2015s221>.
6
7 Swanson, David A., Alan Schlottmann, and Bob Schmidt. 2010. "Forecasting the Population of Census
8 Tracts by Age and Sex: An Example of the Hamilton–Perry Method in Action." *Population*
9 *Research and Policy Review* 29 (1): 47–63.
- 10 Terama, Emma, Elizabeth Clarke, Mark D. A. Rounsevell, Stefan Fronzek, and Timothy R. Carter. 2017.
11 "Modelling Population Structure in the Context of Urban Land Use Change in Europe." *Regional*
12 *Environmental Change*, July, 1–11. <https://doi.org/10.1007/s10113-017-1194-5>.
13
14 Therneau, Terry, Beth Atkinson, Brian Ripley, and Maintainer Brian Ripley. 2015. "Package 'Rpart.'"
15 *Available Online: Cran. Ma. Ic. Ac. Uk/Web/Packages/Rpart/Rpart. Pdf (Accessed on 20 April*
16 *2016)*.
- 17 United Nations. 2014. "World Urbanization Prospects: The 2014 Revision." ST/ESA/SER.A/352. New
18 York, NY: Department of Economic and Social Affairs, Population Division, United Nations.
19 <http://esa.un.org/unpd/wup/Highlights/WUP2014-Highlights.pdf>.
20
21 ———. 2019. "World Population Prospects 2019." New York, NY: Department of Economic and Social
22 Affairs, Population Division. <http://esa.un.org/unpd/wpp/>.
23
24 U.S. Census Bureau, Population Division. 2018. "Census Bureau Regions and Divisions with State FIPS
25 Codes." [https://www2.census.gov/programs-surveys/popest/geographies/2017/state-](https://www2.census.gov/programs-surveys/popest/geographies/2017/state-geocodes-v2017.xlsx)
26 [geocodes-v2017.xlsx](https://www2.census.gov/programs-surveys/popest/geographies/2017/state-geocodes-v2017.xlsx).
27
28 U.S. Environmental Protection Agency. 2017. "Updates to the Demographic and Spatial Allocation
29 Models to Produce Integrated Climate and Land Use Scenarios (ICLUS) (Final Report, Version 2)."
30 EPA/600/R-16/366F. Washington D.C.
31 <https://cfpub.epa.gov/ncea/iclus/recordisplay.cfm?deid=322479>.
32
33 U.S. Global Change Research Program. 2017. "Climate Science Special Report: Fourth National Climate
34 Assessment, Volume I." Edited by D.J. Wuebbles, D.W. Fahey, K.A. Hibbard, D.J. Dokken, B.C.
35 Stewart, and T.K. Maycock. U.S. Global Change Research Program.
36 <https://doi.org/10.7930/J0J964J6>.
37
38 Whelpton, P. K. 1936. "An Empirical Method of Calculating Future Population." *Journal of the American*
39 *Statistical Association* 31 (195): 457–73. <https://doi.org/10.2307/2278370>.
40
41 Zagheni, Emilio, Raya Muttarak, and Erich Striessnig. 2016. "Differential Mortality Patterns from Hydro-
42 Meteorological Disasters: Evidence from Cause-of-Death Data by Age and Sex." *Vienna Yearbook*
43 *of Population Research* 13 (2015): 47–70. <https://doi.org/10.1553/populationyearbook2015s47>.
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60