Demographic change and migration due to sea-level rise in the United States

Mathew E. Hauer [*]

Department of Sociology, Florida State University

and

Sunshine Jacobs

Department of Sociology, Florida State University

and

Scott Kulp

Climate Central

September 24, 2021

Abstract

The warnings of sea-level rise driven migration first appeared in the scientific literature in the late 1970s when increased recognition that disintegrating ice sheets could drive people to migrate from coastal cities. Previous attempts to model this potential sea-level rise driven migration eschew two important considerations. First, sea-level rise migration models lack the crucial feedback loop whereby sea-level rise migrants alter the demographic trajectory in both their origin and destination. If sea-level rise forces millions of people further inland, a potential domino effect could result, further enhancing migration to more distant locations and further suppressing migration to coastal areas. Second, modelers have chosen to model migrants as age-less and sex-less individuals, overlooking well-established relationships between migration propensity and demographic characteristics. Here, we combine demographic models, flood hazard models, and a machine learning model built on 40 years of environmental migration in the US to project the US population distribution of US counties, accounting for anticipated demographic change, migration probabilities, and sea-level rise. Our initial results suggest demographic forces could operate as a significant "drag" on some sea-level rise migration, with some people becoming "demographically trapped" in increasingly hazardous coastal areas. Our results also suggest that the demographic implications of climate migration go far beyond simple displacement estimates, as population momentum from climate migrants multiplies the demographic effects in both origin and destination areas.

---

[*]Thanks y'all!

1 Introduction

Scientists first sounded the alarm of potential "major dislocations in coastal cities" (Mercer 1978) due to a disintegrating West Antarctic Ice Sheet and sea-level rise more than forty years ago. Since those early warnings, sea-level rise continues to be widely regarded as one of the most costly and visible future impacts of global climate change (McGranahan et al. 2007, Nicholls & Cazenave 2010, Strauss et al. 2015). With the global coastal population projected to eclipse one billion people this century (Neumann et al. 2015), sea-level rise is expected to affect and, in many cases, displace hundreds of millions of people (Hauer 2017, Nicholls & Cazenave 2010).

While the scientific community understands that climate change and sea-level rise will affect millions of people in the United States and across the world, we face a critical knowledge gap about the impacts of rising sea levels on human migration. Previous attempts to model sea-level rise and migration eschew two important considerations. First, sea-level rise migration models lack the crucial feedback loop whereby sea-level rise migrants alter the demographic trajectory in both their origin and destination. If sea-level rise forces millions of people further inland, a potential domino effect could result, further enhancing migration to more distant locations and further suppressing migration to coastal areas (Curtis & Schneider 2011, Döös 1997, Hauer et al. 2020).

Second, modelers have chosen to model migrants as age-less and sex-less individuals (Hauer 2017, McGranahan et al. 2007, Neumann et al. 2015, Strauss et al. 2015), ignoring the well-established relationship between migration propensity and demographic characteristics (Black et al. 2011, Clark & Maas 2015, De Longueville et al. 2019, Dieleman 2001). This demographic oversight could be due to the relative youthful research on the topic, a lack of sophisticated demographic modeling, the relative paucity of demographic and migration data in most of the world, or the combination of these factors. Regardless, this demographic ignorance is particularly important given the demographic shifts anticipated this century (Colby & Ortman 2017). Those with the least resources are least likely to migrate when facing environmental pressures (Black et al. 2011) and by ignoring these well-established relationships, past sea-level rise migration models likely overestimate those most likely to migrate and grossly underestimate those who might be "trapped" in coastal

communities. These gaps combine to create divergent estimates of who will be most affected and how sea-level rise will ultimately alter population distributions. The challenge is to develop better coupled natural-human systems for assessing sea-level rise impacts that properly couple climate projections with projections of demographic change.

In this paper, we combine matrix population models, flood hazard models, and a machine learning model built on 40 years of environmental migration in the US to project the US population distribution of US counties, accounting for anticipated demographic change, migration probabilities, and SLR. We ask the two fundamental research questions: To what extent will sea-level rise migration alter the distribution of the demographic trajectory of the US population? And How might changes in future age structures alter migration dynamics for those impacted by sea-level rise? In addressing these questions, this project answers the Intergovernmental Panel on Climate Change's (IPCC) call to "construct impact assessments in which biophysical futures are coupled with socioeconomic futures"[1].

2   Background and Overview

The implications of sea-level rise on human migration first appear in the scientific literature more than forty years ago (Mercer 1978, Schneider & Chen 1980), when scientists started to sound the alarm of a disintegrating West Antarctic Ice Sheet and the potential "major dislocations in coastal cities" (Mercer 1978). Since those early warnings, sea-level rise and associated impacts continue to be widely regarded as one of the most costly and irreversible future consequences of global climate change (on Climate Change IPCC). Scientific projections of global mean sea-level rise for the year 2100 range from a low of 0.4 meters to a high of 2.5 meters (Jevrejeva et al. 2012, Sweet et al. 2017), depending on assumptions of future anthropogenic greenhouse gas emission levels, thermal expansion of ocean masses, sea surface temperatures, rates of melting and decomposition of Antarctic and Greenland ice sheets, and isostatic adjustment as large ice sheets disappear. These sea-level rise projections are likely conservative and high-end sea-level rise is increasingly realistic (Bamber et al. 2019). With the global coastal population projected to eclipse one billion people this century (Neumann et al. 2015), sea-level rise is expected to affect and,

---

[1]IPCC Fifth Assessment Report, p.213.

in many cases, displace hundreds of millions of people (Hauer et al. 2020).

Since the first studies to quantify population displacement due to sea-level rise and document changes in human migration related to sea-level rise, fundamental understanding of sea-level rise and human migration has rapidly advanced with the development of basic theory on climate change migration (Black et al. 2011), empirical case studies of historical analogues for future sea-level rise (e.g. Arenstam Gibbons & Nicholls (2006), Bailey et al. (2014), Trincardi et al. (2016)), integrated economic analysis and modelling of sea-level rise retreat (Yohe & Schlesinger 1998), explicit models of sea-level rise migration (Hauer 2017, Chen & Mueller 2018), as well as contentious policy discussions on the need for coastal retreat and adaptive infrastructure (Hino et al. 2017). In some cases, studies even question *if* sea-level rise will spur widespread migration at all (Kniveton 2017, Laurice Jamero et al. 2017).

Sea-level rise influences human migration in multiple ways. The most straightforward, apparent, and existential migration pressure involves permanent, irreversible inundation of low-elevation areas, under which a certain rise in sea level renders the built environment uninhabitable (Hauer et al. 2016, Poulter & Halpin 2008). Migration is widely understood as a choice in which individuals weigh the economic, social, and emotional costs and benefits of moving (Black et al. 2011). Thus, the decision to migrate is multidimensional (Clark & Maas 2015).

Most sea-level rise migration studies only model migration along environmental factors (Davis et al. 2018, Nicholls & Cazenave 2010), or couple environmental factors with economic (Chen & Mueller 2018) or social factors (Robinson et al. 2020). Conspicuously missing from these models is the integration of demographic factors. This omission is particularly egregious given the availability of high-quality demographic data in the United States and the relative simplicity of integrating demographic change into migration models.

Despite these rapid advancements, clear significance of this dimension of global environmental change in the US, the lively debate surrounding the topic, and the variety of perspectives and methodologies offered in the literature, several scientific challenges remain. Notably, a fully integrated, dynamic demographic/sea-level rise model.

This paper fills this gap by incorporating migration modeling within a demographic

change model called the Demographic Metabolism (Lutz & Muttarak 2017, Ryder 1985). To answer our research questions, we employ two main modules. The first is the development of a parsimonious, one-dimensional, age-specific displacement model. The second is a multiregional Leslie matrix specification.

3   Methods and Materials

To answer our research questions, we employ three main modules. The first is the development of a parsimonious, one-dimensional, age-specific displacement model. The second is a multiregional Leslie matrix specification. And the third is a sea-level rise exposure model.

3.1   A parsimonious one-dimensional, age-specific displacement model

Our displacement model makes use of a statistical time series outlier detection technique to first identify anomalous demographic behavior in a time series and then verify that this anomalous behavior is associated with an environmental event. In the following sections, we describe this algorithm, the data sources we use, and finally the devleopment of the displacement model itself.

3.1.1   Statistical Outlier Detection

We use a statistical time series outlier detection algorithm (Chen & Liu 1993), implemented in the R programming language (R Core Team 2019) via the tsoutliers package (López-de-Lacalle 2019). This algorithm iteratively uses ARIMA models to 1) identify potential outliers and 2) refit the ARIMA with the outliers removed to produce a counter-factual time series. Here we briefly summarize and describe the method.

Often, the behavior of a time series can be described and summarized in ARIMA models. If a series of values, y_t^*, is subject to m interventions or outliers at time points $t_1, t_2, \ldots, t_m$ with weights ω then y_t^* can be defined as

$$y_t^* = \sum_{j=1}^{m} \omega_j L_j(B) I_t(t_j) + \frac{\theta(B)}{\phi(B)\alpha(B)} a_t \qquad (1)$$

Where $I_t(t_j)$ is an indicator variable with a value of 1 at observation t_j and where the jth outlier arises,

$\phi(B)$ is an autoregressive polynomial with all roots outside the unit circle,

$\theta(B)$ is a moving average polynomial with all roots outside the unit circle,

and $\alpha(B)$ is an autoregressive polynomial with all roots on the unit circle.

We examine three types of outliers at time point t_m:

1. additive outliers (AO), defined as $L_j(B) = 1$;

2. level shift outliers (LS), defined as $L_j(B) = 1/(1 - B)$; and

3. temporary change outliers (TC), defined as $L_j(B) = 1/(1 - \delta B)$ where δ is equal to 0.7.

Colloquially, additive outliers arise when a single event causes the time series to unexpectedly increase/decrease for a single time period; level shift outliers arise when an event causes the time series to unexpectedly increase/decrease for multiple time periods; and temporary change outliers arise when an event causes the time series to unexpectedly increase/decrease with lingering effects that decay over multiple time periods.

An outlier is detected with the estimated residuals using a regression equation

$$\pi(B)y_t^* \equiv \hat{e} = \sum_{j=1}^{m} \omega_j \pi(B) L_j(B) I_t(t_j) + a_t \qquad (2)$$

where $\pi(B) = \sum_{i=o}^{inf} \pi_i B^i$.

Equations 1 and 2 allow for an automatic detection of outliers iterated over a two-stage process.

In stage 1, outliers are located. First, an ARIMA model is fit to the time series using the `forecast` package in R (Hyndman et al. 2019, Hyndman & Khandakar 2008) where the best performing ARIMA model is selected based on the Bayesian information criterion (BIC). Next, the residuals from the forecast are checked for their significance using equation 2 where only outliers above a critical t-static are considered "true" outliers ($|\tau| \geq 4$; p-value < 0.000063). We chose this threshold to minimize the probability of committing a Type I error (or claiming an outlier is true when it is in fact not). Finally, two additional rules

are implemented: If multiple outliers are detected at the same time point, only the most significant outlier is selected and if outliers of the same type at consecutive time periods are detected, only the outlier with highest t-statistic is selected.

In stage 2, outliers are removed from the time series and a new ARIMA model is chosen and fit. The selection of the initial ARIMA model could have been affected by the presence of the outliers, making some outliers spuriously identified. To correct for this, a new ARIMA model is fit accounting for additional regression effects in equation 1 from the list of candidate outliers identified in stage 1, effectively removing the outliers from the time series. Each outlier is then reassessed under the new model and those outliers that are no longer significant are removed.

These two stages are then iterated until no additional outliers are detected.

Figure 1 shows a toy example for outlier detection in a time series using the example of Hurricane Katrina in Orleans Parish. On August 23 2005, Hurricane Katrina, a category 5 hurricane, struck southern Louisiana causing widespread damage and destruction in Orleans Parish in particular. The displacement from the Hurricane and the federal response were well documented (Hori et al. 2009, Fussell et al. 2014) and Census estimates suggest Orleans Parish lost more than 200,000 residents between 2005 and 2006. What would have been Orleans' population estimate had Hurricane Katrina *not* occurred? A simple counter-factual estimate might keep population size just under 500,000 people ($\hat{y}$).

In the Hurricane Katrina example, we have knowledge of the impact of Hurricane Katrina after the fact or ex-post-facto in order to create the counter-factual time series $\hat{y}$. But is this outlier detectable without knowledge of the Hurricane Katrina? In other words, can we detect the reduction of Orleans Parish's population between 2005 and 2006 using only the time series? Absolutely. The tsoutliers package identifies 2006 as an extremely strong level-shift outlier (t-stat = -96.39).

3.1.2 Data

For the development of our parsimonious, one-dimensional displacement model, we use two primary datasets: the National Vital Statistics System (NVSS) U.S. Census Populations
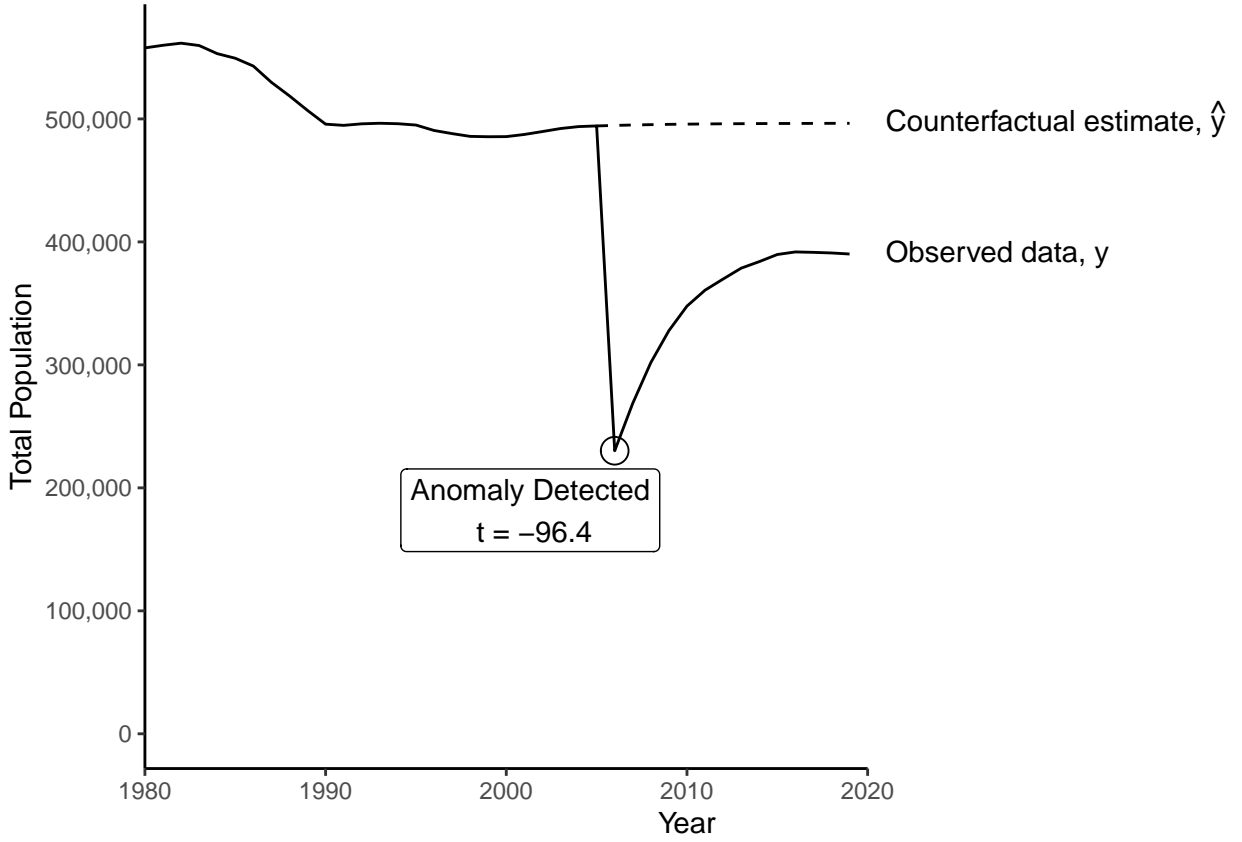
Figure 1: **An example of outlier detection using Hurricanes Katrina and Rita in Orleans Parish Louisiana in 2005.** Hurricane Katrina struck Louisiana in 2005 and we use it as a toy example to illustrate our approach. This figure shows the annual time series of total population in Orleans Parish between 1980 and 2019. Between 1990 and 2005, Orleans Parish total population changed from 495k to 494k, suggesting a possible 'plateau' in the population (illustrated with the dotted 'counterfactual'). Hurricane Katrina and the widespread population loss of more than 200k people represent a very strong outlier (t=-96.39).

with Bridged Race Categories data set[2] and the Spatial Hazard Events and Losses Database for the United States (SHELDUS).

The NVSS Bridged Race Categories data set harmonizes racial classifications across disparate time periods to allow population estimates to be sufficiently comparable across space and time. Importantly, all county boundaries are rectified to be geographically consistent across all time periods. We use the the 1969-2018 dataset which includes annual population estimates in five year age groups (0-4,..., 85+), two sex groups (male and female), and three race groups (White, Black, Other).

In our statistical outlier analysis (detailed below), we only consider counties created prior to 2000 and contained in the NVSS data. NVSS aggregated all counties in Hawaii to the state-level in the 1969-2018 NVSS bridged race data and we exclude them from our analysis. Several counties were created after 2000 (most notably is Broomfield County, Colorado). The 15 counties excluded from our analysis due to boundary changes or other reasons are Hoonah-Angoon Census Area AK 02105, Kusilvak Census Area AK 02158, Prince of Wales-Outer Ketchikan Census Area AK 02201, Skagway-Hoonah-Angoon Census Area AK 02232, Wrangell-Petersburg Census Area AK 02280, Adams County CO 08001, Boulder County CO 08013, Broomfield County CO 08014, Jefferson County CO 08059, Weld County CO 08123, Hawaii County HI 15001, Honolulu County HI 15003, Kalawao County HI 15005, Kauai County HI 15007, and Maui County HI 15009.

We use these data in two separate steps. In our statistical outlier analysis, we aggregate all county-level estimates into annual total population estimates for each county for the period 1969-2016. The historical population estimates prior to 1980 display unusual volatility, so we consider only the time periods 1980-2018. We use the NVSS population estimates disaggregated by age/sex to calculate cohort-change ratios, which we describe below.

The second data source we use is SHELDUS. SHELDUS is a county-level hazard data set for the US which contains information about the direct losses (property and crop losses, injuries, and fatalities) caused by a hazard event (thuderstorms, hurricanes, floods, wildfires, tornados, flash floods, earthquakes, etc.) for the period 1960 to the present. We

---

[2]Data can be downloaded here: `https://seer.cancer.gov/popdata/download.html`

use this database to ensure the county time periods we identify as statistical outliers with population losses contain experienced an environmental hazard in that county-year. This is to ensure the outlier population losses that we detect are associated with a hazard rather than other forces, such as economic forces.

3.1.3 Development of the Displacement Model

We search all US counties for negative statistical outliers (indicating population losses) between 1980 and 2018. We detect population losses of magnitude 4σ or greater. We then searched the SHELDUS database to see if these county-years experienced per capita hazardous losses in excess of the 50th percentile. Four county-periods either were not in the SHELDUS database or experienced per capita hazard losses below the 50th percentile. Additionally, one county-period contained age-sex groups with 0 people, necessitating exclusion. These 48 environmental events include tornados, wind damage, winter weather, earthquakes, flooding, tropical cyclones, hail, and other environmental hazards.

With 48 county-periods exhibiting large population declines after verified hazard losses, across over nearly 40 years and across a wide variety of environmental hazards, we are able to overcome the first limitation of previous environmental migration modelling: that of isolated case studies of individual environmental events. With this large sample of environmental displacement, we next overcome the second limitation: that of only examining total population. To examine environmental migration across age, we build a flexible, one-dimensional, age-specific displacement model.

To link population displacement with age-specific population changes, we calculate cohort-change ratios in each county. Using the demographic accounting equation, the population at some time period t is equal to the $Pop_{t-1} + Births_{t-1} - Deaths_{t-1} + Migrants_{t-1,in} - Migrants_{t-1,out}$. For all age groups older than 0, $Births_{t-1}$ must be 0.

The calculation of a cohort-change ratio, as its name suggests, is relatively straightforward:

$$CCR_{x,t} = \frac{{}_nP_{x,t}}{{}_nP_{x-y,t-y}}$$

Where $_nP_{x,t}$ is the population aged x to $x+n$ in time t and $_nP_{x-y,t}$ is the population aged $x-y$ to $x+n-y$ in time t where y refers to the time difference between time periods. Since mortality must decrement a population, any CCR above 1.0 implies a net-migration rate in excess of the mortality rate, and a growing population.

We build the following model based on the relationship between the change in CCRs at age x, $\Delta CCR_x = CCR_{x,t}/CCR_{x,t-1}$, and the percentage decline in the total population compared to the counter-factual in the outlier detection method, $\Delta P_t = \hat{P}_t/P_t$:

$$log(\Delta CCR_x) = a_x + b_x h$$

Here, h is the $log(\Delta P_t)$ and shows a linear relationship with the logarithm of the change in CCRs by age. x refers to five-year age groups: 0-4, 5-9,..., 85+. This is a similar model and approach to Wilmoth et al's (2012) flexible, one-dimensional mortality model based on the linearity between age-specific mortality rates and infant-mortality. **Table 1** depicts the correlation coefficients between $log(\Delta CCR_x)$ and $log(\Delta P_t)$. The age groups with the lowest correlation coefficients are young adult males aged 20-39 and those in the open-ended age interval (80+), suggesting these age/sex groups react to environmental signals the least predictably.

Figure 2 shows the relationship for selected age-groups between $log(\Delta CCR_x)$ and $log(\Delta P_t)$. Here we can see the near linear relationship between the percentage reduction in the population and the percentage reduction in CCR_x. Note the two county-years with the greatest reductions in total population refer to Louisiana counties most impacted by Hurricane Katrina.

3.1.4 Environmental Migration estimation using the fitted model

Using $h = log(\Delta P_t)$, we can estimate age-specific changes in CCRs after an environmental event by simply applying the following formula:

$$\hat{CCR}_{x,t} = CCR_{x,t-1} \cdot e^{\hat{\beta}_x h} \tag{3}$$

Where $e^{\hat{\beta}_x h}$ provides the percentage change in CCR_x based on the $log(\Delta P_t)$. In this case, we drop the intercept (a) from the estimation procedure to ensure a 0% decline in

Table 1: **Correlation coefficients of changes in Cohort Change Ratios vs. Total Population Change (n=48).**

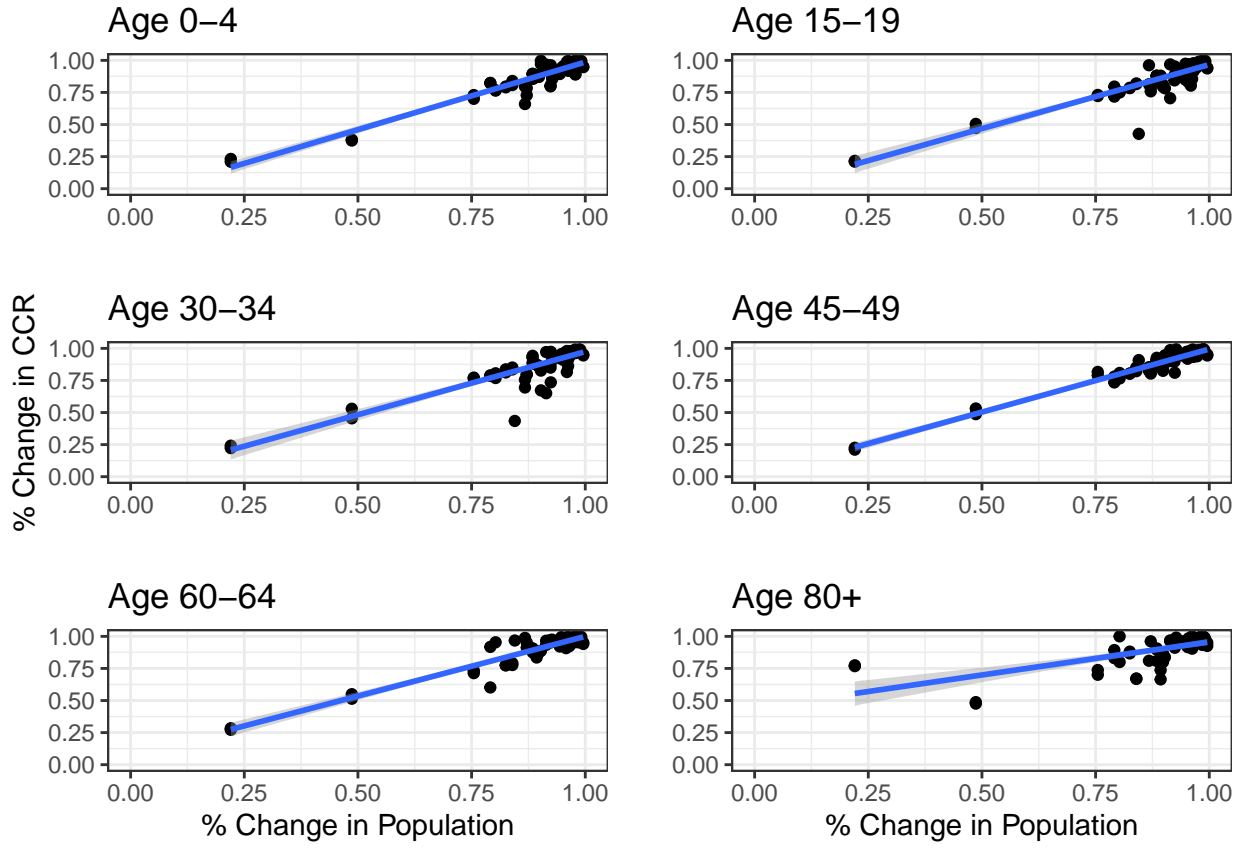| Age Group	Male	Female
0-4	0.931	0.944
5-9	0.954	0.934
10-14	0.963	0.965
15-19	0.806	0.930
20-24	0.650	0.919
25-29	0.710	0.934
30-34	0.752	0.959
35-39	0.874	0.975
40-44	0.936	0.964
45-49	0.971	0.963
50-54	0.969	0.977
55-59	0.965	0.958
60-64	0.896	0.888
65-69	0.809	0.912
70-74	0.937	0.826
75-79	0.839	0.795
80+	0.478	0.651

Figure 2: **Age-specific changes in cohort-change ratios (ΔCCR_x) vs. change in total population (ΔP_t).**
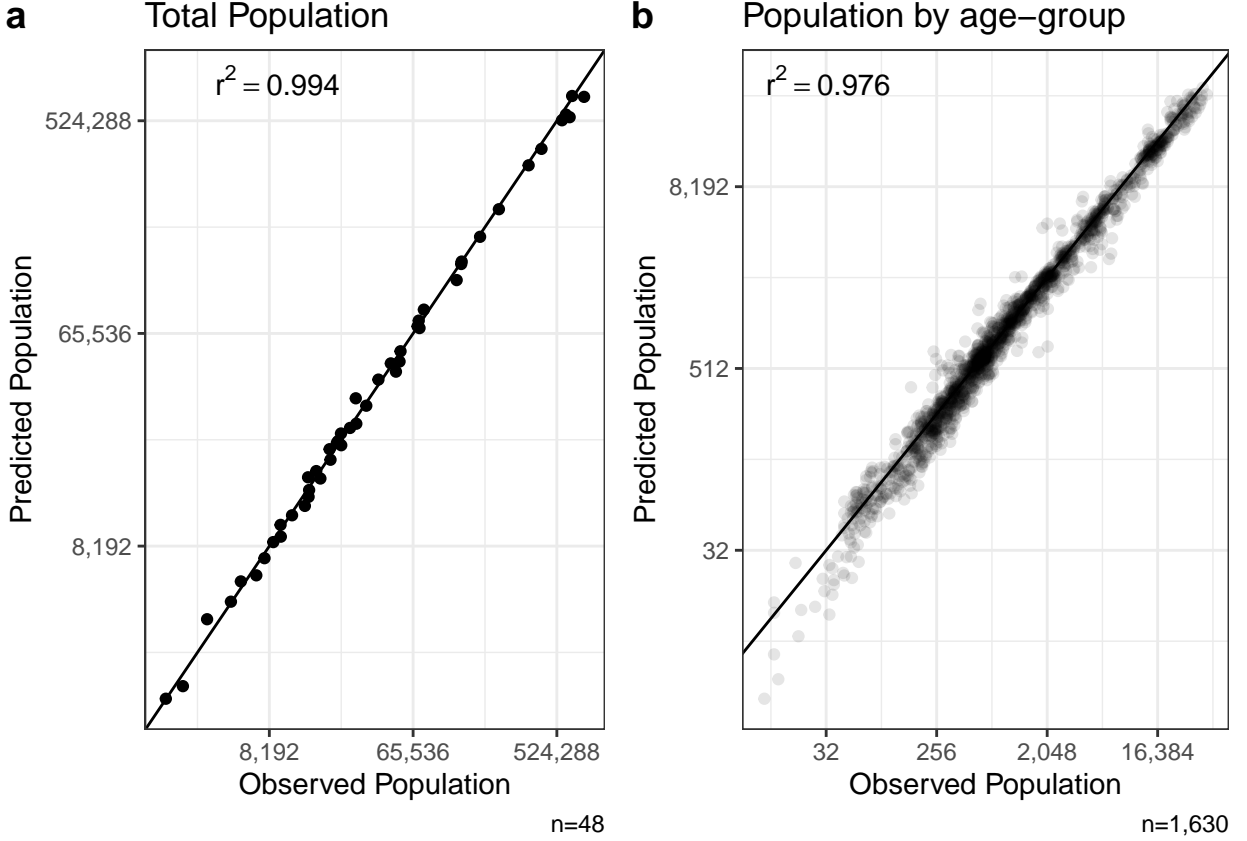
Figure 3: **Relationship between predicted populations using our model and observed populations in the 48 counties in our selection.** (a) shows the total population and (b) shows the populations for each age/sex group. The diagonal solid line is y=x. The model produces good fits for both the total population and by age/sex group.

population yields a corresponding 0% change in the CCR. Simply multiplying the result from the model with the CCR in the year prior yields the anticipated change in the CCR. These changes in CCRs can then be applied to any time series of population values to generate an anticipated population.

Figure 3 shows the accuracy of the our fitted, one-dimensional model. We estimate the predicted population using the equation outlined above and then compare against the observed population. Regarding total population in our 48 counties, our model performs well with an r^2 value of 0.994 and performs well regardless of population size. Regarding

each individual P_x group, our model performs slightly worse, but still performs quite well with an r^2 of 0.976. And just like with total population, the accuracy of our model does not depend on population size.

3.2   Matrix Population Models

We employ the use of multiregional leslie matrices in our population projections (Rogers 1966). To illustrate our leslie matrix and projection specification, we use an example of a three-age group, two region population.

In a typical Leslie matrix,

$$\mathbf{P_{t+1}} = \mathbf{S_t} \cdot \mathbf{P_t} \tag{4}$$

Where $\mathbf{P_t}$ refers to the population matrix containing k age groups and $\mathbf{S_t}$ contains the age-specific fertility and mortality rates.

Thus, to produce a population projection, the operation looks akin to

$$\begin{bmatrix} 0 & F_2 & F_3 \\ S_1 & 0 & 0 \\ 0 & S_2 & S_3 \end{bmatrix} \cdot \begin{bmatrix} P_1 \\ P_2 \\ P_3 \end{bmatrix} = \begin{bmatrix} 0 & 0.2 & 0.2 \\ 0.62 & 0 & 0 \\ 0 & 0.35 & 0.35 \end{bmatrix} \cdot \begin{bmatrix} 100 \\ 100 \\ 100 \end{bmatrix} = \begin{bmatrix} 40 \\ 68 \\ 71 \end{bmatrix} \tag{5}$$

In a multi-regional projection model, we can use a "super-matrix."

A 2-region model would take the following general form for the $\mathbf{S_t}$ matrix,

$$\mathbf{MS_{t+1}} = \left[\begin{array}{c|c} \mathbf{S_i} & \mathbf{M_{j \to i}} \\ \hline \mathbf{M_{i \to j}} & \mathbf{S_j} \end{array} \right] \tag{6}$$

Where the diagonal $\mathbf{S}$ matrices contain combined survival, fertility, and migration probabilities while the off-diagonal $\mathbf{M}$ matrices contain the probabilities of each age-group moving from one region to the other.

In practice, we create these matrices by multiplying a migration matrix, $\mathbf{M}$, which contains the probability of migrating from the ith region to the jth region in the diagonal by the $\mathbf{S}$ matrix above. In this example, 15% of age group 2 from region i moves to region j.

16

$$\mathbf{MS_{t+1}} = \left[\begin{array}{ccc|ccc} & \mathbf{MS} & & & \mathbf{MS_{j \to i}} & \\ \hline & \mathbf{MS_{i \to j}} & & & \mathbf{MS} & \end{array} \right] \tag{7}$$

$$\mathbf{S_t} = \left[\begin{array}{ccc|ccc} 0 & 0.2 & 0.2 & 0 & 0 & 0 \\ 0.62 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.35 & 0.35 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0.2 & 0.2 \\ 0 & 0 & 0 & 0.67 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.33 & 0.33 \end{array} \right] \cdot \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.85 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0.15 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right] = \tag{8}$$

$$\left[\begin{array}{ccc|ccc} 0 & 0.2 & 0.2 & 0 & 0 & 0 \\ 0.527 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.35 & 0.35 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0.2 & 0.2 \\ 0.093 & 0 & 0 & 0.67 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.33 & 0.33 \end{array} \right] \tag{8}$$

Where the migration rate from i to j in age group 2 is 0.15.

3.2.1   Data

We populate our Leslie matrices with the 3,143 US counties, 18 five-year age groups, and 2 sex-groups. We use CCRs to populate our S values in each matrix and child-woman ratios to populate our F values. The values come from the NVSS Bridged Race Categories. To project the CCRs, We employ the use of an autoregressive integrated moving average (ARIMA) model for forecasting equally spaced univariate time series data. We use an `ARIMA(0,1,1)` model which produces forecasts equivalent to simple exponential smoothing. All projections were undertaken in **R** (R Core Team 2019) using the `forecast` package (Hyndman et al. 2019).

Where an ARIMA(0,1,1) model is

$$Y_t = Y_{t-1} + e_t - \theta e_{t-1} \tag{9}$$

$$\hat{Y}_{t+1} = Y_t - \theta e_{t-1} \tag{10}$$

17

where e_t is independent and identically distributed as $N(0, \sigma_e^2)$. It can be shown that $\hat{Y}_{t+1}$ is an exponentially weighted moving average of the observations $Y_t, Y_{t-1}, ...$ with weights $\theta_1^k(1 - \hat{\theta}_1)k = 0, 1, ...,$ and that the additional forecasts $\hat{Y}_{t+j}$ for $j > 1$ remain constant at $\hat{Y}_{t+1}$ (**?**, p.158).

Migration We also calculate the probability of migrating from each county to each county. These data come from the IRS county-to-county migration files for the period 1990-2018 [see Hauer & Byars (2019) and DeWaard et al. (2021) for descriptions of this data). The IRS began using tax data to estimate migration in the 1970s and 1980s (Engels & Healy 1981, Franklin & Plane 2006) and began releasing migration data in 1990. The IRS uses individual federal tax returns, matches these individual returns between two tax years (for instance tax year 2000 and tax year 2001), and identifies both migrants and non-migrants. Beginning with tax year 1991 (migration year 1990), the IRS produces these data in conjunction with the US Census Bureau using the IRS Individual Master File which contains every Form 1040, 1040A, and 1040EZ (Gross 2005). Migration is identified when a current years' tax form contains an address that is different from the matched preceding years' return. A non-migrant is identified when there is no change in address between two years. For the 2002 tax year, the IRS migration data contained approximately 130 million returns (Gross 2005).

The annual series of county-to-county migration data cover 95 to 98 percent of the tax-filing universe (or approximately 87% of US households (Molloy et al. 2011)) and their dependents making these data the largest migration data source for count flows between counties in the United States. The IRS derives migration information from tax-filings making those who do not file taxes most likely to be underrepresented in the migration data (Gross 2005, DeWaard et al. 2016), namely undocumented populations, the poor, the elderly, and college students (Gross 2005). However, the overwhelming majority of householders file US tax returns in the United States (Molloy et al. 2011).

To capture changes in the migration system, we employ the use of an ETS model (Error, Trend, Seasonal), a univariate time series forecasting model (Hyndman et al. 2008). We use an ETS model for migration as opposed to an ARIMA(0,1,1) as we do for the CCRs is because the CCRs are subject to multiplication during a drift whereas the migration

system is not. A CCR that drifts from $1.1 \rightarrow 1.3$ represents more than a 10-fold increase in a projected population over our time horizons ($1.1\char`^17 = 5x$, $1.3\char`^17 = 86x$). Projecting the migration system itself is not subject to such exponential drift.

We fit individual ETS models for each county's numeric migrants between each origin-destination dyadic pair using the `forecast` package in R (Hyndman et al. 2019). This approach allows for the underlying migration system to evolve and change over the projection horizon, allowing dyadic pairs to either wax or wane. The numeric projections are then converted to fractions of the total projected migrants in order to populate the diagonal in the **M** matrix above where non-migrants (ie those migrating from $i \rightarrow i$) are included. The result is the fraction of individuals surviving from age group a who migrant from $i \rightarrow j$.

3.3   Sea-level Rise Exposure

To estimate the populations at-risk to SLR and thus likely migrants, we employ the use of inundation modelling (Hauer et al. 2020) which assumes that people who are underwater 100% of the time must migrate. We estimate these populations using airborne lidar-derived digital elevation models (DEMs) produced by NOAA and supplemented with both the USGS Northern Gulf of Mexico Topobathymetric DEM in Louisiana and the USGS National Elevation Dataset in the fraction of land not covered by other sources (see Kulp & Strauss (2019) for details on the construction of the DEMs).

Using a bathtub model of inundation, we calculate the land area under a given water height to generate binary inundation surfaces. SLR exposure is hyperlocalized and we generate this inundation area in the Census Block Groups (CBG; n=81,815) located in coastal counties (n=406) expected to experience any probability of flooding. We use probabilistic SLR projections (Sweet et al. 2017) that are closely aligned with the IPCC for our water heights. To calculate the land area under a given water height, we simply threshold the DEM to find pixels below SLR_{yt} where y is the projected height of SLR in a given year t. For each CBG, we simply calculate the percentage of its pixels on dry land (defined in the National Wetland Inventory (U.S. Fish and Wildlife Service 2012)) covered by the inundation surface and multiply this percentage by the total population in the CBG in 2010 to produce the total number of people at-risk to a given amount of SLR in a given

year. We then aggregate these CBGs to the county-level to calculate the percentage of people in a given county at-risk of inundation.

3.4   Full Model

Our full model uses a multiregional leslie matrix with the dimensions 3143 counties, 18 five-year age groups, and 2 sex groups.

Recall that in a typical leslie matrix $\mathbf{P_{t+1}} = \mathbf{S_t}\mathbf{P_t}$ and a multiregional leslie matrix is $\mathbf{P_{t+1}} = \mathbf{M_t}\mathbf{S_t}\mathbf{P_t}$ where $\mathbf{S_t}$ contains the age-sex specific probabilities of surviving and fertility information and $\mathbf{M_t}$ contains the probability of migrating between regions. Since our survival probabilities in $\mathbf{S_t}$ are comprised of cohort-change ratios and our displacement model generates alterations in CCRs, we can specifify three different sets of demographic projections: a projection without incorporating climate migration, a projection incorporating climate migration, and a projection incorporating simple displacement.

A simple demographic projection simply uses the specifications of the leslie matrix laid out in Equation 4. This produces a set of demographic projections without accounting for climate migration. We refer to this projection as the *"base projection."*

We also generate a second, simple projection that incorporates simple displacement. Many climate migration models employ simple displacement models and we recreate those models here. We produce this by simply multiplying the total projected population at time step t in each county by the percentage of population expected to be inundated at each time step. This allows us to evaluate how our integrated models compare to existing approaches in the literature. We refer to this projection as the *"inundation projection."*

Finally, we produce an integrated population projection which we refer to as the *migration projection*. We produce this projection in several steps but the full specification is

$$\mathbf{P_{t+1}} = \mathbf{M_t}(\hat{\mathbf{S}}_\mathbf{t}\mathbf{P_t} - \mathbf{S_t}\mathbf{P_t}) + \hat{\mathbf{S}}_\mathbf{t}\mathbf{P_t}$$

Where $\mathbf{M_t}$ is the matrix containing the probabilities of migrating from county i to county j, $\hat{\mathbf{S}}_\mathbf{t}$ contains the survival probabilities for each age group a for each county c for each time period t that is reduced following Equation 3. In this case, h is the percentage of

the population at-risk to SLR. $\mathbf{S_t}$ is the matrix containing the unreduced, or base, survival probabilities.

Essentially, we populate the S_{act} in matrix $\hat{\mathbf{S}}_\mathbf{t}$ or the survival probabilities for each age group a for each county c for each time period t using Equation 3.We calculate $\hat{CCR}_{act}$ using the percentage of the population at-risk to SLR as the h value. This produces a second set of survival probabilities that are reduced based on SLR and we populate a new leslimatrix ($\hat{\mathbf{S}}_\mathbf{t}$) with these new reduced survival probailities ($\hat{S_{act}}$). The difference between the projected population using $\hat{\mathbf{S}}_\mathbf{t}$ and using $\mathbf{S_t}$ ($\mathbf{D_t} = \hat{\mathbf{S}}_\mathbf{t}\mathbf{P_t} - \mathbf{S_t}\mathbf{P_t}$ with $\mathbf{D_t}$ as the displaced population) represents the anticipated number of people who are inundated by SLR and thus likely to migrate. We then multiply the migration matrix $\mathbf{M_t}$ by the displaced population to produce the number of people migrating between counties by age-sex group. Finally, we simply add these migrants back into the projected population produced with $\hat{\mathbf{S}}_\mathbf{t}$.

4 Results

References

Arenstam Gibbons, S. J. & Nicholls, R. J. (2006), 'Island abandonment and sea-level rise: An historical analog from the Chesapeake Bay, USA', *Global Environmental Change* **16**(1), 40–47.

Bailey, C., Gramling, R. & Laska, S. B. (2014), Complexities of Resilience: Adaptation and Change within Human Communities of Coastal Louisiana, *in* J. W. Day, G. P. Kemp, A. M. Freeman & D. P. Muth, eds, 'Perspectives on the Restoration of the Mississippi Delta: The Once and Future Delta', Estuaries of the World, Springer Netherlands, Dordrecht, pp. 125–140.

Bamber, J. L., Oppenheimer, M., Kopp, R. E., Aspinall, W. P. & Cooke, R. M. (2019), 'Ice sheet contributions to future sea-level rise from structured expert judgment', *Proceedings of the National Academy of Sciences* **116**(23), 11195–11200.
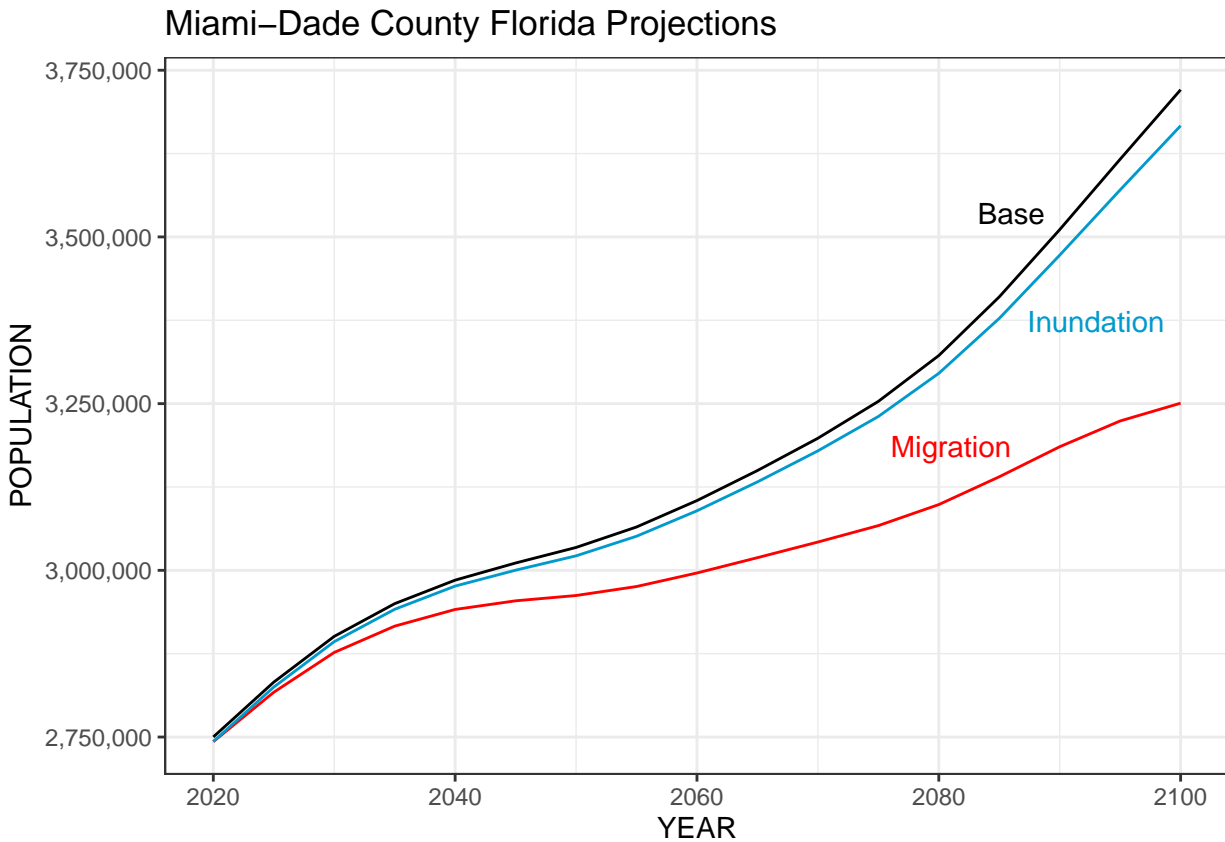
Figure 4: **A comparison of base, inundation, and migration projections for Miami-Dade County in Florida over the next century** The inundation model projection refines the previous base model as it accounts for sea-level rise displacement. Including demographic processes further sophisticates the migration projection. In the case of Miami-Dade County, the migration model reflects both those who emigrated and their lost progeny, indicating a substantial decline in the projected population for 2100.

Black, R., Bennett, S. R. G., Thomas, S. M. & Beddington, J. R. (2011), 'Climate change: Migration as adaptation', *Nature* **478**, 447–449.

Chen, C. & Liu, L.-M. (1993), 'Joint estimation of model parameters and outlier effects in time series', *Journal of the American Statistical Association* **88**(421), 284–297.

Chen, J. & Mueller, V. (2018), 'Coastal climate change, soil salinity and human migration in bangladesh', *Nature Climate Change* **8**(11), 981–985.

Clark, W. A. & Maas, R. (2015), 'Interpreting migration through the prism of reasons for moves', *Population, Space and Place* **21**(1), 54–67.

Colby, S. L. & Ortman, J. M. (2017), 'Projections of the size and composition of the us population: 2014 to 2060: Population estimates and projections'.

Curtis, K. J. & Schneider, A. (2011), 'Understanding the demographic implications of climate change: Estimates of localized population predictions under future scenarios of sea-level rise', *Population and Environment* **33**(1), 28–54.

Davis, K. F., Battachan, A., D'Odorico, P. & Suweis, S. (2018), 'A universal model for predicting human migration under climate change: Examining future sea level rise in Bangladesh', *Environmental Research Letters* **13**.

De Longueville, F., Zhu, Y. & Henry, S. (2019), 'Direct and indirect impacts of environmental factors on migration in burkina faso: application of structural equation modelling', *Population and Environment* **40**(4), 456–479.

DeWaard, J., Curtis, K. J. & Fussell, E. (2016), 'Population recovery in new orleans after hurricane katrina: exploring the potential role of stage migration in migration systems', *Population and environment* **37**(4), 449–463.

DeWaard, J., Hauer, M., Fussell, E., Curtis, K. J., Whitaker, S. D., McConnell, K., Price, K., Egan-Robertson, D., Soto, M. & Castro, C. A. (2021), 'User beware: Concerning findings from the post 2011–2012 us internal revenue service migration data', *Population Research and Policy Review* pp. 1–12.

Dieleman, F. M. (2001), 'Modelling residential mobility; a review of recent trends in research', *Journal of housing and the built environment* **16**(3-4), 249–265.

Döös, B. R. (1997), 'Can large-scale environmental migrations be predicted?', *Global Environmental Change* **7**(1), 41–61.

Engels, R. A. & Healy, M. K. (1981), 'Measuring interstate migration flows: an origin—destination network based on internal revenue service records', *Environment and Planning A* **13**(11), 1345–1360.

Franklin, R. S. & Plane, D. A. (2006), 'Pandora's box: The potential and peril of migration data from the american community survey', *International Regional Science Review* **29**(3), 231–246.

Fussell, E., Curtis, K. J. & DeWaard, J. (2014), 'Recovery migration to the City of New Orleans after Hurricane Katrina: A migration systems approach', *Population and Environment* **35**(3), 305–322.

Gross, E. (2005), Internal revenue service area-to-area migration data: Strengths, limitations, and current trends, *in* 'Proceedings of the Section on Government Statistics', p. 2005.

Hauer, M. & Byars, J. (2019), 'Irs county-to-county migration data, 1990–2010', *Demographic Research* **40**, 1153–1166.

Hauer, M. E. (2017), 'Migration induced by sea-level rise could reshape the US population landscape', *Nature Climate Change* **7**(5), 321–325.

Hauer, M. E., Evans, J. M. & Mishra, D. R. (2016), 'Millions projected to be at risk from sea-level rise in the continental United States', *Nature Climate Change* **6**(7), 691–695.

Hauer, M. E., Fussell, E., Mueller, V., Burkett, M., Call, M., Abel, K., McLeman, R. & Wrathall, D. (2020), 'Sea-level rise and human migration', *Nature Reviews Earth & Environment* **1**(1), 28–39.

Hino, M., Field, C. B. & Mach, K. J. (2017), 'Managed retreat as a response to natural hazard risk', *Nature Climate Change* **7**(5), 364–370.

Hori, M., Schafer, M. J. & Bowman, D. J. (2009), 'Displacement Dynamics in Southern Louisiana After Hurricanes Katrina and Rita', *Population Research and Policy Review* **28**(1), 45–65.

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E. & Yasmeen, F. (2019), *forecast: Forecasting functions for time series and linear models*. R package version 8.10.
URL: *http://pkg.robjhyndman.com/forecast*

Hyndman, R. J. & Khandakar, Y. (2008), 'Automatic time series forecasting: the forecast package for R', *Journal of Statistical Software* **26**(3), 1–22.
URL: *http://www.jstatsoft.org/article/view/v027i03*

Hyndman, R., Koehler, A. B., Ord, J. K. & Snyder, R. D. (2008), *Forecasting with exponential smoothing: the state space approach*, Springer Science & Business Media.

Jevrejeva, S., Moore, J. C. & Grinsted, A. (2012), 'Sea level projections to AD2500 with a new generation of climate change scenarios', *Global and Planetary Change* **80**, 14–20.

Kniveton, D. (2017), 'Sea-level-rise impacts: Questioning inevitable migration', *Nature Climate Change* **7**(8), 548–549.

Kulp, S. A. & Strauss, B. H. (2019), 'New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding', *Nature communications* **10**(1), 1–12.

Laurice Jamero, M., Onuki, M., Esteban, M., Billones-Sensano, X. K., Tan, N., Nellas, A., Takagi, H., Thao, N. D. & Valenzuela, V. P. (2017), 'Small-island communities in the Philippines prefer local measures to relocation in response to sea-level rise', *Nature Climate Change* **7**, 581.

López-de-Lacalle, J. (2019), *Tsoutliers: Detection of Outliers in Time Series*. R package version 0.6-8.

Lutz, W. & Muttarak, R. (2017), 'Forecasting societies' adaptive capacities through a demographic metabolism model', *Nature Climate Change* **7**(3), 177–184.

McGranahan, G., Balk, D. & Anderson, B. (2007), 'The rising tide: Assessing the risks of climate change and human settlements in low elevation coastal zones', *Environment and Urbanization* **19**(1), 17–37.

Mercer, J. H. (1978), 'West Antarctic ice sheet and CO 2 greenhouse effect: A threat of disaster', *Nature* **271**(5643), 321.

Molloy, R., Smith, C. L. & Wozniak, A. (2011), 'Internal migration in the united states', *Journal of Economic perspectives* **25**(3), 173–96.

Neumann, B., Vafeidis, A. T., Zimmermann, J. & Nicholls, R. J. (2015), 'Future Coastal Population Growth and Exposure to Sea-Level Rise and Coastal Flooding - A Global Assessment', *PLOS ONE* **10**(3), e0118571.

Nicholls, R. J. & Cazenave, A. (2010), 'Sea-Level Rise and Its Impact on Coastal Zones', *Science* **328**(5985), 1517–1520.

on Climate Change (IPCC), I. P. (2019), 'Special report on the ocean and cryosphere in a changing climate'.

Poulter, B. & Halpin, P. N. (2008), 'Raster modelling of coastal flooding from sea-level rise', *International Journal of Geographical Information Science* **22**(2), 167–182.

R Core Team (2019), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

Robinson, C., Dilkina, B. & Moreno-Cruz, J. (2020), 'Modeling migration patterns in the usa under sea level rise', *PloS one* **15**(1), e0227436.

Rogers, A. (1966), 'The multiregional matrix growth operator and the stable interregional age structure', *Demography* **3**(2), 537–544.

Ryder, N. B. (1985), The cohort as a concept in the study of social change, *in* 'Cohort analysis in social research', Springer, pp. 9–44.

Schneider, S. H. & Chen, R. S. (1980), 'Carbon Dioxide Warming and Coastline Flooding: Physical Factors and Climatic Impact', *Annual Review of Energy* **5**(1), 107–140.

Strauss, B. H., Kulp, S. & Levermann, A. (2015), 'Carbon choices determine US cities committed to futures below sea level', *Proceedings of the National Academy of Sciences* **112**(44), 13508–13513.

Sweet, W. V., Kopp, R. E., Weaver, C. P., Obeysekera, J., Horton, R. M., Thieler, E. R. & Zervas, C. (2017), 'Global and regional sea level rise scenarios for the United States'.

Trincardi, F., Barbanti, A., Bastianini, M., Benetazzo, A., Cavaleri, L., Chiggiato, J., Papa, A., Pomaro, A., Sclavo, M., Tosi, L. & Umgiesser, G. (2016), 'The 1966 Flooding of Venice: WHAT TIME TAUGHT US FOR THE FUTURE', *Oceanography* **29**(4), 178–186.

U.S. Fish and Wildlife Service (2012), 'National wetlands inventory website. u.s. department of the interior, fish and wildlife service, washington, d.c. http://www.fws.gov/wetlands/'.

Wilmoth, J., Zureick, S., Canudas-Romo, V., Inoue, M. & Sawyer, C. (2012), 'A flexible two-dimensional mortality model for use in indirect estimation', *Population studies* **66**(1), 1–28.

Yohe, G. W. & Schlesinger, M. E. (1998), 'Sea-level change: the expected economic cost of protection or abandonment in the united states', *Climatic Change* **38**(4), 447–472.