Migration in the 1930s: Beyond the Dust Bowl

Myron P. Gutmann,
Department of History and Institute of Behavioral Science, University of Colorado Boulder

Daniel Brown,
School of Natural Resources and Environment, University of Michigan

Angela R. Cunningham,
Department of Geography, University of Colorado Boulder

James Dykes,
Institute of Behavioral Science, University of Colorado Boulder

Susan Hautaniemi Leonard,
ICPSR, University of Michigan

Jani Little,
Institute of Behavioral Science, University of Colorado Boulder

Jeremy Mikecz,
Department of History, University of California, Davis Institute of Behavioral Science, University of Colorado Boulder

Paul W. Rhode,
Department of Economics, University of Michigan

Seth Spielman, and
Department of Geography and Institute of Behavioral Science, University of Colorado Boulder

Kenneth M. Sylvester
University of Michigan Flint

Abstract

This paper analyzes in detail the role of environmental and economic shocks in the migration of the 1930s. The 1940 U.S. Census of Population asked every inhabitant where they lived five years earlier, a unique source for understanding migration flows and networks. Earlier research documented migrant origins and destinations, but we will show how short term and annual weather conditions at sending locations in the 1930s explain those flows, and how they operated through agricultural success. Beyond demographic data, we use data about temperature and precipitation, plus data about agricultural production from the agricultural census. The widely known migration literature for the 1930s describes an era of relatively low migration, with much of the migration that did occur outward from the Dust Bowl region and the cotton South. Our work

about the complete U.S. will provide a fuller examination of migration in this socially and economically important era.

Introduction

People on the move area an enduring image of the U.S. in the 1930s, from photographs(Agee and Evans 1941), literature (Steinbeck 1939), and history (Gregory 1989, Egan 2006). People *did* move in the 1930s, spurred by the economic difficulties of the Great Depression, by heat and drought and by a multitude of other pressures. The scale of migration in the 1930s is visible in Figure 1, which shows the rate of out-migration from counties between 1935 and 1940, and in Table 1, which shows outmigration rates for states, divided into moves across county boundaries within states, and those that crossed state boundaries[1]. Figure 1 also outlines the area usually recognized as the extent of the dust storm activity in the 1930s and 1940, which provides a sense of ways that weather and agricultural stress acted on the lives of U.S. residents in this era.[2]

Despite the lore of the Dust Bowl and the "Great Migration" from the South to the North, the volume of internal U.S. state-to-state migration was not all that great from the early 20th century until after World War II (U.S. Bureau of the Census 1946). Modest interstate migration rates belie a continuing mobility made up of streams of people moving relatively short distances from one type of community to another (U.S. Bureau of the Census 1946, Bogue, Shryock et al. 1957, Hall and Ruggles 2004, Ferrie 2006), and as Figure 1 shows, areas with substantial out-migration were not limited to the Dust Bowl or the deep South. Nonetheless, migration was a frequent subject of discussion, important enough that the U.S. Census tracked migration for the first time in 1940, asking exactly where every person enumerated had lived five years earlier (U.S. Bureau of the Census 2002). With this wealth of data, research about migration in the 1930s has explored many questions but left even more unanswered (Lively and Taeuber 1939, U.S. Bureau of the Census 1946, Bogue and Hagood 1953, Bogue, Shryock et al. 1957, Tolnay, White et al. 2005, White 2005, White, Crowder et al. 2005, Fishback, Horrace et al. 2006, McLeman and Smit 2006, Boustan, Fishback et al. 2010, Boustan, Kahn et al. 2012, Hornbeck 2012, Long and Siu 2013, McLeman 2013).

How much do we really know about the causes of migration in the 1930s? To what extent do we know that the big forces that are supposed to have been important actually were significant? This is an important question, and the availability of Census data for 1940 that shows where virtually every American lived in 1935 makes it possible to think about the factors that drove migration at a very refined scale. Moreover, the availability of data about weather, agriculture, and employment -- among other factors -- with comparable refinement make it possible to add real insight to our understanding. Unlike previously utilized sample or tabulated population data they show the migration experience of almost every American resident at the county level, and tell us where they went (although for this article we are only

---

[1]These figures only include adults aged 20 years and older and not living in group quarter in 1940, with a known county of origin based on our county definitions. See Appendix C.
[2]These extents are digitized from Cunfer (2005).

interested in whether they moved across county or state lines). Managing these detailed and complex data -- there were more than 130 million American inhabitants in 1940 -- created significant challenges, and describing how we overcame those challenges constitutes a significant part of this article. But our findings are important as well.

To a great extent, the conventional story of the 1930s is right but too limiting: people left areas where the weather was challenging for agriculture in the mid-1930s, but that experience was not limited to the Dust Bowl region. The weather was hot and dry in a much larger part of the U.S., and migrants escaped those areas as well. Moreover, much more was happening. There were regional processes that intensified the agriculture and weather effects, and broader economic processes that are predictable. Our research confirms much that we should have known all along, but with much more detailed data that give us confidence that we understand what was happening.

Theoretical background

A large theoretical literature (Ravenstein 1885, Ravenstein 1889, Roy 1951, Lee 1966, Massey 1990, Massey, Arango et al. 1993, Greenwood 1997, Massey, Arango et al. 1998) has sought to explain migration. Though much of it focuses on international migration, its core elements have also been applied to internal migration.

Neoclassical economics provides the dominant perspective, informed by contributions from the New Economics of Migration, by Massey's notion of Cumulative Causation theory and by the descriptive richness of Migration Systems Theory (Massey 1990, Massey, Arango et al. 1993, Bakewell 2013, Fussell, Curtis et al. 2014). In these approaches, a person's likelihood of migrating is a function of his or her individual characteristics but also of the characteristics of their places of origin and destination, including the distribution of income, land, and human capital; the organization of agriculture and industry; public policy; and cultural frameworks, reflected in local ethnic, religious and racialconditions(Massey, Arango et al. 1993, Fishback, Horrace et al. 2006). Migration is also shaped by the structure of social networks. Because these processes are not independent of spatial context, we see migration not merely as people moving in unconnected and location-free contexts but as interactions between people and locations followingspecific pathways.

Most research sees migration as a tension between pushes and pulls, and theorizes that migration serves as a mechanism to restore equilibrium between competing forces. Disaster-related demographic theory builds on these patterns of movement to identify the migratory systems that existed prior to the disaster and to gauge whether a given shock transforms the pre-existing migration system (Black, Adger et al. 2011, McLeman 2013, Fussell, Curtis et al. 2014). The disaster literature also adds the concept of vulnerability to the determinants of migration (Adger 2006, McLeman and Smit 2006, McLeman 2013). The vulnerability paradigm focuses on the exposure of people to stress over time, prior to the period of crisis, based on the condition of the economic and ecological systems they inhabit. Assessments of vulnerability have focused on both short-term moves and long-term reorganizations of human-environment systems (Berkes, Carl et al. 1998, Adger 2006, Black, Arnell et al. 2013), asking whether a specific "demographic signature of disaster" exists (DeWaard,

Curtis et al. 2014). In the U.S., the Dust Bowl story has driven research about environmental migration in the 1930s, yet drought and land degradation were not limited to the southern plains. Extreme heat also led to agricultural stress elsewhere (Gregory 2005, Olmstead and Rhode 2008, Giesen 2011, McEwan, Pederson et al. 2014).

Our analysis is not limitedto environmentally-driven agricultural shocks. The U.S. was in the midst of a depression in the 1930s, with high unemployment, low wages, and a weak economic recovery. We also ask whether people left places with comparatively poor economic conditions, or stayed in places with relatively good economic conditions, and did so in a way that reveals the characteristics of places as drivers of migration. The United States still had a significant rural population in the 1930s, with more than 40% of its inhabitants living in places of less than 2,500 residents in both 1930 and 1940 (U.S. Bureau of the Census 2012). The West, South and Midwest were significantly more rural, while the Northeast was significantly less rural. Farming was still a major industry in many parts of the country, and weather had a major impact on agriculture, and consequently, the rural economy. The limits of urbanization are visible in Figure 2, which shows where all the counties that had an urban population of 50,000 or more were located. The map is sparse.

For the analysis reported here, we focus on the attributes of counties rather than the attributes of individuals. We are interested in how likely it was for someone living in a given U.S. county in 1935 to migrate to another county by the time of the 1940 census, and the attributes of counties that made them more or less likely to send their inhabitants elsewhere by 1940. We realize that this is a substantial simplification, because the migration literature is heavily focused on the idea first raised by Roy (1951) that migrants self-select for upward mobility. Moreover, Borjas(1987) expanded selection theory by arguing that it is the more highly skilled who migrate, and Kanbur and Rapoport(2005) argue that selectivity by education is key. While we recognize that individual attributes are as important as those of context in determining who migrates, from which communities, and where they go, our approach is an important first step in understanding how migration operated in this era, and how county characteristics shaped out-migration flows.

Understanding relative levels of out-migration

The main questions we explore revolve around the role of the environment and economy in encouraging or discouraging out-migration from counties between 1935 and 1940. We discuss the data that we rely on, and questions raised by the nature of those data, in a later section. We begin by discussing the factors that may have led one county to experience more out-migration than another. Our focus is mostly on processes that are important for the less urbanized parts of the United States, where natural phenomena, such as precipitation and temperature, may have had a strong effect, either acting on their own or acting through agriculture, although we also include measures related to employment by industry and unemployment, which we hypothesize play a role in determining migration even in urban areas. While other researchers have examined the role of New Deal support programs in explaining migration (Fishback, Horrace et al. 2006), our preliminary analysis suggested that they were less important than other factors, and they are not included in our statistical models.

Given the severe drought of the mid-1930s, which has been described recently as the worst drought of the last millennium (Cook, Seager et al. 2014), we begin with an examination of the role of weather in influencing levels of out-migration from U.S. counties. We measure weather by looking at annual total precipitation for various time periods, and average daily maximum temperature as a percentage of a longer-term (1920-1940) average. The drought was worst in1934 (and to some extent similarly severe in 1933 and 1935), so we have created relative weather measures for 1934, 1933-1935, and 1933-1939. We show the scale of the drought for these time periods in Figures 3 and 4. Both temperature (higher than normal) and precipitation (lower than normal) diverged most significantly from expected patterns in 1934, somewhat less so in 1933-35, and came closer to normal for the 7-year period from 1933 to 1939. What is also clear is that the spatial patterns were rather different, with the highest sustained temperatures in the front range of the Rocky Mountains and the northwestern Great Plains (see especially figure 4.b.), and the lowest sustained precipitation in the classic Dust Bowl areas in Oklahoma, Texas, Kansas, New Mexico and Colorado, plus parts of the northern Plains (figure 3.c), and to a lesser extent in the intermountain west (California, Nevada, Utah, Wyoming, and Idaho - figure 3.b.).

We utilizea variety of ways to measure changes in agriculture during the 1930s, making use of data from the 1930, 1935, and 1940 censuses of agriculture, which represent agricultural results in 1929, 1934 and 1939, respectively (Haines, Fishback et al. 2014). The agricultural census includes one direct measure, the percent of land with failed crops. We display this in Figure 5. The two panels of the figure show the consequences of the most severe weather, with significant failure levels throughout the central U.S. in 1934, and in areas of Kansas, Colorado, Arizona, New Mexico, and South Dakota (and spots elsewhere) in 1939.

In order to attempt to find new ways to gauge the impact of the weather on agricultural production, we developed three other measures. The most ambitious of our measures estimates the percent change in crop production from 1929 to 1934 and from 1929 to 1939, for each county's three largest crops as indicated in the 1930 census. The methods we used to derive these estimates are described in Appendix A. This measure captures the overall fall-off in production, due to farmers not having the resources to plant or believing that the crop would fail anyway. As we show in figure 6, a very large portion of the U.S. experienced major production short-falls in 1934, something that still had not been reversed in 1939. The spatial pattern changes between 1934 and 1939, with a greater fall-off in production in the later period in Georgia, Mississippi, Alabama, and Kentucky, plus New England and New York, and less in the corn belt states (Illinois, Iowa, northern Missouri), and the upper Midwest (Wisconsin, Minnesota).

We also experimented with two other measures. In one, we made the same calculation we did for the three largest crops, and instead focused only on corn, wheat, and cotton. In a second, we attempted to gauge the impact of severe weather and poor agricultural conditions on livestock by estimating livestock inventories in 1934 and 1939 as a percentage of what they were in 1929 (figure 7).[3] The results shown in figure 7 are interesting because they show that in most parts of the U.S. livestock had increased between 1929 and 1934 (figure 7.a), with some exceptions in areas of the central U.S. with the worst weather. The situation

worsened between 1934 and 1939 (figure 7.b), but livestock nonetheless continued to increase in numbers in most of the country.

Although we hypothesize that changes in agriculture drove most of the out-migration in the 1930s, other economic factors played a role. One of these is unemployment, which was enumerated in a special census of "partial employment, unemployment, and occupations," published in 1937 (Biggers and United States 1938)[4]. Figure 8 displays the spatial distribution of unemployment, with the highest levels of unemployment in the deep South, Appalachia, in the northeast, along the northern tier of the U.S., and in Utah and New Mexico. As we might predict this does not appear to align easily with either weather or agriculture, We believe the economy played a broader role, so we have also examined employment in various industries as potential indicators that the economy was capable of doing better or worse in different areas of the U.S., with an impact on migration.

Our hypotheses are simple and straightforward. We expect that rural counties that experienced severe weather, poor agricultural results, and high unemployment should have had higher levels of out-migration than those that did not, and that urban counties and counties with higher levels of employment in manufacturing and retail sales should have had lower levels of out-migration, all other things being equal.

Data and Methods

The variety of datanecessary for this analysis, originally gathered at different spatial and temporal scales, were transformed in order to produce a data set that may be analyzed at the county level. Our dependent variable is the rate of out-migration from each county of the U.S. between 1935 and 1940, based on the data in the 1940 U.S. Census of Population. The independent variables draw on environmental, agricultural, and economic data. The methods we use include both descriptive and multivariate approaches.

The list of counties of the United States has changed over time, even during as brief a period as the decade from 1930 to 1940. For this analysis we began with the 1940 list of counties and their geography, and modified that geography to take into account changes in the list of counties and the ways that we and others have aggregated counties to optimize analysis. First, there are counties that existed in 1930 but not in 1940 (or vice-versa). Campbell and Milton Counties were merged into Fulton County, Georgia between 1930 and 1940; we combined these three counties into one. In another case, we aggregated spatial units in

[3]We chose cattle and swine for our analysis of livestock because these data are most suitable for comparison across census years. The data for other stock animals made comparison more difficult. Work animals such as horses, donkeys, and mules had been in steady decline for some time due to the mechanization of farms; thus, they are not reliable indicators of the health of farms. Livestock units are calculated using ratios developed by the FAO, which counts each cattle as one unit and each swine as one-quarter of a unit(FAO. 2011). Livestock numbers are inflated for the census year of 1935 in comparison to the other two years (1930 and 1940). Due to the collection of date earlier in the year and slightly different questions posed to farmers for 1935, this year's data does not include those animals that died or were slaughtered between Jan. 1 and Apr. 1 of that year. Thus, the top map above shows a greater increase (or lesser decrease) in livestock than that actually experienced between 1930 and 1935.
[4]The data are available in digital format in Haines and Inter-university Consortium for Political and Social Research (2010). The number of unemployed is the number of persons 15-74 years of age who reported being totally unemployed but wanting and able to work on the Unemployment Report Card. This does not include emergency workers. Farmers and farm laborers were included; cards were given to each household, but there does not seem to have been any follow-up to ensure compliance. In order to estimate an unemployment rate in 1937, we estimated a total working age population for each county by summing the population aged 15 and over in 1930 and 1940, and interpolating to a 1937 population (70% of the difference).

Virginia in order to combine independent cities with their surrounding counties.[5] In a third modification, we needed to merge counties from a single metropolitan area where confusing naming practices made it impossible to distinguish separate counties (New York City's five counties and Saint Louis City and Saint Louis County, Missouri). These modifications are summarized in Appendix B. The resulting dataset contains 3,069 counties from the contiguous 48 states that existed in 1940. In the cases where we combined counties and the data were counts, we summed the counts across all geographic units. In the case of data that were rates or averages, we calculated means, which were spatially weighted when appropriate.

Census Data

The 1940 U.S. Census of Population full count data have a wealth of information about individuals, including demographic, social and cultural, economic, and location data (Ruggles, Alexander et al. 2010), and constitute the primary data source for our analysis. In the 1940 census, respondents were asked to provide their county of residence and that of all their household members in 1935. At the time we undertook this analysis, the full-count 1940 data were available in two forms, and we made use of both of them. The University of Minnesota has released a preliminary version of the 1940 data in IPUMS coded format, which includes every person in the U.S. but not every variable from the census. The IPUMS version of the 1940 dataset includes standardized variables for a person's 1935 state of residence and the scale of their residential movements since 1935, delineating those who moved within or between counties, states, and countries, but the coded data do not yet contain a variable for *which* county a person lived in in 1935. That information is available from a restricted release of a version of the 1940 data that does include the detailed response text for every person. We have merged these data sets, and coded -- to the extent possible -- origin county from the 1935 textual residence variable, in addition to recoding 1940 places of residence that were erroneously coded in the IPUMS data (for example coding that conflated Brown and Boone Counties in Indiana, and Richmond City and Richmond County in Virginia). That turned out to be a challenging task, which we describe in Appendix C.

Table 1 presents a subset of the data, tabulated by state of residence in 1935. This allows us to show the level of out-migration from counties within states, for migrants who stayed within their state of residence ("intercounty move") and those who left their state of residence ("interstate move"). We exclude people under age 20 in 1940, people living in group quarters in 1940, those living outside the U.S. in 1935, and those whose residence in 1935 is unknown. For the contiguous U.S. as a whole, using our data set, 5.5 percent of the population moved across county boundaries (but stayed in-state)[6], and 4.1 percent moved across state lines, a total of 9.6 percent. Variation from state to state is substantial, as we would expect.

---

[5]Following the example of Fishback et al. (2003) we combined independent cities with their surrounding counties to ensure consistent and useful data. In doing so, we assigned new fips codes to these aggregated units.
[6]We treat moves within New York City but across borough lines as being within the same county, despite the fact that the five New York City boroughs are each a separate county.

Weather and climate data

We derived our climatological variables from datasets developed by the PRISM Climate Group at Oregon State University. ThesePRISM data are 4km-grid rasters of temperature and precipitation, modeled at a monthly resolution, stretching back to the late nineteenth century and covering the contiguous United States. Assuming that physiographic factors such as elevation and aspect have influenced local climate in a similar way in both the more recent and more distant past, the PRISM group incorporated a 'climate fingerprint' from thirty year normals for 1971-2000 into their expert system to fill in the gaps between scattered weather monitoring stations for earlier years. Further details on PRISM's method of topographically-informed interpolation may be found in the group's publications and web documentation(Daly, Gibson et al. 2002, Daly, Halbleib et al. 2008). For our analysis , weaggregated the PRISM data to counties by calculating zonal statistics, using 1940 county boundaries from the National Historical Geographical Information System (NHGIS) (Minnesota Population Center, 2011), and aligned with the PRISM rasters.[7] In order to do this we averaged the grid cell values within each county to calculate the maximum and minimum temperatures and the total precipitation.[8]

Urban and Regional Status

We recognize that there is regional variation in how the demographic, economic, agricultural and environmental characteristics affected the migration patterns of the 1930s. In order to understand this variation, we decided to classify counties into broader groups, using agro-ecological categories defined by the United States Forest Service in 1997(Bailey 1997).[9] Bailey's report divides the United States into a hierarchical system of ecoregions. These levels of ecoregions include: 4 climatic domains, 15 divisions, and 63 provinces. Each of these regional classifications was developed at an increasing level of climatic precision. The division level has the appropriate amount of resolution for our analysis, with eleven divisions for the continental U.S., and an additional nine mountain regime sub-divisions. To simplify our analysis, we have joined these mountain regimes with their climatic lowland counterpart (i.e. "Subtropical Regime Mountains" is joined with "Subtropical Division"). (See Figure 9) We also merged division 4 (Savannah) with division 3 (Subtropical), both of which represent parts of the far southeastern U.S., because division 4 only contains four counties at the southern tip of the state of Florida.

Along with regional variation, we assume that the dominance of agricultural factors should be less in areas where agriculture plays a smaller role in the economy. One way to do that is to identify urban areas (Figure 2). Our starting point for this was a set of historical classifications of metropolitan statistical areas available at the Minnesota Population Center's IPUMS website ("County Composition," in (Ruggles, Alexander et al. 2010)). We

---

[7]For counties that were too small to contain even a single 4km by 4km raster cell and for which zonal statistics could not be calculated, temperature and precipitation values were obtained by creating a centroid for each county and intersecting that centroid with the raster data.

[8]We employed a combination of ArcGIS's arcpy, the statistical package R, and Beyer's Geospatial Modeling Environment (GME-http://www.spatialecology.com/gme/index.htm), software that acts as a go-between for the two programming languages. For instance, GME allowed us to process our data county by county, thus circumventing some of the known issues with the way ArcGIS performs its rasterization and statistical processes.

[9]This report is a revised and updated version of a map published in 1981 by the United States Fish and Wildlife Service (Bailey and Cushwa 1981).

created an urban scale variable from the metropolitan area data for 1930. This variable has three values, one for rural/non-metro, one for urban/non-metro (all non-metro counties with an urban population greater than 50,000 in 1930) and a third for urban/metro, which are counties classified as metropolitan in 1930, which we further aggregated to two values (rural vs. urban) for our analysis.

Methods

At this stage in our research, most of our data analysis has been descriptive and visual, using maps to show the spatial patterns of migration alongside other spatial patterns -- weather, agricultural production and failure, unemployment, farm population, and New Deal program investments. The results, as we will show, are striking, within the usual constraints that it is very difficult to really understand correlation or causation visually. What looks like a striking relationship might just be something that catches our eye.

In order to provide a more conventional statistical presentation, we have taken a variety of directions. We begin with OLS regression models with county as the unit of analysis and the natural log of the number of adults (aged 15 or over in 1935) who were resident in the county in 1935 and moved outside the county by 1940 as the dependent variable. All models include the log of the estimated base 1935 population (those who moved plus those who did not) among the explanatory variables. Each model presented explores different combinations of potential explanatory variables. Because there is reason to believe that the relationships of our explanatory variables to our dependent variable vary in different parts of the U.S., wealso explore the role of ecological divisions in the models. We first consider ecological divisions as a set of binary independent variables. In these OLS models, we find significant heteroskedasticity and a handful of outliers, so we also estimate the models using robust regression methods. While the robust regression models we use might have addressed these issues, in our models it did not do so, as indicated by the Breusch-Pagan p-values in Tables 4 and 5.

When residuals from adjacent counties are correlated, i.e., spatial autocorrelation, standard OLS estimates of standard errors are artificially small and the goodness of fit measures and chances of finding statistical significance are inflated. Spatial autocorrelation can be measured with Moran's I, and our results show that there is significant spatial autocorrelation in the standard OLS models. It is not possible to calculate Moran's I for the Robust Regression techniques we use, but we suspect that there are spatial effects present.

The third approach we use is to consider ecological divisions as characteristics that interact with all the other variables, and use a method that allows those interactions to uncover how the effects of the other independent variables may depend upon ecological context. Following Anselin(2007) we refer to the latter as "regime" models because each ecodivision constitutes a separate spatial regime with its own set of regression coefficients. Despite the advantages of the regime models, treatingour ecological divisions as different spatial regimes still does not account fully for the spatial effects in the system. This requires a spatial regression approach to reveal the nature and extent of the effects from neighboring counties. Anselin's(2007) decision tree is the accepted method for choosing between alternative specifications of spatial effects, leading us to choose the "spatial error" model (as

opposed to a spatial lag or combined model), because it outperformed other spatial regression and OLS models in our tests. A spatial error model specifies that the unexplained outmigration in a focal county is directly affected by the residual outmigration in the surrounding counties. Regression analyses were largely performed in R, and work that explored patterns in the residuals for different models was conducted in GeoDa(Anselin, Syabri et al. 2006).

Results

We begin our discussion of results with a visual presentation of adult out-migration by county of residence in 1935 (the results are tabulated by state in Table 1). These results are presented in Figure 1. These are data where the numerator is the number of people known to live in a specific county in 1935 and who had left by 1940. The denominator for these computations is the sum of all adults (over age 15 in 1935) whose residence was known in 1935, and was within the U.S. All individuals living in group quarters in 1940 or living outside the United States in 1935 are excluded.

People who lived in the western U.S. in 1935 appear more likely to have moved in the next five years than those in the east (and especially New England and the Middle Atlantic states), with the greatest likelihood of out-migration in two north-south bands, one from western Texas and New Mexico north to the Dakotas and Montana (roughly what we consider the Great Plains), and the second away from the coasts in states on the western edge of the U.S., especially Arizona, Nevada, California, Idaho, Oregon, and Washington.

What causes these patterns? Certainly, looking back to figures 3 (temperature in 1934) and 6 (precipitation in 1934), we see possible connections. There is a lot of out-migration from those places that were hottest and driest in the worst year of the drought. Is this a real relationship? And how does it work? Is it a direct connection, or one that works through agriculture? And what is the relationship between migration and other factors, such as unemployment, the extent to which the county's population is engaged in farming, or New Deal public programs?

It is possible to quantify the relationships we see with a multivariate statistical analysis. We estimated a series of multivariate OLS models, with various combinations of independent variables. We estimate multivariate models for two families of regressions, one ("Crop Failure") where the main independent variable is the level of crop failure, and the other ("Climate"), where the main independent variables are precipitation and temperature. We began by estimating univariate regressions between the amount of migration and each of the potential independent variables. For the crop failure model we achieved stronger associations by dividing the values into four categories (as opposed to a single continuous variable), including 0% failure as the reference category. For the climate models the data for 1934 as a percentage of 1920-1940 averages was most predictive of migration. We also include the weighted percent change in production of a county's three top crops as reported in 1930 and 1940 in these models. This measure was more predictive of migration than the other measures, such as the change from 1930 to 1935, or 1935 to 1940, or changes in livestock. In both the crop failure and climate models we include per capita retail and

manufacturing employment in 1930, the percent of the working age population unemployed in 1937, and a binary category delineating whether a county is urban (having an urban population in excess of 50,000). For the ecodivisions, we chose the warm continental region in the northeastern United States as the reference category; ecodivisions were added before the other independent variables (except for the log of the total population) when incorporated into models as a main effect.

We hypothesize that there are attributes of counties that encourage or discourage migration, and that these are generally linked to their impact on livelihood. At the core of our analysis are agriculture and the climate forces that shape it, such as drought. We hypothesize that poor agricultural conditions lead to greater out-migration from a county, and that better agricultural conditions are associated with less out-migration. The same holds for unemployment, with greater unemployment associated with higher out-migration. On the other side, we hypothesize that economic activities outside of agriculture will be protective and associated with less out-migration. In that category we have employment in retail sales and manufacturing, plus the urban status of the county.

The first and most basic of these results are included in Table 2. We begin with a model that just includes the natural log of the baseline population (the denominator in a migration rate) on the right side of the equation. We then add the ecodivisions to the model to show regional effects. All regions (except the Hot Continental division) are significantly different in their level of migration from the Warm Continental reference category. Looking back to figures 3.A and 4.A, we see that the largest coefficients are in areas with the hottest and driest weather in 1934, with the complication that heat and drought did not overlap perfectly -- it was hot but not especially dry in the Pacific northwest, and dry but not especially hot in California.

In Table 3 we show three versions of a model in which the main independent variable is crop failure, divided into four categories (zero, 1 to 5%, 5 to 25% and 25% and over). We present two versions of the crop failure OLS model, one with and one without the ecodivisions, as well as a model using robust regression. We present the results as odds ratios. Overall, the OLS model with the ecodivision regions shows a better fit than the one without the ecodivisions, and the robust regression model has an r-squared that is still larger. The results in Table 3 largely confirm our assumptions about how environmental and economic stress contributed to migration flows. As expected, higher levels of crop failure led to more migration, as did higher unemployment. On the other hand, counties with more manufacturing employment and counties with large urban populations had slightly lower (and not always significantly different) levels of out-migration. The one result that does not necessarily confirm our assumptions is the percent of the population employed in retail, often suggested as an indicator of overall economic activity. Our results show that higher levels of employment in retail sales are associated with higher levels of out-migration from the county, perhaps because of the ubiquity of retail employment throughout the United States in 1930.

The U.S. ecodivisions are consistently significant in their impact on out-migration, although there is little difference between the two continental divisions in the northeast and northern

Midwest. What the different odds ratios show us is that there was more out-migration in some regions than others (generally in the west), even after we take into account the rest of the model. This is an issue that we will return to later.

In Table 4 we replicate the analysis from Table 3, with the main independent variables reflecting a combination of climate and agriculture. These variables are precipitation and temperature in 1934 (compared with 1920-1940), plus production of the three main crops in the county in 1939, as a percentage of the production of those crops in 1929. These Climate model results confirm what we saw with the Crop Failure models, with a slightly better overall model fit. In these models higher temperatures and lower precipitation, as well as lower agricultural production, are also associated with more out-migration. The other variables generally behave in the same way as they did in the crop failure models. One interesting finding is that the unemployment variable only becomes significant when ecodivision is included in the model.

We report the results of our spatial regime models in summary form in Table 5, which compares model diagnostics using the ecodivision regions in three ways: first as categorical independent variables, then as a spatial regime model, and finally as a spatial regime model with a spatial error term. Because the spatial regime models involve interactions between the ten ecodivisions and all the other variables, the coefficients are voluminous. We have chosen not to report them here, but rely below on residual maps to show the spatial characteristics of fit. The diagnostics in Table 5 show that by including all of the interactions between ecodivision and the other independent variables we are able to develop a model that explains virtually all of the variance in out-migration, based on an r-squared greater than 0.99. When we then add the spatial error term, we improve the fit still further (as indicated by the significantly reduced AIC), and we eliminate evidence of spatial autocorrelation (as indicated by an insignificant Moran's I).

When we introduce the full set of interactions in the regimes model we achieve a strong model fit, but reveal dramatic heterogeneity among the effects for each combination of independent variable values. We illustrate this heterogeneity in figures 10 and 11 by mapping model residuals. In figure 10 we map the residuals for four versions of our crop failure models, and in figures 11 we do this for the climate models: without ecodivisions, with ecodivisions included as categorical variables, with ecodivisions incorporated as regimes, and finally with both the ecodivision regimes and a spatial error term to explicitly correct for spatial autocorrelation. We use LISA (local indicators of spatial association) statistics, calculated in GeoDa using contiguity to define neighbor relationships, to better convey the patterns of geographical dependence in our data and our model fits. These four steps allow us to use increasingly effective means for understanding the role of spatial dependence in our data and models. In the maps, white counties are those where the model fits well. Blue counties are areas where the measured migration is significantly lower than predicted by the model, while red counties are ones where the measured migration is significantly higher than predicted by the model.

As we progress towards more nuanced appreciations of spatial effects, our overall model fit improves; the number of counties with measured migration significantly different from that

predicted by the models (the blue and red counties) decreases. This is a good outcome, confirming the statistical results in Table 6, where r-squared, AIC, and other diagnostics improve as we add spatial information. Some areas of poor model fit persist, however. An area in the southwest remains blue, indicating measured migration is lower than expected and suggesting that an unidentified factor may be providing protection against migration drivers, while red patches scattered across the south indicate that our variables are not accounting for all the drivers of outmigration. These maps also begin to show multi-county patterns that are more localized than what is captured by the ecodivisions. Along the Gulf Coast of Texas, for instance, Harris County, the location of Houston, appears in blue bordering counties in red in the final maps, indicating a place of lower than expected outmigration bordered by counties of higher than expected outmigration, perhaps suggestive of an especially strong rural to urban pull due to the rapid growth of the oil industry in Harris County in the 1930s.

The last thing we note in discussing these results is the visual relationship between the residuals and the location of the various periods of dust activity in the 1930s and 1940. In our best models those areas appear to fit the model relatively well, suggesting that their behavior is well explained by our models, and may not be unusual when compared with other areas of high temperature and low precipitation.

Conclusion

We began this article by asking how much we really knew about the causes of migration in the 1930s. There was a relationship between weather and migration in the 1930s, and it operated beyond the borders of the Dust Bowl. In most of the U.S., people left places that were very hot or very dry, and stayed in places that were relatively cool and wet, although that was not true everywhere. Much of the migration-weather process worked through agriculture, but it didn't always operatein ways that can be generalized across the whole U.S. That's why the models that use temperature and precipitation appear to explain more of the variation in migration than most of the agricultural variables we could find or estimate.

The factors that determine levels of migration during an era of environmental and economic stress are both national and regional in scale. The U.S. is a large country with strong regional variations in climate, agriculture, and economy, which our analysis reveals. While temperature and precipitation had generalizable impacts on migration in the second half of the Depression, there were noticeable exceptions to the general pattern of how temperature and precipitation related to migration in various parts of the U.S. Moreover,while spatial error models captured unobserved spatial processes and reduced the number of counties with unexplained migration outcomes, significant patterns remain, noticeably in the form of counties with high migration but low crop failure in Georgia and Florida (shaded in red in Figure 10), and in counties with high migration and low failure in central Florida and along the Mississippi Delta in (shaded in red in Figure 10). These patterns suggest a wage decline mechanism, rather than an environmental push mechanism, so that people left areas where wages declined. In the Great Plains, heat and drought reduced production, lowering wages and leading to migration. In Georgia, Florida, and along the Mississippi delta, higher

production more than production failures drove down wages, which also led to out-migration.

In revealing both national and regional patterns, our work does not discredit the visceral conventional story of 1930s migration that's beautifully illustrated by Steinbeck, or Lange, or described historically by Worster and Gregory; rather, our analysis grounds the drama of what we know about specific cases within a wider context, the nuances of which may only be sketched with the refined and complex data we now have at our disposal. The coexistence of spatially dependent and general processes suggest that future work must continue to explore the nature of multiscalar interactions, and to treat the responses of individual migrants as influenced by places of origin and destination. Even as the work presented here anticipates individual-level analysis, it reminds us that the motivations migrants share, even during an era of widely shared environmental and economic suffering, remain influenced by history and the characteristics of place.

Acknowledgments

Appendix A

Deriving Variables about Changes in Agricultural Production for Each County's Three Largest Crops

Using agricultural census data for the years 1930, 1935, and 1940 (which respectively represented farming in 1929, 1934, and 1939) (Haines, Fishback et al. 2014), we have examined several different types of data as indicators of agricultural production during the period. These data include

- Percent of total cropland that failed,

- Change in production of major market crops (corn, cotton, and wheat),

- Change in production of major livestock (cattle and pigs). We chose cattle and pigs for our analysis of livestock because these data are most suitable for comparison across census years. The data for other stock animals made comparison more difficult. Work animals such as horses, donkeys, and mules had been in steady decline for some time due to the mechanization of farms; thus, they are not reliable indicators of the health of farms.

- An agricultural production composite index of the top three crops for each county.

The first three datasets are relatively self-explanatory. The fourth dataset, however, requires more explanation. Identifying the most important crops for each of the approximately 3,100 counties in the contiguous United States and then calculating the percent change in production from 1929 to 1939 required several steps. First, we identified the top three crops in each county by the acreage harvested in 1929 (Table A.1). The Agricultural Census data for the period contains acreage for nearly all *significant* crops, meaning any crop that appears in the top three for at least one county. The one major exception is fruit trees: census-takers only recorded the number of trees, rather than the amount of acreage. However, twenty years later, the 1950 Census recorded both acreage and number of trees. This allowed us to estimate the number of acres devoted to fruit trees in each county, assuming that the average number of trees per acre was stable from the 1920s to 1950.

Second, for each county, we identified the amount of production for each of the top three crops (whether recorded in bushels, bales, lbs. or tons), which we then used to calculate changes in production for the period. The data posed several problems in this effort. Census-takers of the period did not record any production values for some crops, most notably vegetables, for which enumerators noted farms reporting, acreage, and dollar values for each vegetable, but not the quantity of production (for example, bushels). Vegetables were only reported as one of the top three crops in 126 counties, making this gap less problematic. For other crops, the unit of measurement used to record production changed between 1930 and 1940. Most of the time, a simple conversion was sufficient to make the data comparable, an example being that one bushel of cherries weighs approximately 56 pounds, allowing us to convert production in bushels into production in pounds, or vice versa. In a few cases, the data required more complicated conversion calculations. Peanuts are a more complicated example, because they were recorded in bushels in 1930 and 1935, but in pounds in 1940. This posed a problem as different types of peanuts had significantly different rates of lbs./bushel. We worked around this by applying different conversion rates for each state depending on the dominant type of peanut found in that state.

Once we had accounted for all discrepancies in the data, we calculated the percent change in production for each crop across three time periods (1929-1934, 1935-1939, and 1929-1939). In addition, we created a composite index of the percent change of the three crops combined. We created this composite measure by calculating a weighted percent change. For example, in Sutter County, CA, barley, wheat, and hay represented 48.8%, 40.2%, and 11%, respectively, of the county's acreage devoted to the top three crops. We multiplied the percent change in production of each crop by the crop's relative size as a weight to create a composite weighted percent change figure.

Appendix B

Counties Combined in the Analysis

Original Counties	New Units (Following Fishback, Kantor, and Wallis (2003).
Campbell (13041), Fulton (13121), and Milton (13203) Counties, GA (for 1930 only; these counties were already merged by the 1940 census)	Fulton (13121)

Original Counties	New Units(FollowingFishback, Kantor, and Wallis (2003).
New York City: Bronx (36050), Kings (36470), New York (36610), Queens (36810), and Richmond (36850) counties	"New York City County" (36150)
St. Louis City (29510) and County (29189)	"St. Louis City and County" (29300)
Virginia Independent Cities and Surrounding Counties:	
Albemarle (51003) and Charlottesville City (51540)	"Albemarle (incl. Charlottesville)" (51200)
Allegheny (51005) and Clifton Forge City (51560)	"Allegheny (incl. Clifton Forge)" (51210)
Arlington (51013) and Alexandria City (51510)	"Arlington (incl. Alexandria City)" (51215)
Augusta (51015) and Staunton City (51790)	"Augusta (incl. Staunton)" (51220)
Campbell (51031) and Lynchburg City (51680)	"Campbell (incl. Lynchburg)" (51230)
Dinwiddie (51053) and Petersburg City (51730)	"Dinwiddie (incl. Petersburg City)" (51240)
Hampton (51630) and Elizabeth City (51055)	"Elizabeth City (incl. Hampton)" (51250)
Frederick (51069) and Winchester City (51840)	"Frederick (incl. Winchester City)" (51260)
Henrico County (51087) and Richmond City (51760)	"Henrico (incl. Richmond City)" (51270)
Martinsville (51690) and Henry (51089)	"Henry (incl. Martinsville)" (51280)
James City County (51095) and Williamsburg City (51830)	"James City County (incl. Williamsburg City" (51290)
Montgomery County (51121) and Radford City (51750)	"Montgomery (incl. Radford City)" (51300)
Nansemond (51123) and Suffolk City (51800)	"Nansemond (incl. Suffolk City)" (51310)
Norfolk (51129), Norfolk City (51710), South Norfolk City (51785), and Portsmouth City (51740)	"Norfolk County (combined)" (51320)
Danville (51590) and Pittsylvania (51143)	"Pittsylvania (incl. Danville)" (51330)
Prince George (51149) and Hopewell City (51670)	"Prince George (incl. Hopewell City)" (51340)
Roanoke (51161) and Roanoke City (51770)	"Roanoke (incl. Roanoke City)" (51350)
Rockbridge (51163) and Buena Vista City (51530)	"Rockbridge (incl. Buena Vista City)" (51360)
Rockingham (51165) and Harrisonburg City (51660)	"Rockingham (incl. Harrisonburg City)" (51370)
Fredericksburg (51630) and Spotsylvania (51177)	"Spotsylvania (incl. Fredericksburg)" (51380)
Warwick (51189) and Newport News City (51700)	"Warwick (incl. Newport News)" (51390)
Bristol (51520) and Washington (51191), VA	"Washington (incl. Bristol)" (51400)

Appendix C

Managing places of origin and matching the coded and uncoded versions of the 1940 full-count census

The process of matching the coded and uncoded (raw text) 1940 census data appears simple, requiring that the data user match on references to the original microfilm reel and manuscript page and line number; all are reported in both datasets. What appears simple turns out to be extremely difficult because IPUMS coding rules (developed for the older sample-based data sets) call for the page number to be "recoded" in order to ensure that all members of a household have the same page number, even when they span multiple pages of the original manuscript. Put another way, when a household spans multiple pages, the page number in the coded data set stays the same, but in the raw text data set page number

changes as the page turns. This is a particularly troublesome characteristic if the "household" is an institution (group quarters, military base, hospital), spanning multiple pages or if intermediate pages are missing. We have developed a script in SAS that corrects the majority of these problems, resulting in a dataset that contains 131,438,236observations (not including Alaska and Hawaii, which have limited other data);the cases lost through our merging process are primarily residents of group quarters, who would be systematically excluded from our analysis in any event. With two exceptions, our data represent between 99 and 101 percent of the official population of each state – and almost 99.9 percent of the official population of the contiguous United States as a whole – as reported in the Historical Statistics of the United States (Carter, Gartner et al. 2006).

Effectively merging the two versions did not produce useable information about the 1935 county of origin of migrants, however. We needed to assign a unique identifier (state and county FIPS code) to the 1935 place of residence for everyone enumerated in 1940. For most people, enumerated as being in the "same house" or "same place" and coded as having stayed in their origin county by IPUMS' migrate5 variable, this task was easy: we coded their 1935 county of residence FIPS to the 1940 county of residence FIPS coded by IPUMS.Individuals classified by IPUMS as moving between US counties (21<migrate5<40) were more challenging. We coded the 1935 county of individuals who IPUMS reported as making an intercounty movebut who remained within our modified county boundaries (for instance, moves between any of the five counties encompassed by New York City which we have combined into a single "county") directly from the 1940 county.For individuals who made intercounty or interstate moves by our definition we began by creating a dictionary of unique state and county text strings (from official lists), and used those to assign county IDs for individuals for whom the enumerator had written down a 1935 county of residence. This worked reasonably well, giving us an exact 1935 county of residence for roughly half of all migrants. Two sorts of problems remained: a combination of clerical mistakes by the enumerator and misspelled or incorrectly identified counties, plus respondents reporting only the city of previous residence, and not the county. We resolved this issue by attempting to match the enumerated city-state combination in the uncoded data with a dictionary of unique city-county-state combinations drawn from the coded IPUMS version of the data.

We composed thiscity-county-state dictionary by extracting all the unique 1940 city-county-state combinations from the coded IPUMS dataset using the dataset's MIGCITY variable, and removing any entries for cities that spanned more than one county, thus preventing the ambiguous assignment of a county name. Residents with an unknown county but a known city that spanned two or more counties were assigned to the county with the largest area within the city limits (for example, Amarillo residents with an unknown 1935 county were assigned to Potter County). These changes are shown in Appendix D. Changes made to align the individual level data with county-level data, described below, also resolved county assignment issues for New York City, St Louis and Virginia's independent cities. We have also fixed the miscoding in IPUMS's coded data that coded Richmond County, VA as Richmond City, VA.

At the end of these processes, among our population of interest (non-group quarter adults who remained within the contiguous United States between 1935 and 1940), less than 5% had an unknown county of origin.

Appendix D

Cities in Multiple Counties

1935 City	Possible 1935 Counties	1935 County Assigned If Unknown
Amarillo, TX (160)	Potter (48375), Randall (48381)	Potter (48375)
Atlanta, GA (350)	Fulton (13121), DeKalb (13089)	Fulton (13121)
Bethlehem, PA (730)	Northampton (42095), Lehigh (42077)	Northampton (42095)
Centralia, IL (1021)	Marion (17121), Clinton (17027)	Marion (17121)
Elgin, IL (2030)	Kane (17089), Cook (17031)	Kane (1789)
Elwood City, PA (2061)	Lawrence (42073), Beaver (42007)	Lawrence (42073)
Fostoria, OH (2351)	Seneca (39147), Hancock (39063)	Seneca (39147)
Huntington, WV (2910)	Cabell (54011), Wayne (29099)	Cabell (54011)
Joplin, MO (3210)	Jasper (29097), Newton (29145)	Jasper (29097)
Rocky Mount, NC (5974)	Nash (37127), Edgecombe (37065)	Nash (37127)
St Cloud, MN (6693)	Stearns (27145), Sherburne (27141), Benton (27009)	Stearns (27145)
Watertown, WI (7311)	Jefferson (55055), Dodge (55027)	Jefferson (55055)

Table A.1

Frequency Distribution of the top three crops for each county (only five most frequent 1st, 2nd, and 3rd crops are shown). For example, corn is the predominant crop in 34.61% of all counties in 1930 (by acreage), was the second-most in 29.76% of counties and was third in 8.73%. Thus, corn was a top three crop in 73.1% of all counties in 1930.

Largest Crop (by acreage)	%	2nd Largest Crop	%	3rd Largest Crop	%
Corn	34.61	Corn	29.76	Hay	22.57
Hay	31.49	Hay	20.74	Wheat	14.13
Cotton	19.21	Oats	17.55	Oats	13.94
Wheat	10.16	Wheat	11.85	Corn	8.73
Oats	0.72	Cotton	6.02	Barley	8.01
All Others	3.81	All Others	14.07	All Others	32.63

References

Adger WN. Vulnerability. Global Environmental Change. 2006; 16(3):268–281.

Agee, J., Evans, W. Let us now praise famous men. Boston: Houghton Mifflin company; 1941.

Anselin, L. Spatial Regression analysis in R: a workbook. Santa Barbara: Center for Spatially Integrated Social Science; 2007.

Anselin L, Syabri I, Kho Y. GeoDa: an introduction to spatial data analysis. Geographical Analysis. 2006; 38(1):5–22.

Bailey, RG. Map: Eco-regions of North America (rev.). Digitized shapefiles of this data are available at: "Bailey's Ecoregions of the Conterminous United States". In: Bailey, Robert, editor. United States Geological Survey. Washington, DC: USDA Forest Service in cooperation with The Nature Conservancy and the U.S. Geological Survey; 1997. https://www.sciencebase.gov/catalog/item/54244abde4b037b608f9e23d

Bailey, RG., Cushwa, CT. Map: Ecoregions of North America FWS/OBS-81-29. Washington, D.C: U.S Fish and Wildlife Service: 1:12,000,000; 1981.

Bakewell O. Relaunching migration systems. Migration Studies. 2013

Berkes, F., Carl, F., Johan, C. Linking social and ecological systems management practices and social mechanisms for building resilience. Cambridge, U.K, New York, NY, USA: Cambridge University Press; 1998.

Biggers, JD. United States. Census of partial employment, unemployment, and occupations: 1937. Washington, U.S: G.P.O.: For sale by the Supt of Docs; 1938.

Black R, Adger WN, Arnell NW, Dercon S, Geddes A, Thomas D. The effect of environmental change on human migration. Global Environmental Change. 2011; 21(Supplement 1(0)):S3–S11.

Black R, Arnell NW, Adger WN, Thomas D, Geddes A. Migration, immobility and displacement outcomes following extreme events. Environmental Science & Policy. 2013; 27:S32–S43.

Bogue, DJ., Hagood, MJ. Subregional migration in the United States, 1935-40. Oxford, Ohio: Scripps Foundation, Miami University; 1953.

Bogue, DJ., Shryock, HS., Jr, Hoermann, SA. Subregional migration in the United States, 1935-40. Oxford, Ohio: Scripps Foundation, Miami University; 1957.

Borjas GJ. Self-Selection and the Earnings of Immigrants. The American Economic Review. 1987; 77(4):531–553.

Boustan LP, Price V, Fishback, Kantor S. The Effect of Internal Migration on Local Labor Markets: American Cities during the Great Depression. Journal of Labor Economics. 2010; 28(4):719–746.

Boustan LP, Kahn ME, Rhode PW. Moving to Higher Ground: Migration Response to Natural Disasters in the Early Twentieth Century. The American Economic Review. 2012; 102(3):238–244. [PubMed: 25067842]

Carter, SB., Gartner, SS., Haines, MR., Olmstead, AL., Sutch, R., Wright, G. Historical Statistics of the United States Millennial Edition Online. New York: Cambridge University Press; 2006.

Cook BI, Seager R, Smerdon JE. The worst North American drought year of the last millennium: 1934. Geophysical Research Letters. 2014; 41(20):7298–7305.

Daly C, Gibson WP, Taylor GH, Johnson GL, Pasteris PP. A knowledge-based approach to the statistical mapping of climate. Climate Research. 2002; 22(2):99–113.

Daly C, Halbleib M, Smith JI, Gibson WP, Doggett MK, Taylor GH, Curtis J, Pasteris PP. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology. 2008; 28(15):2031–2064.

DeWaard, J., Curtis, KJ., Fussell, E. Demographic Signatures of Migration Systems: Population Recovery after Hurricanes Katrina and Rita. Minneapolis: Minnesota Population Center; 2014.

Egan, T. The worst hard time: the untold story of those who survived the great American dust bowl. Boston: Houghton Mifflin Co; 2006.

FAO. Guidelines for the preparation of livestock sector reviews. Rome: UN Food and Agriculture Organization; 2011.

Ferrie, JP. Internal Migration Historical Statistics of the United States Millennial Edition Online. Carter, SB.Gartner, SS.Haines, MR., et al., editors. New York: Cambridge University Press; 2006.

Fishback PV, Horrace WC, Kantor S. The impact of New Deal expenditures on mobility during the Great Depression. Explorations in Economic History. 2006; 43(2):179–222.

Fussell E, Curtis K, DeWaard J. Recovery migration to the City of New Orleans after Hurricane Katrina: a migration systems approach. Population and Environment. 2014; 35(3):305–322. [PubMed: 24729651]

Giesen, JC. Boll weevil blues: cotton, myth, and power in the American South. Chicago ; London: The University of Chicago Press; 2011.

Greenwood, MJ. Internal migration in developed countries. In: Mark, RR., Oded, S., editors. Handbook of Population and Family Economics. Vol. 1. Elsevier; 1997. p. 647-720.

Gregory, JN. American exodus : the Dust Bowl migration and Okie culture in California. New York: Oxford University Press; 1989.

Gregory, JN. The southern diaspora: how the great migrations of Black and White Southerners transformed America. Chapel Hill: University of North Carolina Press; 2005.

Haines, MR., Price, V., Fishback, Rhode, PW. United States Agriculture Data 1840 - 2010; Ann Arbor, MI. Inter-university Consortium for Political and Social Research (ICPSR); 2014.

Haines, MR. Inter-university Consortium for Political and Social Research. Historical, Demographic, Economic, and Social Data: The United States, 1790-2002. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]; 2010.

Hall PK, Ruggles S. "Restless in the Midst of Their Prosperity": New Evidence on the Internal Migration of Americans, 1850-2000. The Journal of American History. 2004; 91(3):829–846.

Hornbeck R. The Enduring Impact of the American Dust Bowl: Short- and Long-Run Adjustments to Environmental Catastrophe. American Economic Review. 2012; 102(4):1477–1507.

Kanbur R, Rapoport H. Migration selectivity and the evolution of spatial inequality. Journal of Economic Geography. 2005; 5(1):43–57.

Lee ES. A Theory of Migration. Demography. 1966; 3(1):47–57.

Lively, CE., Taeuber, C. Rural migration in the United States. Washington, U.S: Govt Print Off; 1939.

Long J, Siu H. Refugees from Dust and Shrinking Land: Tracking the Dust Bowl Migrants, Wheaton College: 25 pgs. 2013

Massey DS. Social Structure, Household Strategies, and the Cumulative Causation of Migration. Population Index. 1990; 56(1):3–26. [PubMed: 12316385]

Massey DS, Arango J, Hugo G, Kouaouci A, Pellegrino A, Taylor JE. Theories of International Migration: A Review and Appraisal. Population and Development Review. 1993; 19(3):431–466.

Massey, DS., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., Taylor, JE. Worlds in motion: understanding international migration at the end of the millenium. Oxford : New York: Clarendon Press ; Oxford University Press; 1998.

McEwan RW, Pederson N, Cooper A, Taylor J, Watts R, Hruska A. Fire and gap dynamics over 300 years in an old-growth temperate forest. Applied Vegetation Science. 2014; 17(2):312–322.

McLeman, RA. Climate and human migration : past experiences, future challenges. New York: Cambridge University Press; 2013.

McLeman RA, Smit B. Migration as an Adaptation to Climate Change. Climatic Change. 2006; 76(1-2):31–53.

Olmstead, AL., Rhode, PW. Creating abundance: biological innovation and American agricultural development. New York, NY: Cambridge University Press; 2008.

Ravenstein EG. The Laws of Migration. Journal of the Statistical Society of London. 1885; 48(2):167–235.

Ravenstein EG. The Laws of Migration. Journal of the Royal Statistical Society. 1889; 52(2):241–305.

Roy AD. Some Thoughts on the Distribution of Earnings. Oxford Economic Papers. 1951; 3(2):135–146.

Ruggles, S., Alexander, JT., Genadek, K., Goeken, R., Schroeder, MB., Sobek, M. Integrated Public Use Microdata Series: Version 5.0. U. o. Minnesota; Minneapolis: 2010.

Steinbeck, J. The grapes of wrath. New York: Modern Library; 1939.

Tolnay SE, White KJC, Crowder KD, Adelman RM. Distances Traveled during the Great Migration: An Analysis of Racial Differences among Male Migrants. Social Science History. 2005; 29(4):523–548.

U.S. Bureau of the Census. Population Internal migration, 1935 to 1940 Economic characteristics of migrants. Washington: U.S Govt Print Off; 1946. Sixteenth census of the United States: 1940.

U.S. Bureau of the Census. Measuring America: the decennial censuses from 1790 to 2000. Washington, DC: U.S Dept of Commerce, Economics and Statistics Administration, U.S Census Bureau : For sale by the Supt of Docs., U.S G.P.O; 2002.

U.S. Bureau of the Census. United States summary, 2010 population and housing unit counts. Washington, D.C.: U.S Dept of Commerce, Economics and Statistics Administration, U.S Census Bureau : Supt of Docs., U.S G.P.O (distributor); 2012.

White KJC. Women in the Great Migration: Economic Activity of Black and White Southern-Born Female Migrants in 1920, 1940, and 1970. Social Science History. 2005; 29(3):413–455.

White KJC, Crowder K, Tolnay SE, Adelman RM. Race, gender, and marriage: destination selection during the great migration. Demography. 2005; 42(2):215–241. [PubMed: 15986984]

Biographies

Myron P. Gutmann is Professor of History and Director of the Institute of Behavioral Science at the University of Colorado Boulder. His research focuses on the relationship between population and environment and on effective management and preservation of research data. He is a fellow of the AAAS.

Daniel G. Brown is Professor and Interim Dean in the School of Natural Resources and Environment at the University of Michigan. His work has aimed at understanding human-environment interactions through a focus on land-use and land-cover changes, modeling these changes, and spatial analysis and remote sensing methods for characterizing landscape patterns. He is a fellow of the AAAS.

Angela R. Cunningham is a PhD candidate in geography and a research assistant at the University of Colorado Boulder. Her research interests lie in the long nineteenth century, spatial history, and multiscale analysis, with her dissertation focused on using big microdata, event history analysis, and geographic visualization to understand the connections made by American soldiers between home and the Western Front.

James Dykes is Director of Computing and Research Services at the Institute of Behavioral Science at the University of Colorado Boulder. He worked as a Statistician for 10 years at Dartmouth College and a Technologist in storage technology and system reliability for 15 years at Seagate Technology. His work focuses on system reliability, small-area analysis and spatial methods.

Susan Hautaniemi Leonard is Associate Research Scientist at ICPSR in the Institute for Social Research at the University of Michigan. Her research is centered on relationships between human populations and their environments, especially how environmental conditions affect family demography and the interaction between family demography and land use, and the development of historical epidemiology and public health infrastructure.

Jani Little is Director of the Rocky Mountain Research Data Center and Senior Research Associate at the Institute of Behavior Science at the University of Colorado Boulder. Her interests include population redistribution processes and the indirect estimation of historical migration.

Jeremy Mikecz is a PhD candidate in history at the University of California Davis, and a research assistant at the University of Colorado Boulder. His research focuses on the application of the tools of digital and spatial history to the study of qualitative, pre-modern

texts. His dissertation applies these techniques to a study of indigenous participation in the conquest of Peru.

Paul W. Rhode is Professor of Economics at the University of Michigan, research associate at the National Bureau of Economic Research, and former co-editor of the Journal of Economic History. His research interests emphasize the economic development of the United States, especially the development of agriculture and its scientific context.

Seth Spielman is Associate Professor in the Geography Department at the University of Colorado, Boulder where he directs the Spatial Sciences Node of the NSF-Census Research Network. His research at the intersection of mapping (GIS), Statistics, and the Social Sciences focuses on the measurement of social landscapes, such as neighborhoods.

Kenneth M. Sylvester is the Director of Research and Sponsored Programs at the University of Michigan-Flint, and until recently was Research Associate Professor at the Inter-university Consortium for Political and Social Research (ICPSR), Institute for Social Research (ISR), University of Michigan, Ann Arbor. His research focuses on demographic and socioeconomic behavior in relation to environmental change, and uses census microdata, geo-spatial data, multivariate and simulation methods to study adaptive behavior, agrarian society and environmental impacts, mainly in western North America.
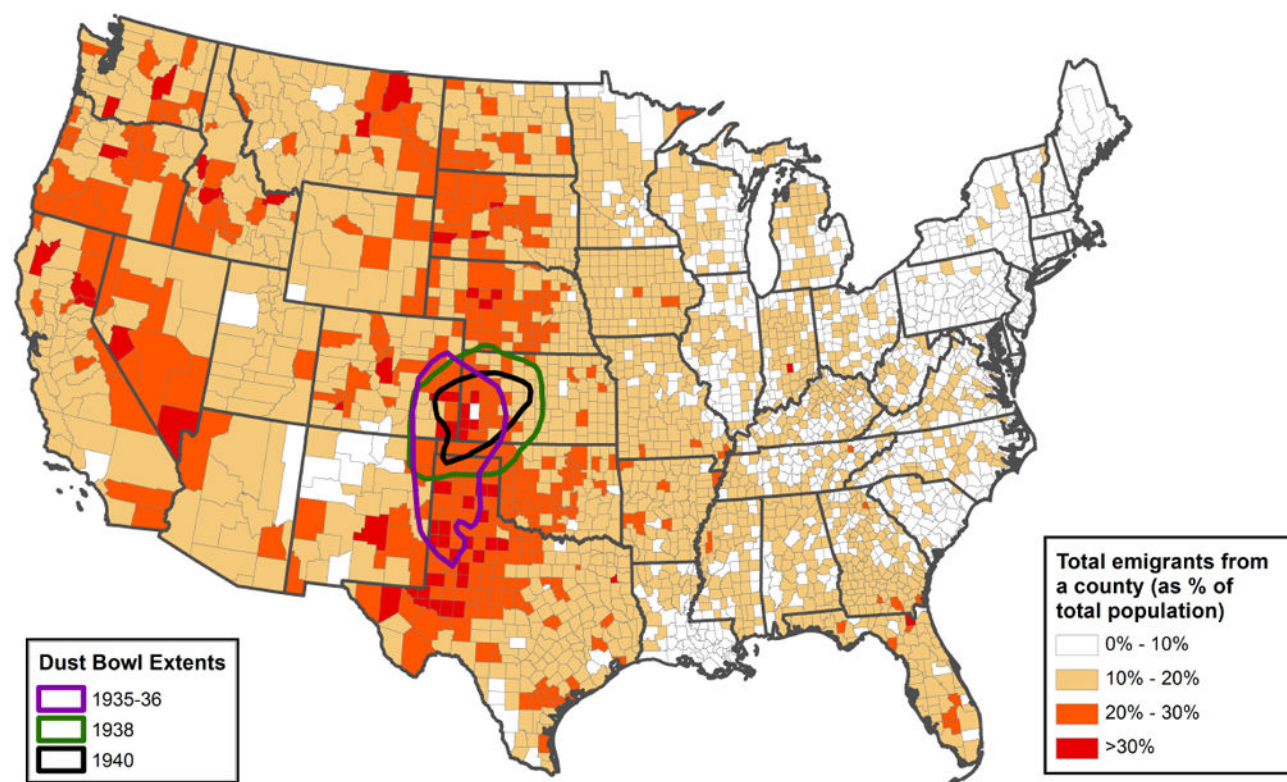
Figure 1.
Total emigrants from county, 1935-1940, as a percentage of total estimated 1935 county population.
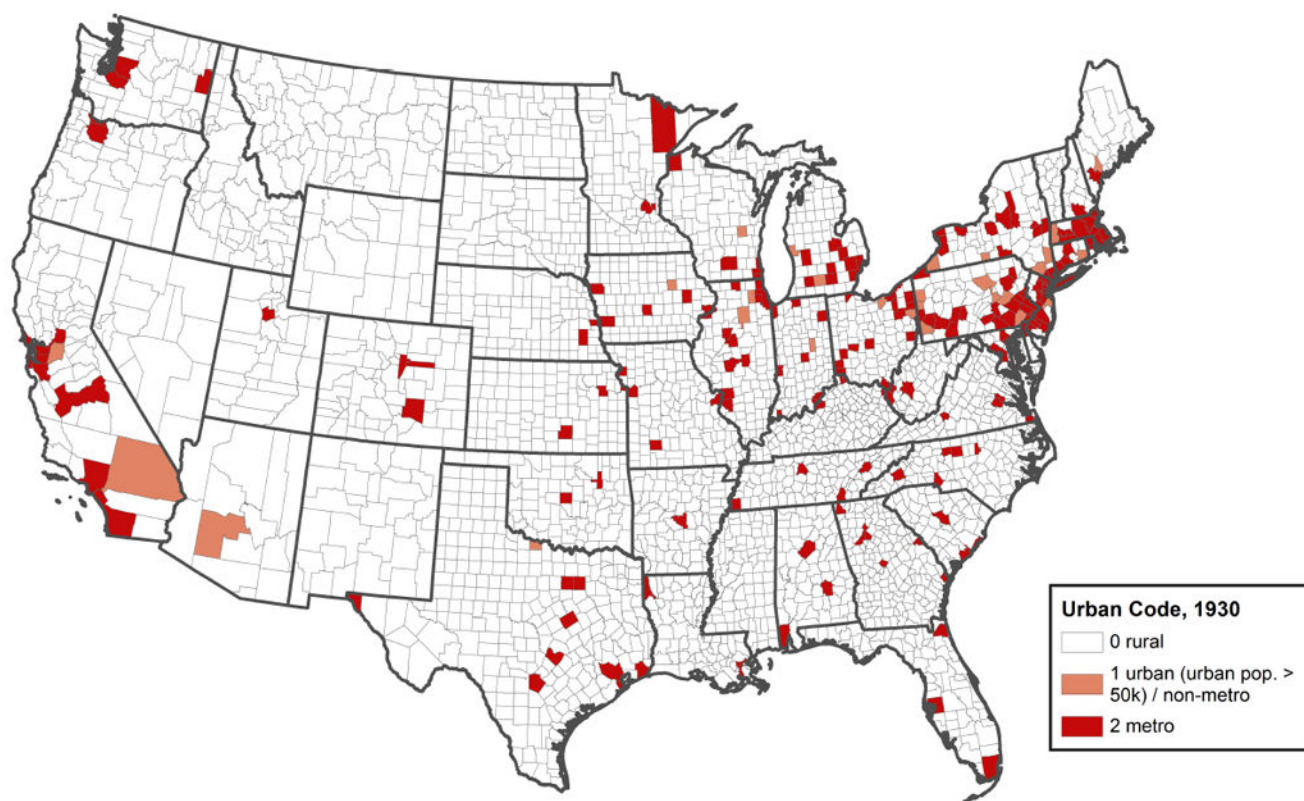
Figure 2.
Counties by urban code, 1930.
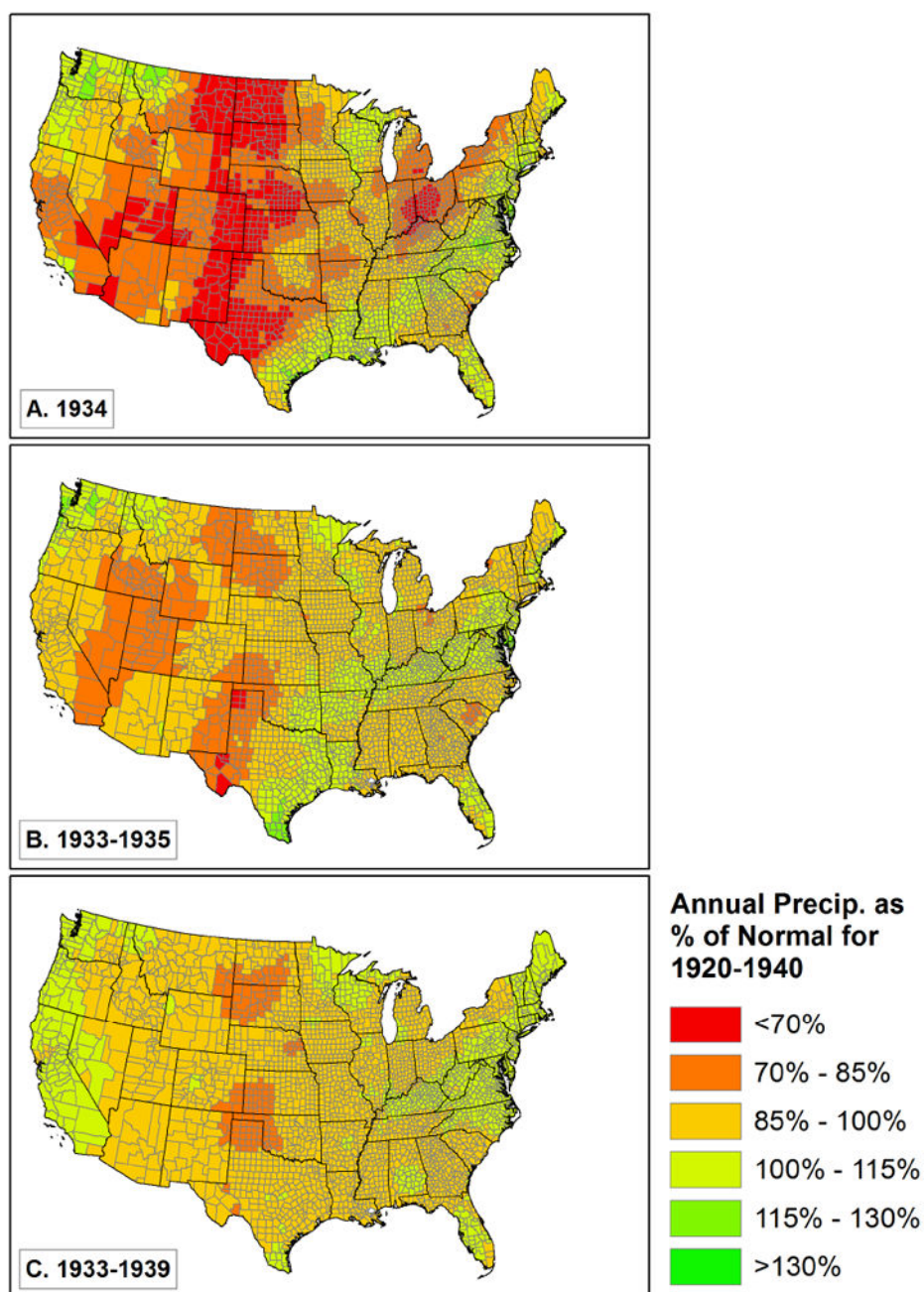
Figure 3.
Annual precipitation as a percentage of twenty-year (1920-1940) normal.
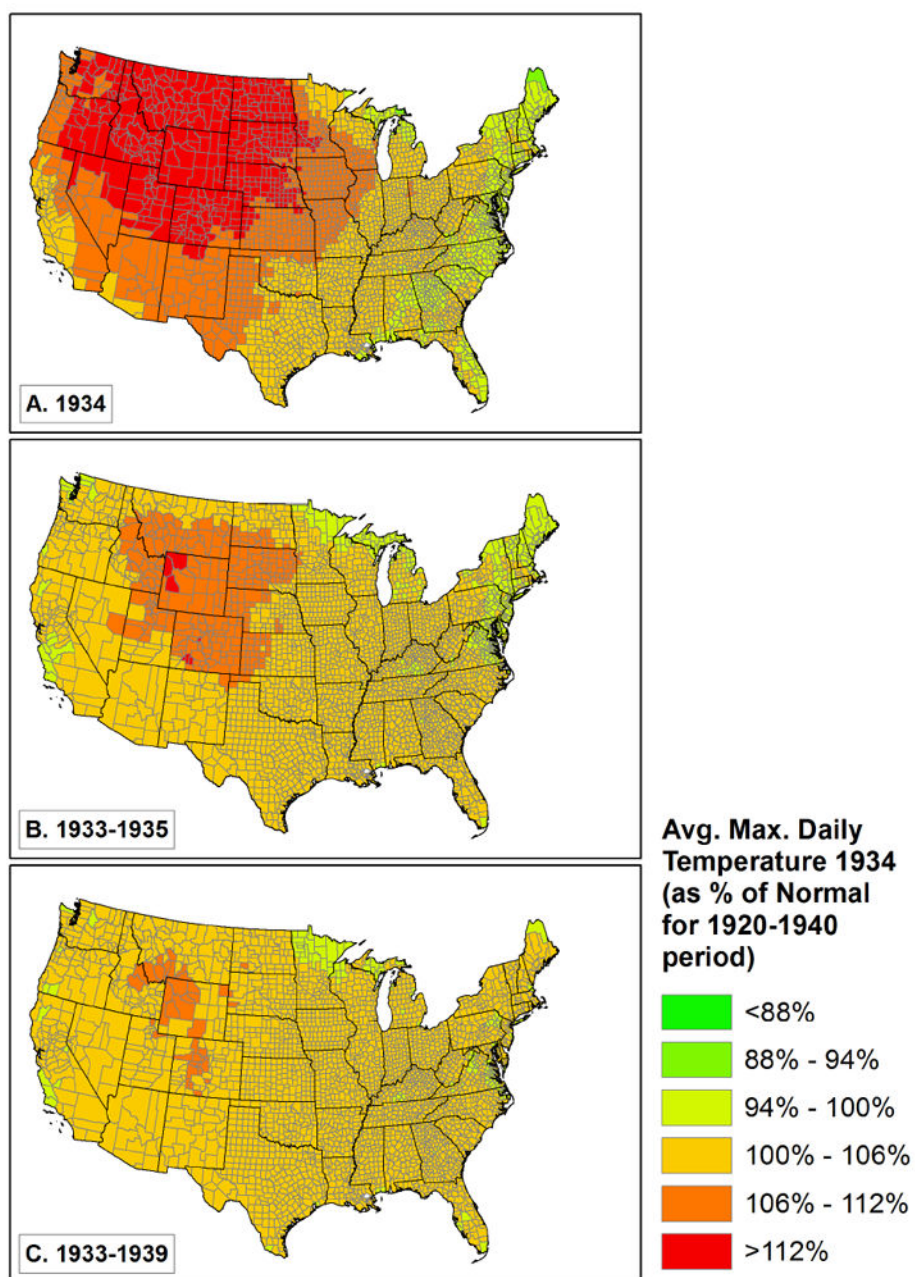
Figure 4.
Average daily maximum temperature as a percentage of twenty-year (1920-1940) normal.
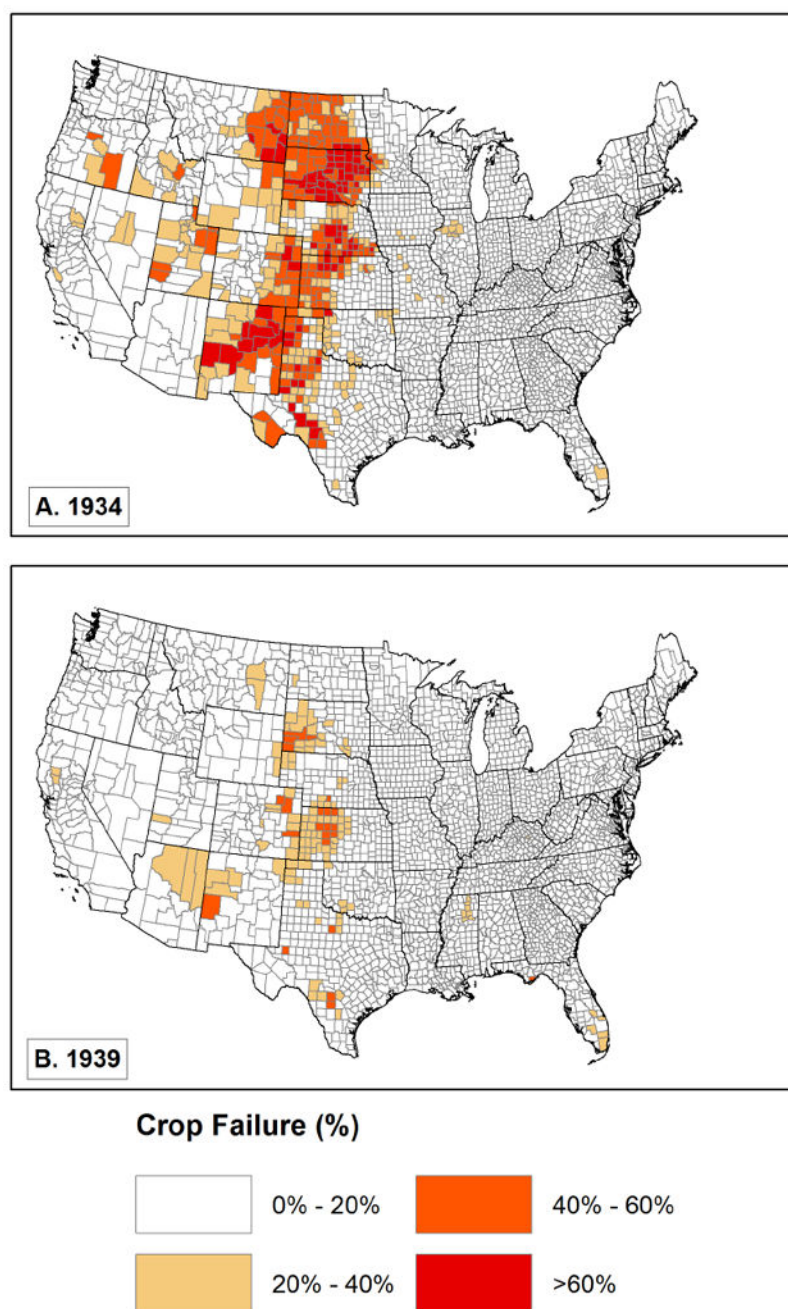
Figure 5.
Crop failure acreage as percentage of total crop acreage.
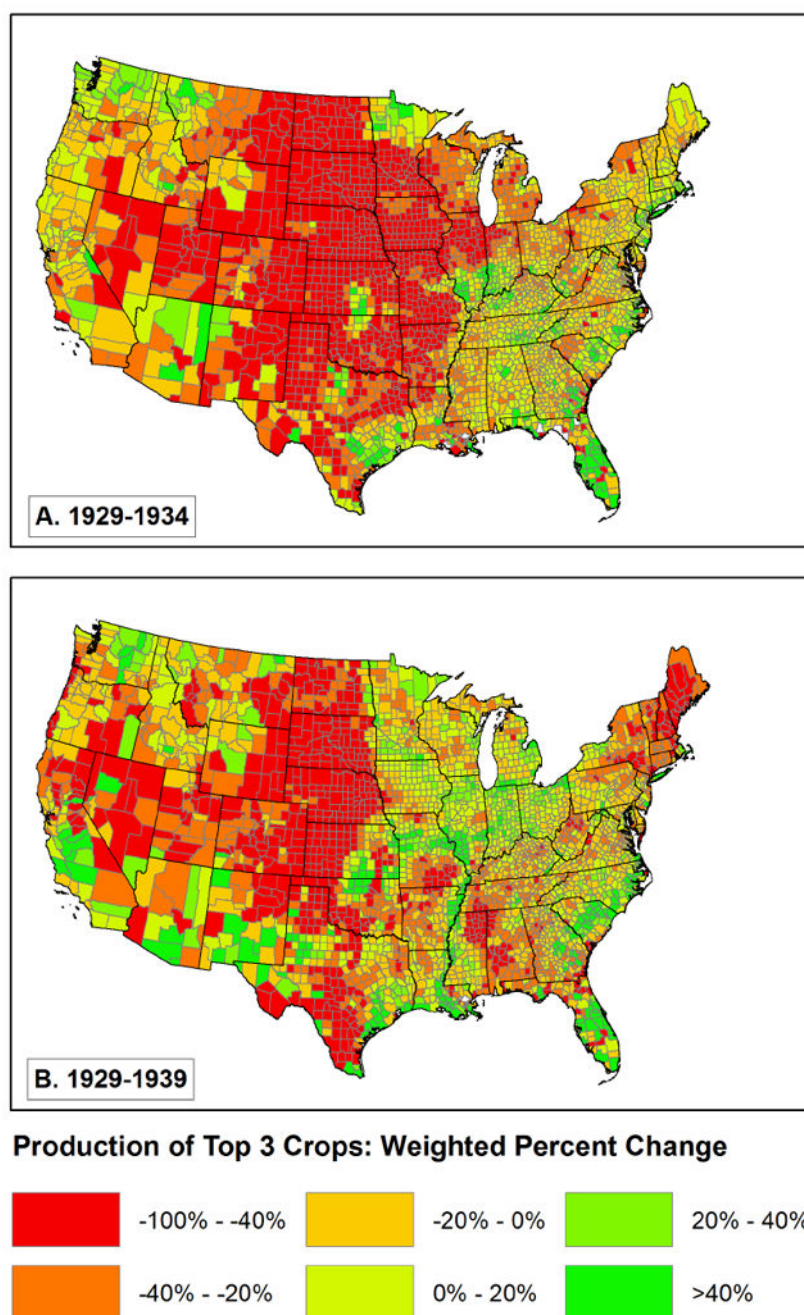
Figure 6.
Weighted percent change in county's three top crops.
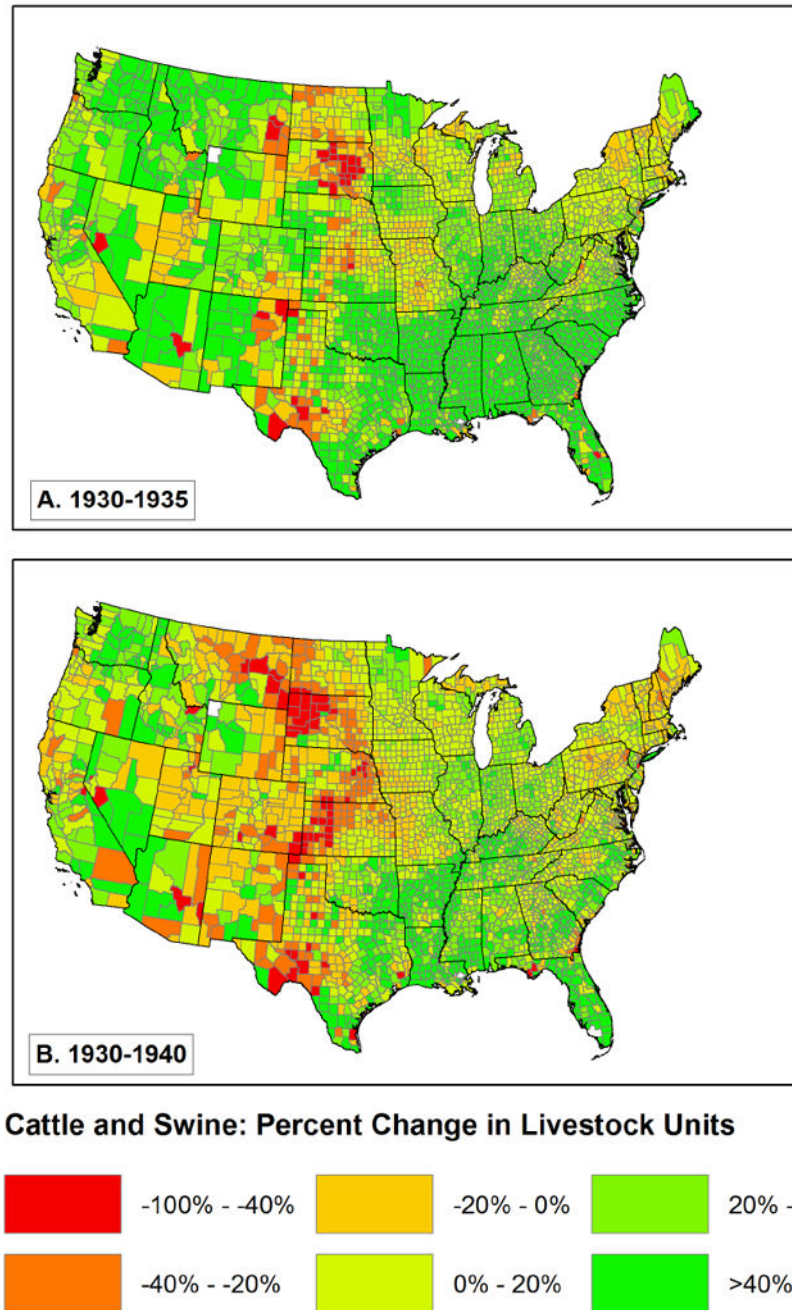
Figure 7.
Percent change in cattle and swine by livestock unit (1930-35 and 1930-40)
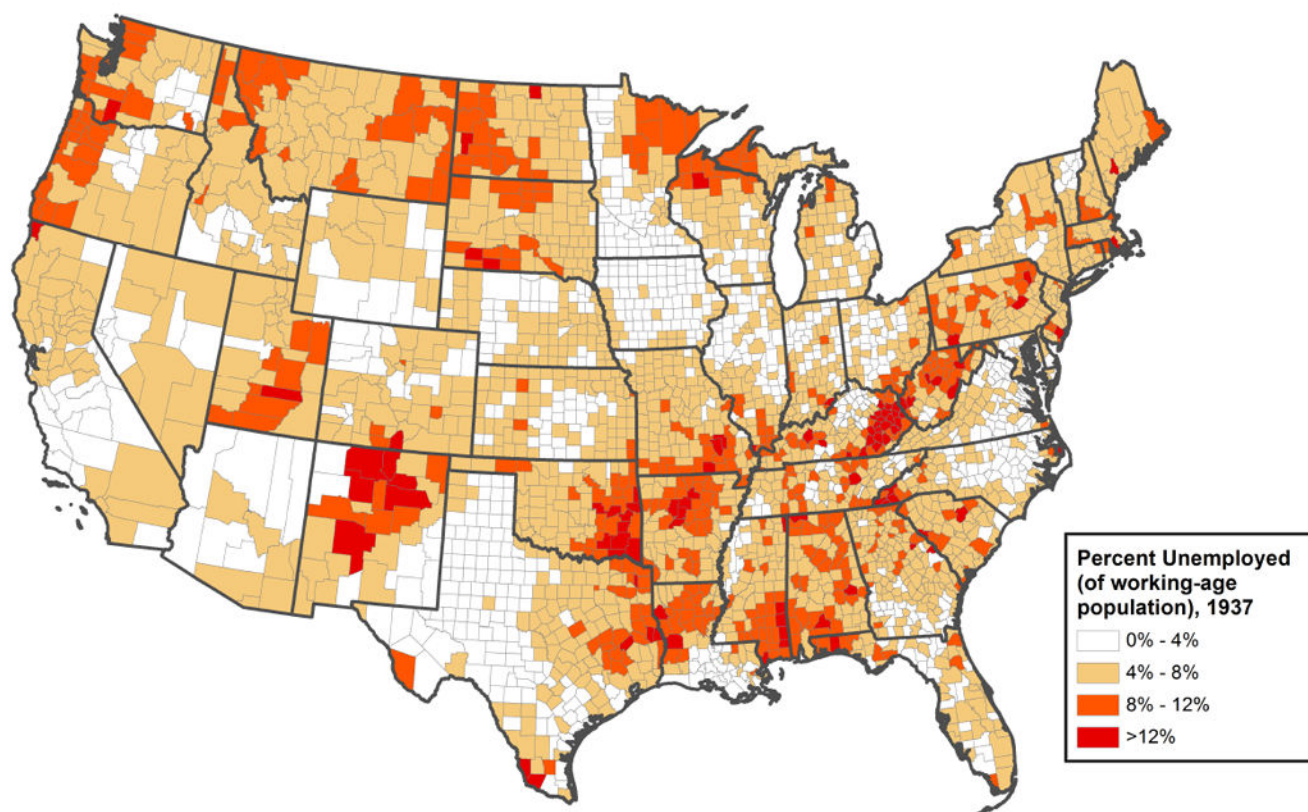
Figure 8.
Percent of working age population unemployed, 1937.
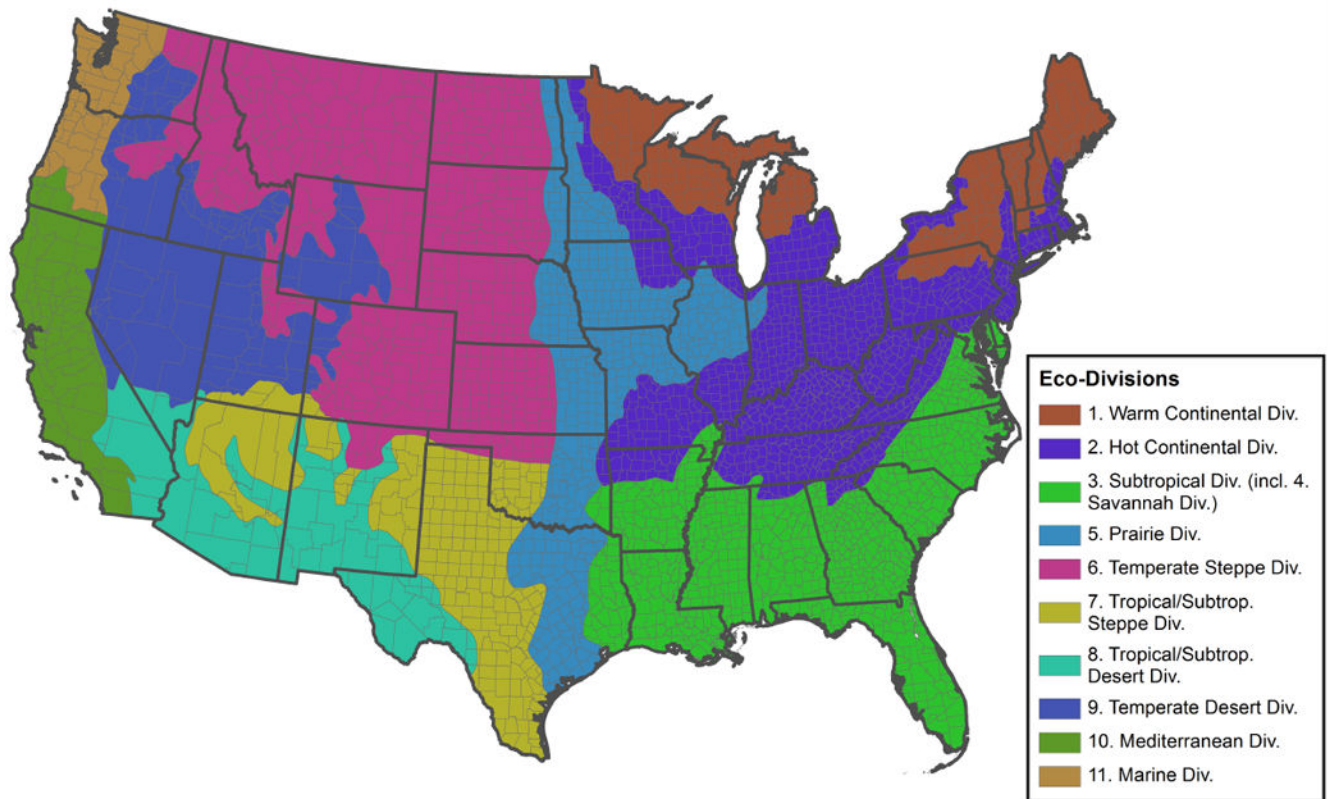
Figure 9.
Ecodivisions as defined by a United States Forest Service report in 1997 (Bailey 1997). *This report divides the United States into a hierarchical system of 4 climatic domains, 15 divisions, and 63 provinces. Excluding Hawaii and Alaska there are 11 ecodivisions which are further reduced to 10 here with the merger of division 4 (only 4 counties) into division 3.*

A. Without ecodivisions

B. With ecodivisions as main effect

C. With ecodivisions as regimes

D. With ecodivisions as regimes & spatial error term

Measured # of emigrants insignificantly different from predicted

Measured # of emigrants significantly higher than predicted

Measured # of emigrants significantly lower than predicted

Dust Bowl extent 1935-36

Dust Bowl extent 1938

Dust Bowl extent 1940

Figure 10.
Mapped residuals of four different implementations of the crop failure model.
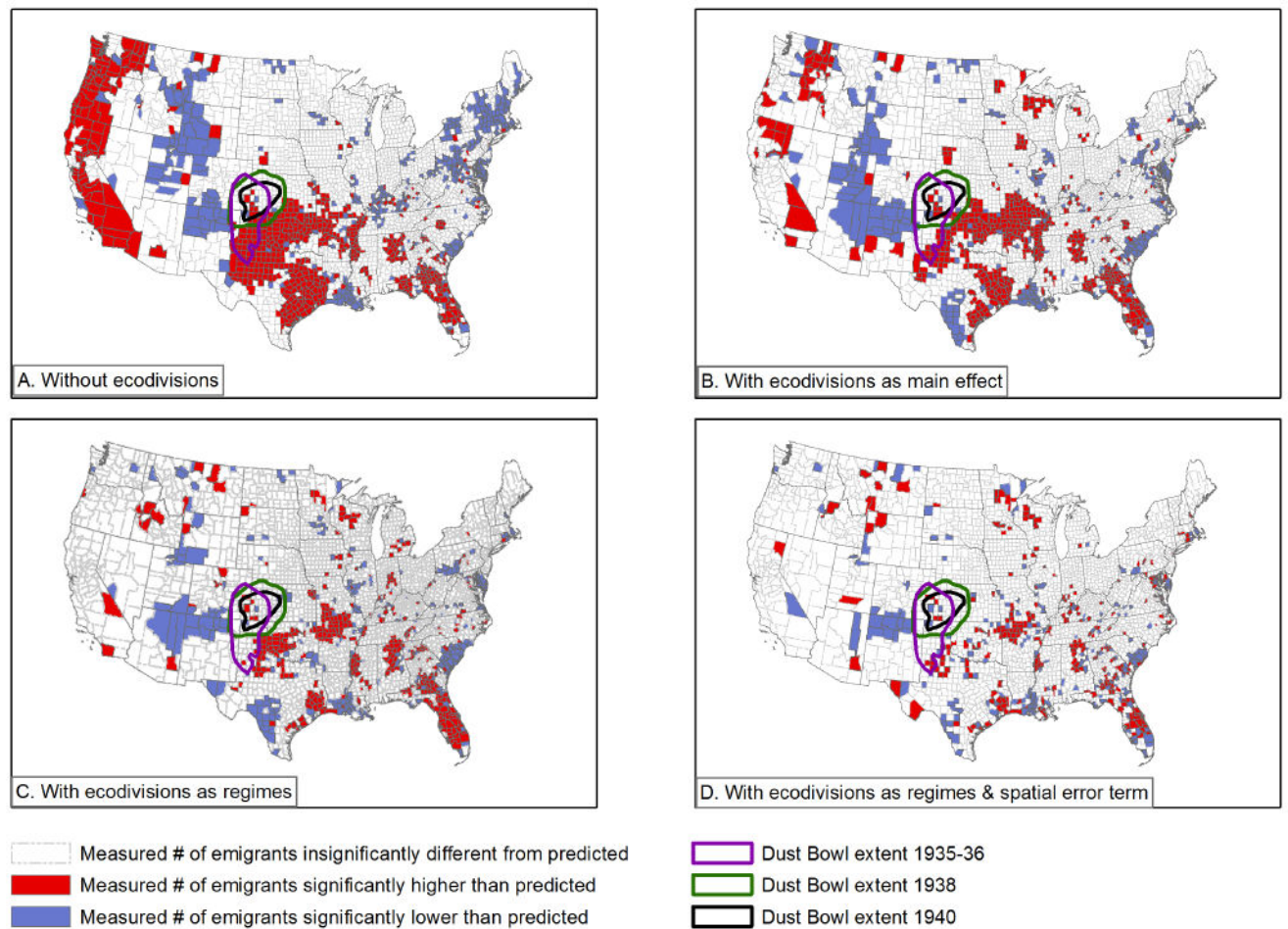
Figure 11.
Mapped residuals of four different implementations of the climate model.

Table 1

Estimated Adult Domestic Migration by 1935 State of Residence

| State | Migration type | | | Total |
| | *% of adults in migration type* | | | |
	Stayed in county	Intercounty/Intrastate move	Interstate move	
Alabama	1,359,598	96,101	68,491	1,524,190
	89.2	*6.31*	*4.49*	
Arizona	196,343	14,777	24,019	235,139
	83.5	*6.28*	*10.21*	
Arkansas	904,154	77,218	79,705	1,061,077
	85.21	*7.28*	*7.51*	
California	3,560,988	367,511	115,219	4,043,718
	88.06	*9.09*	*2.85*	
Colorado	531,664	57,350	55,670	644,684
	82.47	*8.9*	*8.64*	
Connecticut	1,021,383	19,813	29,197	1,070,393
	95.42	*1.85*	*2.73*	
Delaware	149,723	1,742	5,774	157,239
	95.22	*1.11*	*3.67*	
Dist. of Columbia	341,208	0	32,536	373,744
	91.29	*0*	*8.71*	
Florida	885,954	73,582	45,629	1,005,165
	88.14	*7.32*	*4.54*	
Georgia	1,455,648	126,115	63,783	1,645,546
	88.46	*7.66*	*3.88*	
Idaho	223,837	23,408	24,170	271,415
	82.47	*8.62*	*8.91*	
Illinois	4,715,569	177,628	218,927	5,112,124
	92.24	*3.47*	*4.28*	
Indiana	1,906,166	125,113	78,931	2,110,210
	90.33	*5.93*	*3.74*	
Iowa	1,339,286	118,863	90,633	1,548,782
	86.47	*7.67*	*5.85*	
Kansas	952,265	85,748	109,383	1,147,396
	82.99	*7.47*	*9.53*	
Kentucky	1,438,411	80,828	71,302	1,590,541
	90.44	*5.08*	*4.48*	
Louisiana	1,200,842	71,176	35,566	1,307,584
	91.84	*5.44*	*2.72*	
Maine	463,464	21,564	13,541	498,569

| State | Migration type | | | Total |
| | *% of adults in migration type* | | | |
	Stayed in county	**Intercounty/Intrastate move**	**Interstate move**	
	92.96	*4.33*	*2.72*	
Maryland	995,452	29,203	32,617	1,057,272
	94.15	*2.76*	*3.09*	
Massachusetts	2,532,894	84,696	71,260	2,688,850
	94.2	*3.15*	*2.65*	
Michigan	2,886,555	190,914	86,849	3,164,318
	91.22	*6.03*	*2.74*	
Minnesota	1,518,284	110,082	71,136	1,699,502
	89.34	*6.48*	*4.19*	
Mississippi	989,627	95,532	39,849	1,125,008
	87.97	*8.49*	*3.54*	
Missouri	2,108,622	127,600	155,807	2,392,029
	88.15	*5.33*	*6.51*	
Montana	269,676	29,275	27,170	326,121
	82.69	*8.98*	*8.33*	
Nebraska	702,010	67,731	88,426	858,167
	81.8	*7.89*	*10.3*	
Nevada	45,020	4,092	8,082	57,194
	78.71	*7.15*	*14.13*	
New Hampshire	270,340	5,937	10,794	287,071
	94.17	*2.07*	*3.76*	
New Jersey	2,425,833	98,628	79,633	2,604,094
	93.15	*3.79*	*3.06*	
New Mexico	204,125	14,723	19,136	237,984
	85.77	*6.19*	*8.04*	
New York	8,114,585	225,695	228,054	8,568,334
	94.7	*2.63*	*2.66*	
North Carolina	1,639,261	112,437	44,755	1,796,453
	91.25	*6.26*	*2.49*	
North Dakota	313,584	23,175	42,022	378,781
	82.79	*6.12*	*11.09*	
Ohio	4,018,762	209,068	142,849	4,370,679
	91.95	*4.78*	*3.27*	
Oklahoma	1,065,432	127,046	141,546	1,334,024
	79.87	*9.52*	*10.61*	
Oregon	528,513	68,990	44,908	642,411
	82.27	*10.74*	*6.99*	

State	Migration type			Total
	% of adults in migration type			
	Stayed in county	**Intercounty/Intrastate move**	**Interstate move**	
Pennsylvania	5,789,001	197,106	161,113	6,147,220
	94.17	*3.21*	*2.62*	
Rhode Island	421,421	4,666	12,498	438,585
	96.09	*1.06*	*2.85*	
South Carolina	871,526	48,602	32,331	952,459
	91.5	*5.1*	*3.39*	
South Dakota	317,760	31,049	42,232	391,041
	81.26	*7.94*	*10.8*	
Tennessee	1,480,037	86,456	76,406	1,642,899
	90.09	*5.26*	*4.65*	
Texas	3,075,063	464,807	134,326	3,674,196
	83.69	*12.65*	*3.66*	
Utah	265,103	17,161	22,743	305,007
	86.92	*5.63*	*7.46*	
Vermont	187,575	8,509	8,812	204,896
	91.55	*4.15*	*4.3*	
Virginia	1,326,306	64,306	46,343	1,436,955
	92.3	*4.48*	*3.23*	
Washington	904,744	86,388	60,278	1,051,410
	86.05	*8.22*	*5.73*	
West Virginia	932,077	54,051	39,748	1,025,876
	90.86	*5.27*	*3.87*	
Wisconsin	1,780,784	101,204	62,391	1,944,379
	91.59	*5.2*	*3.21*	
Wyoming	113,452	8,660	14,477	136,589
	83.06	*6.34*	*10.6*	
Total	70,739,927	4,336,326	3,211,067	78,287,320
	90.36	*5.54*	*4.10*	

Table 2

OLS Models of County Outmigration with Population and Ecodivisions Only

		Population only	Population and ecodivisions
DIAGNOSTICS	R squared	0.7952	0.8540
	Adj. R squared	0.7952	0.8535
	F statistic	1.192e+04 on 1 and 3069 DF	1790 on 10 and 3060 DF
	Shapiro test W value	0.8670	0.8052
	Breusch-Pagan p value	0.0900	0.0049
	VIF values >3		
	Moran's I (if significant)	0.4847	0.3198
INDEPENDENT VARIABLES		Odds (sig.)	Odds (sig.)
	Population	2.2305 ***	2.4131 ***
	Ecodivisions (*ref= warm continental*)		
	Hot continental		1.0534 ·
	Subtropical		1.1134 ***
	Prairie		1.5662 ***
	Temperate steppe		1.8560 ***
	Subtropical & tropical steppe		2.0443 ***
	Subtropical & tropical desert		1.8300 ***
	Temperate desert		1.7112 ***
	Mediterranean		1.9235 ***
	Marine		1.8707 ***

Significance codes:

*** 0

** 0.001

* 0.01

˙ 0.05

Table 3

OLS Models of County Outmigration – Crop Failure Models

		Crop failure	Crop failure with ecodivisions	Crop failure with ecodivisions (robust)
DIAGNOSTICS	R squared	0.8431	0.8613	0.9185
	Adj. R squared	0.8427	0.8605	0.9180
	F statistic	2056 on 8 and 3062 DF	1115 on 17 and 3053 DF	
	Shapiro test W value	0.8166	0.7868	0.7756
	Breusch-Pagan p value	0.0273	0.0006	0.0006
	VIF values >3		ecodiv	log of pop, ecodiv, crop failure
	Moran's I (if significant)	0.3385	0.2874	
INDEPENDENT VARIABLES		Odds (sig.)	Odds (sig.)	Odds (sig.)
	Population	2.3009 ***	2.4190 ***	2.3741 ***
	Crop failure (*ref=0%*)			
	Crop failure 1-5%	1.1245 ***	1.0970 ***	1.0808 ***
	Crop failure 5-25%	1.3765 ***	1.1976 ***	1.1822 ***
	Crop failure >25%	1.5937 ***	1.2571 ***	1.2270 ***
	% change in 1934 average daily max temperature from 20 year normal			
	% change in 1934 total precipitation from 20 year normal			
	% of population in retail employment, 1930	1.1076 ***	1.0498 ***	1.0647 ***
	% of population in manufacturing employment, 1930	0.9816 ***	0.9860 ***	0.9820 ***
	Weighted % change in 3 top crops production, 1930-1940			
	Est. % of working age (15-64) population unemployed, 1937	1.0033	1.0067 **	1.0080 ***
	County is urban	0.9390 ·	0.9292 *	0.9665
	Ecodivisions (*ref= warm continental*)			
	Hot continental		1.0589 ·	1.0788 ***
	Subtropical		1.1413 ***	1.2341 ***

		Crop failure	Crop failure with ecodivisions	Crop failure with ecodivisions (robust)
	Prairie		1.3862 ***	1.3810 ***
	Temperate steppe		1.5525 ***	1.5249 ***
	Subtropical & tropical steppe		1.7413 ***	1.7568 ***
	Subtropical & tropical desert		1.5797 ***	1.6033 ***
	Temperate desert		1.4833 ***	1.4369 ***
	Mediterranean		1.8149 ***	1.7567 ***
	Marine		1.8791 ***	1.8730 ***

Significance codes:

*** 0

** 0.001

* 0.01

˙ 0.05

Table 4

OLS Models of County Outmigration – Climate Models

		Climate failure	Climate with ecodivisions	Climate with ecodivisions (robust)
DIAGNOSTICS	R squared	0.8458	0.8655	0.9210
	Adj. R squared	0.8454	0.8647	0.9206
	F statistic	2100 on 8 and 3062 DF	1155 on 17 and 3053 DF	
	Shapiro test W value	0.8213	0.7852	0.7728
	Breusch-Pagan p value	<0.0001	<0.0001	<0.0001
	VIF values >3		ecodiv, temperature	log of pop, ecodiv, max temperature
	Moran's I (if significant)	0.3569	0.2688	
		Odds (sig.)	Odds (sig.)	Odds (sig.)
	Population	2.3751 ***	2.4558 ***	2.4000 ***
	Crop failure (*ref=0%*) *Crop failure 1-5%* *Crop failure 5-25%* *Crop failure >25%*			
INDEPENDENT VARIABLES	% change in 1934 average daily max temperature from 20 year normal	1.0175 ***	1.0069 **	1.0074 ***
	% change in 1934 total precipitation from 20 year normal	0.9951 ***	0.9940 ***	0.9950 ***
	% of population in retail employment, 1930	1.0857 ***	1.0447 ***	1.0592 ***
	% of population in manufacturing employment, 1930	0.9815 ***	0.9866 ***	0.9825 ***
	Weighted % change in 3 top crops production, 1930-1940	0.9991 ***	0.9993 ***	0.9995 ***
	Est. % of working age (15-64) population unemployed, 1937	1.0018	1.0067 **	1.0082 ***
	County is urban	0.9253 *	0.9266 *	0.9717
	Ecodivisions (*ref= warm continental*) *Hot continental*		1.0144	1.0306

		Climate failure	Climate with ecodivisions	Climate with ecodivisions (robust)
	Subtropical		1.1941 ***	1.2674 ***
	Prairie		1.3636 ***	1.3479 ***
	Temperate steppe		1.3812 ***	1.3633 ***
	Subtropical & tropical steppe		1.7012 ***	1.7037 ***
	Subtropical & tropical desert		1.4150 ***	1.4617 ***
	Temperate desert		1.3686 ***	1.3221 ***
	Mediterranean		1.7142 ***	1.6475 ***
	Marine		1.9351 ***	1.8888 ***

Significance codes:

*** 0

** 0.001

* 0.01

˙ 0.05

Table 5

Comparison of Models Using Ecodivision with Varying Techniques

	CROP FAILURE MODELS	With ecodivision as categorical independent variable	With ecodivision as regime	With spatial error term and ecodivision as regime
DIAGNOSTICS	R squared	0.8613	0.9976	
	Adj. R squared	0.8605	0.9976	
	AIC	2589.697	2480.67	2105.541
	Shapiro test W value	0.7868	0.7772	0.7238
	Breusch-Pagan p value	0.0006	0.1210	0.5830
	Moran's I (if significant)	0.2874	0.2309	insignificant

	CLIMATE MODELS	With ecodivision as categorical independent variable	With ecodivision as regime	With spatial error term and ecodivision as regime
DIAGNOSTICS	R squared	0.8655	0.9978	
	Adj. R squared	0.8647	0.9977	
	AIC	2496.307	2265.556	2032.81
	Shapiro test W value	0.7852	0.7661	0.7278
	Breusch-Pagan p value	<0.0001	0.0911	0.3594
	Moran's I (if significant)	0.2683	0.1785	insignificant