What makes communities resilient to drought?

May 8, 2016

Dan Blaustein-Rejito (GSPP), Ian Bolliger (ERG), Hal Gordon (ARE),
Andy Hultgren (ARE), Yang Ju (Landscape Arch. and Env. Planning),
Kate Pennington (ARE), Sara Stoudt (Statistics)

Abstract

Drought has affected an unprecedented area of the United States over the past several years. In April 2016, 34 percent of the United States was abnormally dry and 14 percent was in drought.[1] This dryness impacts a large portion of the population: 84.3 million people live in drought-affected areas, and 17.5 million people live in areas experiencing "exceptional drought". [2] But the welfare impacts on the people who live in drought-affected communities are not purely determined by the severity and duration of the drought. What factors predict how severely a drought will impact a community? This paper examines resilience to drought through a two-part analysis. In the first stage, we find the correlation between drought realizations and changes in key measures of welfare, including mortality, unemployment, and crop yield. Counties with low correlation can be thought of as drought-resilient, while high correlation indicates vulnerability. In the second stage, we examine which characteristics predict resilience. Understanding the predictive power of demographic characteristics on the impact of drought on welfare outcomes could have important policy implications. Our results are inconclusive, likely due to data and time limitations. We conclude with a discussion of paths forward to formalize the research methodology.

1 Why resilience matters

Drought has affected an unprecedented area of the United States over the past several years. In April 2016, 14 percent of the United States was in drought and 34 percent was abnormally dry.[3] This dryness impacts a large portion of the population: 84.3 million people live in drought-affected areas, and 17.5 million live in areas experiencing "exceptional drought". [4]

[1] US Drought Monitor 2016.
[2] US Drought Monitor 2016.
[3] US Drought Monitor 2016.
[4] US Drought Monitor 2016.

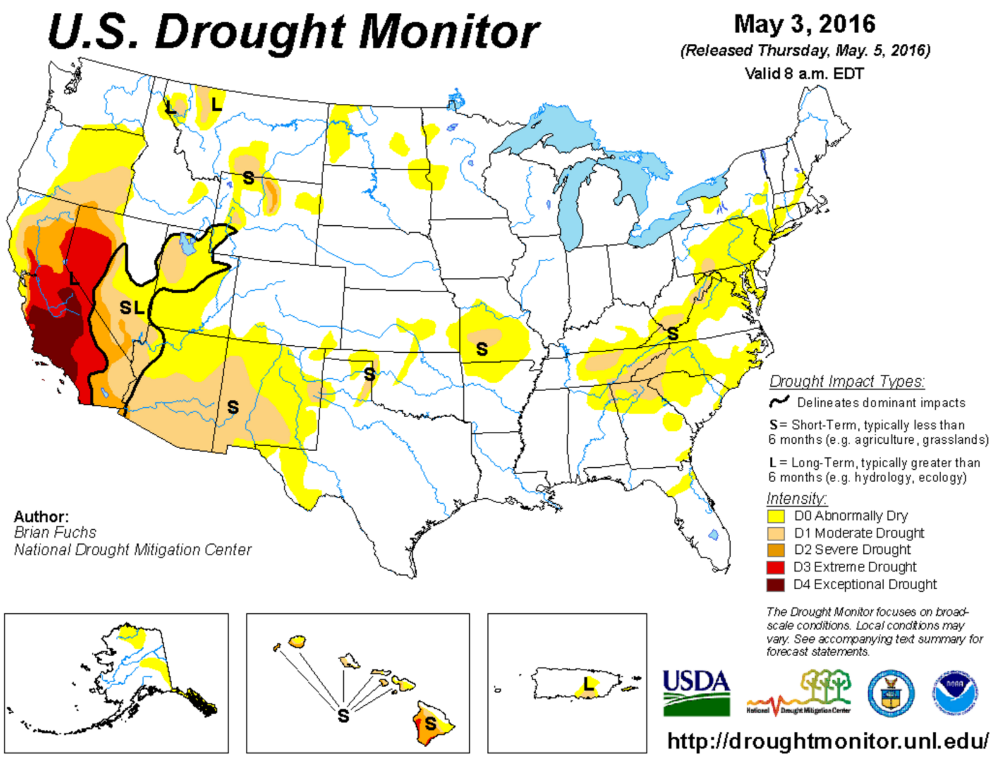


Figure 1

The drought in California demands particular attention given its severity and its impact on national food production. California grows about half of the US's fruit, nuts, and vegetables and 22% of dairy [5]. It has been in severe drought for the past five years. Currently, 90% of the state is in drought. More than 50% is in severe to exceptional drought. [6] Some towns in the Central Valley can no longer supply running water to all their residents. Some communities have undertaken dramatic mandatory water rationing, and Governor Jerry Brown has requested–and achieved–domestic water consumption decreases of around 25%.

But the welfare impacts on the people who live in drought-affected communities are not purely determined by the severity and duration of the drought. Low-income communities, like East Porterville in the CA Central Valley, are more likely to lose tap water than wealthy neighborhoods of Los Angeles. This study sets out to identify the factors that predict how severely a drought will impact a community. A better understanding of risk and resilience can support more effective policy to prevent and counteract the welfare impacts of drought.

2 Data

Towards this goal, we obtained data on drought in the United States, temperature and precipitation, economic outcomes, and demogaphic information. Each data set was cleaned, aggregated to the county level, and then merged by county-year. Brief descriptions of each data set follow.

- Drought Monitor

  The United States Drought Monitor combines a set of measures to categorize the severity of

[5] EPA State Agricultural Profiles, https://www3.epa.gov/region9/ag/ag-state.html

[6] US Drought Monitor 2016.

droughts in the US. Run in partnership between the National Drought Mitigation Center, the US Department of Agriculture, and the National Oceanic and Atmospheric Administration, a set of 250 experts reviews current data and revises the drought map every week. The Drought Monitor was launched in 1999 to be used as an input for policymakers working on issues concerning water supply and drought.

We construct drought indicator variables capturing the cumulative number of weeks a county experiences severe drought (i.e. DM level D2, D3, and D4) in the past 1 year, 3 years, and 5 years. The result is a 1-year drought index, a 3-year drought index, and a 5-year drought index.

To compute the indices, we first downloaded weekly DM polygons between 2000 and 2016 from (http://droughtmonitor.unl.edu/MapsAndData/GISData.aspx), and US county polygons from 2010 TIGER/Line (https://www.census.gov/geo/maps-data/data/tiger-line.html). The weekly DM polygons provide polygons for each DM level (i.e. D0 to D4) across the US. Second, for each week, we intersected the DM polygon with the county polygon to calculate the area under each DM level in each county. Finally, for each county, we summed the area under D2, D3, and D4 across all the weeks in the past 1 year, 3 years, and 5 years, then normalized the sum area by the area of the county.

- NARR

  We obtained and cleaned monthly temperature and precipitation from the NCEP North American Regional Reanalysis (NARR) project, but have not yet incorporated it into the analysis. If we continue this project in the future, we will use it as a first-stage dependent variable, as discussed in the model section.

- Yields

  We use annual, county-level crop yield data from the USDA for corn, soybeans, and wheat from the late 1990s to 2014. Yields are reported in bushels/acre for each crop.

- Mortality

  We use county-level data from the CDC WONDER database for under-5, over-65, and all-ages mortality for 1999-2014. Metrics of mortality include total deaths, the crude mortality rate, and age adjusted mortality rates.

- ACS

  The American Community Survey is the annual survey conducted by the US Census. Data are available at the public use microdata area (PUMA) geographic level and are mapped, or "crosswalked", to counties. Variables are weighted to represent the entire United States. We use the ACS for demographic information such as population, age, race, and sex; employment and income information including employment in agriculture, annual household income, home ownership; and household water bills.

- Employment

  We use county-level data from the United States Bureau of Labor Statistics on the annual average total labor force, number employed, number unemployed, and unemployment rate in each year of study.

- EPA

  We use geospatial data on facilities regulated by the United States Environmental Protection Agency to create a measure of the water intensiveness of the local economy. We extract the

FIPS code and the industrial category (SIC code) for each facility. Then we create a dummy variable for high water use and count the number of high-use facilities per county.

The SICs that are coded high water- use are:

- Water-intensive industries:
- 0100-0999 Agriculture, forestry, fishing
- 2000-3999 Manufacturing
- 4900-4932 Energy

There are several important issues with the use and interpretation of these data. We would like to clearly describe them up front. We decided to use the data despite these issues because it provided another opportunity to practice data management skills and because they may still provide some, albeit limited, insight.

First, these data are a snapshot of facilities managed in March, 2016. There is no way to gauge how the distribution of water-intensive industry has changed over time. This leads to two sample bias problems: first, the facilities contained in this data may not have existed during the window of our study. Second and most importantly, there may be an endogeneity problem where high water-use facilities in high-drought areas have shut down over the window of analysis and are missing from this sample. This would bias our estimates of the importance of water intensity down.

Finally, about two thirds of the observations had to be dropped due to a missing FIPS or SIC code. If this censoring is nonrandom, it introduces an additional source of bias to the study.

- USDA Economic Research Service

  We use the Rural-Urban Continuum Codes from the USDA ERS to classify counties as rural or urban. Counties with a continuum code value greater than 3 are considered "non-metro" by the USDA ERS, and we code these as rural. All other counties are considered urban.

3 Model

We develop a two-stage model. In the first stage, we estimate county-level "vulnerability" to drought by regressing key metrics of well-being on drought measures. In the second stage, we identify predictors of vulnerability.

- **Stage 1: Estimating vulnerability**

  In the first stage, we run separate Ordinary Least Squares (OLS) specifications for three dependent variables: the mortality rate, employment rate, and crop yields. We select mortality and employment because they can be argued to paint a picture of a county's general well-being. We select crop yields because they should respond strongly to drought, and give us insight into the salience of drought for agricultural versus non-agricultural communities.

  Our regression equation is as follows:

$$y_{i,t} = \beta_i D_{i,t} + \alpha_i + \tau_i t + \gamma_{s,t} + \epsilon_{i,t} \tag{1}$$

  where $D_{i,t}$ refers to the number of weeks in U.S. Drought Survey bins 2-4 in county i and year t, α_i are county fixed effects controlling for time-invariant differences between counties, τ_i is the coefficient on a county level linear time trend, and $\gamma_{s,t}$ are state-by-year fixed effects controlling for state level time trends common across all counties $i \in s$. Note that the state-by-year fixed effects will non-parametrically account for national trends in the outcome of interest as well as state-level trends. The identifying variation in this model is within-county annual deviation from the county time trend and from statewide annual average drought levels. We correct standard errors using clustering for serial correlation over space and time, due to the spatial and temporal nature of droughts (neighboring observations in time and space are not independent draws).

The coefficients from these regresssions indicate how responsive our measures of well-being – mortality, employment, and crop yields – are to drought. These measures of vulnerability will be used as the dependent variables in the second stage.

- **Second Stage: Identifying the predictors of vulnerability**

  What accounts for this variation in vulnerability to drought? We use the vulnerability estimates from the first stage as dependent variables in the second stage. We select a vector of potential predictors and test their statistical and economic significance. These predictors include county-level socioeconomic characteristics such as racial composition, age distribution, income, monthly water bill, and measures of home ownership, rural v. urban, and water-intensiveness of the local economy.

  The model is as follows:

$$\beta_i = \rho_0 + \boldsymbol{\delta}\mathbf{X}_i + \nu_i \tag{2}$$

  where the β_i are the coefficients from (1) for a given outcome; $\mathbf{X}_i$ is a vector of county-level demographic and economic information; and $\boldsymbol{\delta}$ denote the associated coefficients.

  This regression is cross-sectional and therefore not well identified from a causal perspective. Any omitted variable that happens to covary with both the levels of β_i and the variables $\mathbf{X}_i$ will bias the coefficients $\boldsymbol{\delta}$. However, this model will illustrate how "drought resilience" (a low value of β_i) covaries with a set of common county socioeconomic characteristics. The goal with this stage is to build insight regarding what characteristics are commonly associated with drought vulnerability.

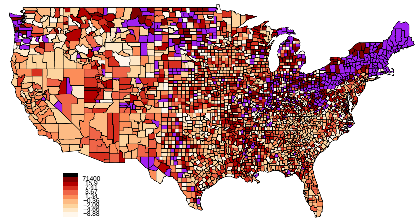
4 Discussion

The results from the first stage are depicted in the following maps. Since we run several specifications of each regression, the clearest way to present our results is visually. For tables of our full regression output, please reference the results folder on the git repository.

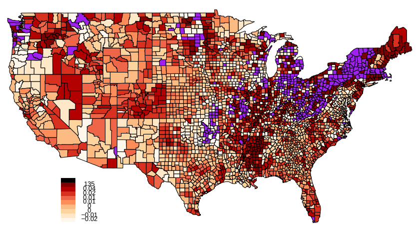
In the figure below, the colors represent the county-level correlation coefficients between three-year drought and corn yields, soybean yields, the elderly mortality rate, and unemployment,. We see comparable coefficients for corn and soy yields, with South Dakota, Indiana, and eastern North Carolina showing the most vulnerability to drought on soy yields and central Texas and southern Iowa showing the most vulnerability for corn. Still, the coefficients are generally small, as we would expect given the impact of irrigation and other technological inputs on decoupling yields from weather outcomes.

The correlation between drought and unemployment is generally low, with Maine and Mississippi showing the greatest correlation. The correlation between drought and elderly mortality is stronger (the coefficient for any-age mortality is consistent, but weaker, as expected). Health and economics research shows strong effects of heat on mortality among the elderly in the absence of air conditioning.

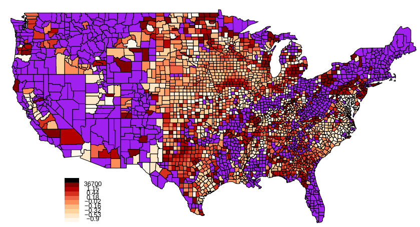
Future work should examine the impact of temperature and precipitation, rather than the composite drought measure, to see whether this effect is mainly driven by temperature or precipitation.



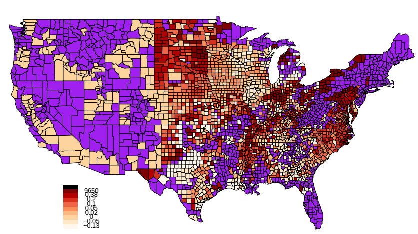
(a) Mortality



(b) Unemployment



(c) Corn



(d) Soy

Figure 2: First Stage Results
Note: Purple = NA

In the second stage of analysis, we examine whether county-level demographic and economic characteristics predict the vulnerability to drought estimated in the first stage. Our explanatory variables include demographic characteristics including percent white and percent Hispanic, age, and sex; and economic characteristics like percent of households who own their home, monthly water bill, percent employment in agriculture, and the water-intensiveness of the local economy.

Full tables including the correlation coefficients and standard errors are available in the results folder on the git repository. They are best explored through the Shiny app, which permits users to select specifications and immediately visualize results at the national and county levels.

We expected to see a positive correlation between percent Hispanic, monthly water bill, percent employed in agriculture and the water-intensiveness of local economic activity and the vulnerability to drought. However, very few of the second-stage coefficients are statistically significant. Still, given the limitations of the data and analysis, this should not be interpreted to mean that there is no statistically significant relationship. Rather, we consider our results to be inconclusive.

For example, it seems very probable that a county where one of the major employers is a clothing manufacturer using hundreds of thousands of gallons of water a day might be more vulnerable to drought, because drought could increase the cost of water and/or decrease the supply. Therefore we would expect to see a positive, significant coefficient on water use. Similarly, we would expect counties with a higher population employed in agriculture to display greater sensitivity to drought.

However, any analysis is limited by the quality of the data used. As discused in the data section, sampling bias in the water intensiveness variable could lead to the serious underestimation of its impact. Facilities that closed due to drought are no longer in the sample, so that the responses of the most vulnerable counties are not observed. This results in censoring of observations, biasing the correlation coefficients down by removing the observations for which the effect is largest.

In the next section, we outline steps that could be taken to formalize the analysis.

5 Conclusion

This paper lays out a methodology for conducting a sensitivity analysis to investigate the factors that make drought more or less harmful to different communities. This is an important research question given the large percent of the United States land area and population which are currently experiencing prolonged drought, especially given predictions of increasing incidence and severity as the climate changes.

We develop a two-stage analysis. First, we identify correlation coefficients between drought and several key measures of human well-being. Second, we identify characteristics that predict the size of those correlations. Perhaps counties with a higher proportion of white residents are less vulnerable to drought, due to a long history of racial inequality in the United States. One might expect that counties in which agriculture is a significant employer would be more vulnerable to drought, or that rural areas are more vulernable than urban areas. Questions like these can have important implications for policies to protect and relieve people from drought.

However, our analysis is limited by the data we were able to procure within the project timeline. We would like to briefly address several of these limitations, and ways to address them.

First, a community is considered to be experiencing drought if that county is in drought. However, counties are not perfectly aligned with watersheds or water sources. Consider a large city that pipes in water from a long distance away, such as the Hetch Hetchy reservoir supplying San Francisco 167 miles away, or a town like Flint, Michigan purchasing water from Detroit. A measure of local drought might have very little to do with local water availability and price. A better variable would be drought at the location of a community's water *source*. However, this would require better data and spatial analysis heavylifting, that were out of the scope of this project.

Next, the variables we selected to represent well-being in the first stage of analysis are imperfect. It would be difficult to argue that drought has a strong effect on mortality in the contemporary United States. A better choice for the impact of drought on health might be ailments affected by the moisture content of the air, such as allergies, or on people's likelihood of exercising outdoors. Given more time, we would think critically about the best ways to capture the impact of drought on health. Similarly, there may be additional variables besides unemployment and crop yields that are appropriate for first stage. Time should be spent thinking through causal chains to identify additional stage one dependent variables.

Finally, better data on agricultural employment and production and on the water-intensiveness of the local economy would enable a more rigorous second-stage.

It would also be valuable to conduct the same analysis using temperature and precipitation rather than drought.

Despite these limitations, this paper demonstrates a useful methodology for gaining traction on an important issue facing the United States today.

References

[1] U.S. Bureau of Labor Statistics. *Labor Force Data by County, Annual Averages.*
`http://www.bls.gov/lau/`

[2] Center for Disease Control (CDC), U.S. Department of Health and Human Services. *Mortality for 1999 - 2014 with ICD 10 codes.*
`http://wonder.cdc.gov/mortSQL.html`

[3] Environmental Protection Agency (EPA). *Facility Registry Service Spatial Data Download.*
`https://www.epa.gov/enviro/geospatial-data-download-service`

[4] Missouri Census Data Center. *American Communities Survey.*
`http://mcdc.missouri.edu/websas/geocorr14.html`

[5] National Centers for Environmental Prediction (NCEP). *NCEP North American Regional Reanalysis.*
`http://www.esrl.noaa.gov/psd/`

[6] National Drought Mitigation Center (NDMC), U.S. Department of Agriculture (USDA), and the National Oceanic and Atmospheric Administration (NOAA). *U.S. Drought Monitor.*
`http://droughtmonitor.unl.edu/`

[7] United States Department of Agriculture, Economic Research Service. 2013 Urban-Rural Continuum Codes. `http://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx`. Accessed 4/23/16.

[8] United States Department of Agriculture, National Agricultural Statistics Service. Qick Stats. `https://quickstats.nass.usda.gov/`. Accessed 4/15/16.