297 USING GIS TO ASSESS SOCIAL VULNERABILITY TO CLIMATE HAZARDS IN THE SOUTHERN UNITED STATES

Robert J. Gottlieb[1]*, Harold E. Brooks[2], Mark A. Shafer[3], Michael B. Richman[1]
[1]University of Oklahoma, [2]National Severe Storms Laboratory, [3]Oklahoma Climatological Survey

1. INTRODUCTION

Over the past few decades, efforts to assess and understand the socio-economic factors that contribute to a population's vulnerability to meteorological hazards have increased. It has become more common to use quantitative methods to approach this issue in State Hazard Mitigation Plans written by state emergency management agencies (e.g Odeh 2002, Simpson and Human 2008). Additionally, a report from the National Climate Adaptation Summit submitted to the White House identified vulnerability assessments as a priority for the future (NCAS Committee 2010). Emergency managers can use the results of these studies to help identify at-risk populations in their location and focus their hazard preparedness efforts.

2. BACKGROUND CONCEPTS

2.1 *Defining Vulnerability*

The term vulnerability has been defined in a staggeringly large of number of different ways during the period it has been researched (O'Brien et al. 2004, Cutter 1996). This can be attributed to the wide range of approaches that have been used to study the concept. The multitude of different approaches is not necessarily a negative consequence. Vulnerability is such a complicated topic that different approaches are necessary to further its understanding. However, the range of methods can also create confusion amongst researchers and decision makers. Therefore, it is very important to explicitly state what is actually being assessed.

This study examines the severity of effects on the human population of the contiguous United States by meteorological and climatological hazards. Some studies in meteorology have looked at deaths as the main result of high vulnerability (e.g. Ashley 2007), but there are many other negative effects, such as property damage, high insurance costs, loss of business, loss of resources, and psychological trauma. Vulnerability has also been occasionally used to discuss the ability of a population to recover of after an event, but this should instead be defined as a population's resilience (Cutter 1996).

For this work vulnerability is viewed as a broad susceptibility to loss that results from both exposure to hazards and societal preparedness for and responses to those hazards. Examining the combination of hazards and social vulnerability has been referred to as the vulnerability of place (Cutter 1996, Cutter et al. 2003). While a vulnerability of place model is the overall framework for this work, social vulnerability will be the main focus in this particular paper.

It is also important to explain some things that this work does not attempt to do. It is not an attempt to determine vulnerability to climate change. It is only a look at the degree to which a population may be affected by an event (e.g. a tornado, flood, or hurricane). It does not examine aggregated effects of expected climate change. Many other indices have been developed to assess vulnerability to climate change. For a review of these indices see Füssel (2009).

Individual sources of social vulnerability have been well documented. Vulnerable groups include the elderly, children, the disabled, the poor, and minorities (for further discussion see Cutter 2003). Areas with a high degree of infrastructure are also vulnerable because more material can be damaged, and it can be more difficult to repair so much. While broad categories are accepted, the specific causes of social vulnerability are generally very difficult to isolate. In the development of vulnerability indices data from the Census or other sources are used as a proxy for complicated processes and interactions. The goal is not to

* *Corresponding author address:* Robert Gottlieb, University of Oklahoma, School of Meteorology, Norman, OK, 73072; e-mail: rgottlieb@ou.edu

explain these processes, but rather to broadly explore the demographics of locations which may require extra attention from emergency managers and decision makers.

3. DATA

Socio-economic data is mostly obtained from the 2000 U.S. Census. Data from this source and others are available from the U.S. Statistical Abstract (U.S. Census Bureau 2009). Detailed results of the 2010 Census will not be available until March 2011. Although interdecadal estimates of Census data are made, it is preferable not to introduce an additional source of error by using estimated data instead of the complete counts performed in the Census.

The county is used as the level of analysis. The census tract level will be used in the future, but since there are fewer counties than census tracts, using counties involves less data and serves as a good test for the methodology which is described here. Data was collected and analyzed for the contiguous United States, but in the future some analysis will focus on a region covering the states of Arkansas, Louisiana, Mississippi, Oklahoma, Tennessee, and Texas.

These states were chosen because they are the study region of the Southern Climate Impacts Planning Program (SCIPP). SCIPP is a member of the National Oceanic and Atmospheric Administration (NOAA) Regional Integrated Sciences and Assessments (RISA) program. SCIPP is operated by groups at the University of Oklahoma and Louisiana State University.

Cutter et al. (2003) derived 42 Census variables from an original list of over 250. Two of these 42 could not be located, so 40 variables were used in this study (Table 1). These variables describe a variety of influences that social science studies have determined affect vulnerability. These influences fall into categories such as age, socio-economic status, race and ethnicity, and infrastructure (Cutter et al. 2003).

4. METHODOLOGY

Researchers and emergency managers have employed several different quantitative models to calculate vulnerability scores. The one that is examined in this study is the Social Vulnerability Index (SoVI), which was developed by Cutter et al. (2003). The SoVI was chosen because it offers an objective method to determine which variables are most important to vulnerability, a quality that most other indices of this type lack. This is needed because researchers generally have not agreed upon which variables to choose subjectively. This objectivity is achieved by identifying the variables that vary the most across the study area.

The SoVI is designed to quantify only the vulnerability of populations that results from socio-economic factors. The SoVI uses principal components analysis (PCA) to empirically compress a large number of variables into a small number of components. Many descriptions of PCA are available (e.g. Wilks 2006). Each component is a linear combination of the original variables. However, the value of each component will be dominated by only a few variables. A small number of components are selected to represent most of the information and variability in the data set. Each component is associated with an eigenvector of the covariance matrix of the original data set. The first eigenvector points in the direction in which the data has the highest variability. Subsequent eigenvectors will be orthogonal to all previous eigenvectors. A varimax rotation of the eigenvectors is used to maximize the loadings for each component onto a small number of variables. For a detailed explanation of the benefits of rotating the eigenvectors see Richman (1986).

North's test is used to select the number of components that are retained (North et al. 1982). North's test requires the retention of only eigenvalues that can be distinguished from the eigenvalues of neighboring components. The sampling error of an eigenvalue is calculated using $\delta\lambda \sim \lambda(2/n)^{1/2}$. The use of North's test differs from Cutter et al. (2003), which used the Kaiser criterion to select the number of eigenvalues to be retained. When using the Kaiser criterion, all components with eigenvalues greater than 1 are retained. This is an arbitrary distinction that does not consider the variability among the eigenvalues. It generally retains too many components. In this analysis North's test kept seven components, while the Kaiser criterion would have kept ten (Figure 1).

After selecting components, a varimax rotation is used to maximize the loadings for each

MED_AGE	Median age
PERCAP	Per capita income
MVALOO	Median value of owner-occupied housing
MEDRENT	Median rent
PHYSICN	Physicians per 1000 people
PCTVOTE	Percentage of votes for the winning party in the presidential election
BRATE	Births per 1000 people
MIGRA	Net international migration over the last decade
PCTFARMS	Percentage of land that is farms
PCTBLACK	Percentage of population that is African-American
PCTINDIAN	Percentage of population that is Native American
PCTASIAN	Percentage of population that is Asian
PCTHISPANIC	Percentage of population that is Hispanic
PCTKIDS	Percentage of population that is under five years old
PCTOLD	Percentage of population that is over 65 years old
PCTVLUN	Percentage of civilian labor force that is unemployed
AVGPERHH	Average people per household
PCTHH75	Percentage of households with income over $75,000
PCTPOV	Percentage of population living in poverty
PCTRENTER	Percentage of occupied housing units that are rented
PCTRFRM	Percentage of population that are rural farmers
DEBREV	Local government general debt to revenue ratio
PCTMOBL	Percentage of households that are mobile homes
PCTNOHS	Percentage of population over 25 without a high school diploma
HODENUT	Number of housing units per sq. mi.
HUPTDEN	Number of housing permits per new residential construction per sq. mi.
MAESDEN	Number of manufacturing establishments per sq. mi.
EARNDEN	Industry earnings (thousands of dollars) per sq. mi.
COMDEVDN	Commercial establishments per sq. mi.
CVBRPC	Percentage of population in the labor force
FEMLBR	Percentage of females participating the civilian labor force
AGRIPC	Percentage of population employed in extractive industries
TRANPC	Percentage of population employed in the transportation industry
SERVPC	Percentage of population employed in the service industry
NRRESPC	Percentage of population living in nursing homes
PCCHGPOP	Population change percentage over the last decade
PCTURB	Percentage of population living in urban areas
PCTFEM	Percentage of population that is female
PCTF_HH	Percentage of households led by single females
SSBENPC	Percentage of population receiving Social Security benefits

Table 1: The list of all socio-economic variables that were included in the analysis.

component onto a small number of variables. Each component is scaled so that a positive score indicates higher vulnerability (e.g. some scores are multiplied by -1). A county's SoVI score is found by taking the sum of each of its adjusted component scores. The result is that a county's SoVI score and its component scores are similar to z-scores. However, there is a slight difference that will be discussed in the next section.

Geographic Information Systems (GIS) provide a convenient way to map the score of vulnerability indices and their individual components and variables. It is additionally useful that hazard data can be mapped on top of these data and combined with them. GIS can be easily used at both the national scale and more local scales. However, it is important to take great care when using and mapping hazard data. Data on many hazards are incomplete or error-prone, and this is especially true of older data.

Once hazards and social vulnerability data are mapped together, the best method to combine them is not obvious. Some studies have chosen to calculate vulnerability as the product of a social

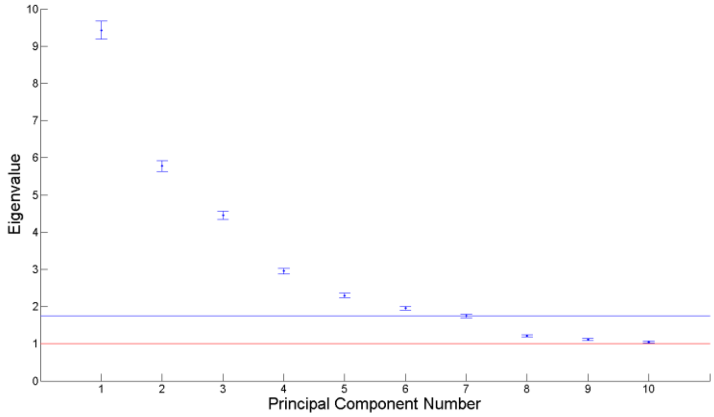


Figure 1: A comparison of North's test (cutoff in blue) and the Kaiser criterion (cutoff in red). The Kaiser criterion retains components that have eigenvalues that are too close to distinguish from one another.

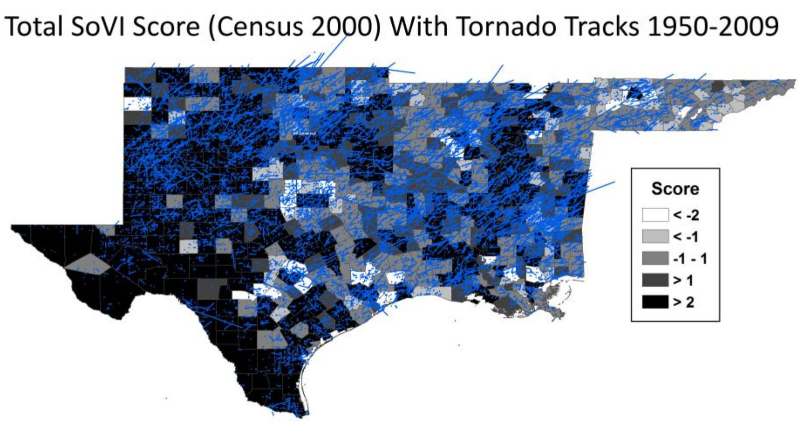


Figure 2: An example of the SoVI score mapped together with tornado tracks.

vulnerability score (sometimes called risk or exposure) and a hazard score (e.g. Simpson and Human 2008). However, this equation has two significant flaws. Suppose this equation were rearranged to calculate the hazard score, i.e. the frequency and/or severity of hazards. The result is now

$$\frac{Vulnerability}{Risk} = Hazard$$

With this equation an increase in an area's risk implies that its hazard score would decrease without any physical reason. The second problem arises if negative numbers are used, as they are in the SoVI. A location with a negative SoVI score would have a more negative vulnerability score if it had a higher positive hazard score. A different method of quantitatively combining hazard and vulnerability information is recommended, but it has not yet been developed for this study. Although this framework is not yet in place, a qualitative illustration of merging hazard and social vulnerability data can still be helpful (Figure 2). This figure shows an overlay of the SoVI and tornado tracks between 1950 and 2009. This figure is presented only as an example of a map made with GIS containing both social vulnerability and hazard data. It is known that hazard and loss datasets have significant flaws, particularly early in their records (Gall et al. 2009). For these reasons, no attempt is made to draw conclusions from it at this time. Tate et al. (2010) have presented an outline for mapping vulnerability to multiple hazards, but work following from this paper may deviate from those recommendations.

5. PRELIMINARY RESULTS

5.1 *General Results of PCA*

The use of PCA organized the data into seven components. These components, along with the variables that have the highest correlation with the components are listed in Table 2. The components explain 71.6% of the variance in the data. The variance of each component is computed by dividing a component's eigenvalue by the sum of the eigenvalues for all 40 components. The seven components can be roughly described as representing socio-economic status, age, infrastructure, rural agriculture, gender, growth, and employment security. The content of the components is largely consistent with the results of Cutter et al. (2003). The components that are identified match many of the key contributions to vulnerability that social scientists have identified.

The SoVI score and the scores of the components for each county are mapped using GIS (Figure 3). A few general regions of high

Principal Components	Variance Explained	Variables	Correlation
Socio-economic Status	23.6%	PCTPOV	0.89
		CVBRPC	-0.88
		FEMLBR	-0.86
		PCTNOHS	0.84
		PERCAP	-0.76
		PCTMOBL	0.69
		PCTHH75	-0.68
		PCTVLUN	0.62
		MEDRENT	-0.60
Age	14.4%	PCTKIDS	0.90
		MEDAGE	-0.88
		AVGPERHH	0.85
		BRATE	0.83
		PCTOLD	-0.81
		SSBENPC	-0.78
Infrastructure	11.1%	MAESDEN	0.96
		EARNDEN	0.96
		HODENUT	0.93
		HUPTDEN	0.87
Rural Agriculture	7.4%	PCTFRM	0.80
		AGRIPC	0.79
		PCTFARMS	0.76
Gender	5.7%	PCTFEM	-0.71
		PCTF_HH	-0.66
Growth	4.9%	MIGRA	-0.90
		COMDEVDN	-0.90
		PHYSICN	-0.87
		PCTASIAN	-0.60
Employment Stability	4.4%	SERVPC	-0.68
		PCTRENTER	-0.64

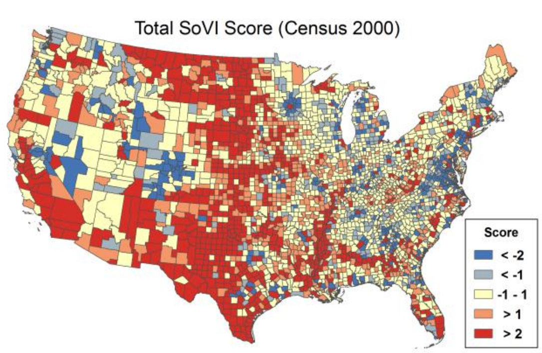
Table 2: The principal components of social vulnerability found, the amount of variance of the data set they explain, their main variables of the components, and correlation of each variable with its component.

vulnerability are revealed in the map of the total SoVI score. The first region is large cities. This is largely the result of a very high amount of infrastructure and high net migration. However, many surrounding counties of large cities (e.g. Dallas, Atlanta, Nashville, TN, Indianapolis, and Milwaukee) have low vulnerability scores. Another area of high vulnerability is the lower Mississippi River Valley. This is the connected to relatively low socio-economic status and a high percentage households led by females with no spouse present. A third area of high vulnerability is the rural Great Plains. This is mostly related to the population's high dependence on rural agriculture for its livelihood. Agriculture is among the occupations that is most affected by natural disasters. Lastly, southern California has high vulnerability connected to high net migration.

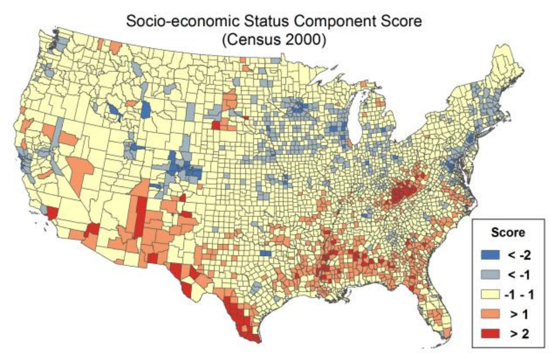
5.2 *Results of Individual Components*

A further look at the scores for the individual components reveals some interesting patterns and raises points for discussion aside from the ones mentioned in the previous section. First, a further discussion of the age component is needed. The age component is evenly split between variables that are positively correlated with the component (e.g. the percentage of people who are under five years of age) and the variables that are negatively correlated with the component (e.g. the percentage of people who are over 65 years of age). However, high values for both of these variables correspond to a higher vulnerability. For this reason the absolute value of the age component is used in calculation of the SoVI. Therefore the scale used in this map is different

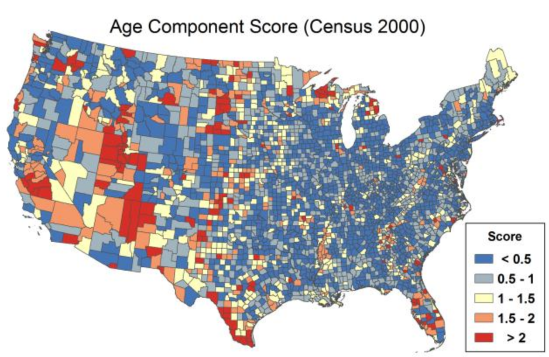
a



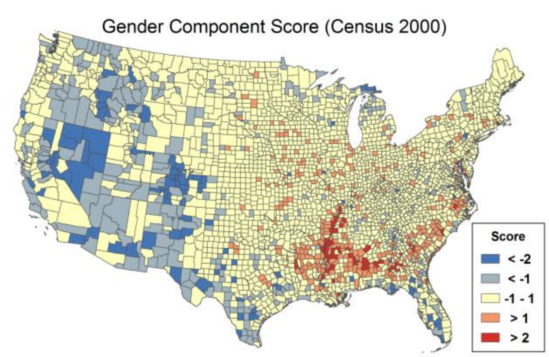
b



c



d



e f

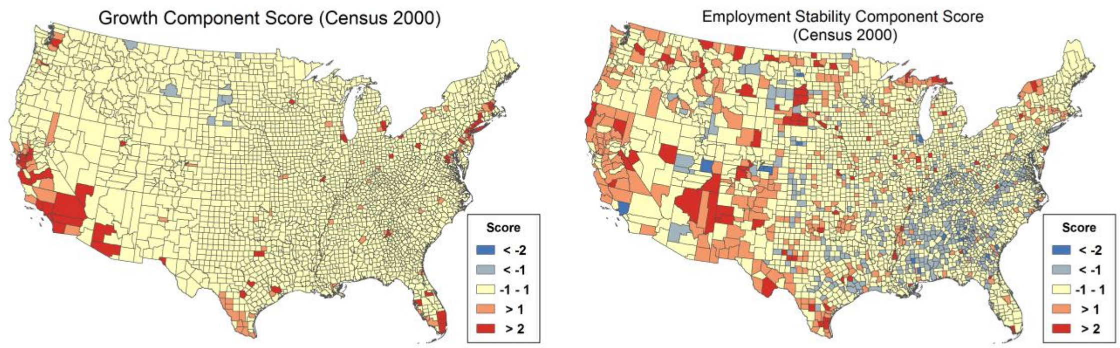


Figure 2: Maps of the total SoVI score and five of the seven components found using PCA. All maps are based on data from the 2000 Census (a) Total SoVI; (b) socio-economic status component; (c) age component; (d) gender component; (e) growth component; (f) employment stability score. Note that the age component has been adjusted so that all of its scores are positive.

from the other maps. The vast majority of the age component scores are close to zero. Interestingly, there is both low socio-economic status and high employment stability in the southeastern United States. The employment stability component consists of only the percentage of occupied housing units that are rented and the percentage of people who are employed in the service industry.

6. CONCLUSIONS AND FUTURE WORK

This paper has described a framework for assessing vulnerability caused by the occurrence of meteorological hazards and disaster preparedness based on socio-economic circumstances. Census data has been used to create a social vulnerability index modeled after Cutter et al. (2003), but some problems with previous designs of vulnerability indices have also been described. Some general areas of high social vulnerability have been identified, but more work is needed before more specific conclusions are made.

This work is still at an early stage and much more will be added in the near future. Once the most detailed data from the 2010 Census is released an analysis will be performed at the census tract level. More hazard data will be added to this assessment and combined with the SoVI. An analysis focusing on the southern United States will also be conducted. The possibility of adding other information to a model of vulnerability will also be explored, but if may prove very difficult. For example, information about tornado siren coverage would be desirable, but a national database of tornado sirens may not exist. Changes in vulnerability over a period of time may also be examined. Also, a reliable method of validating a vulnerability index is needed.

7. ACKNOWLEDGEMENTS

The authors wish to thank James Hocker for his helpful advice in preparing this manuscript.

8. REFERENCES

Ashley, W. S., 2007: Spatial and temporal analysis of tornado fatalities in the United States: 1880-2005. *Wea. Forecasting*, **22**, 1214-1228.

Cutter, S. L., 1996: Vulnerability to environmental hazards. *Prog. Hum. Geogr.*, **20**, 529-539.

Cutter, S. L., B. J. Boruff, and W. L. Shirley, 2003: Social vulnerability to environmental hazards. *Soc. Sci. Quart.*, **84**, 242-261.

Füssel, H. M., 2009: Review and quantitative analysis of indices of climate change exposure, adaptive capacity, sensitivity, and impacts.

Background note to the World Development Report 2010. World Bank, 34 pp.

Gall, M., K. A. Borden, S. L. Cutter, 2009: When do losses count?: Six fallacies of natural hazards loss data. *Bull. Amer. Meteor. Soc.*, **90**, 799-809.

National Climate Adaptation Summit Committee, 2010: National Climate Adaptation Summit Committee Report.

North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.*, **110**, 699-706.

O'Brien, K., S. Eriksen, A. Schjolden, L. Nygaard, 2004: What's in a word?: Interpretations of vulnerability in climate change research. CICERO Working Paper.

Odeh, D. J., 2002: Natural hazards vulnerability assessment for statewide mitigation planning in Rhode Island. *Nat. Hazards Rev.*, **3**, 177-187.

Richman, M. B., 1986: Rotation of principal components. *J. Climatology*, **6**, 293-335.

Simpson, D. M. and R. J. Human, 2008: Large-scale vulnerability assessments for natural hazards. *Nat. Hazards*, **47**, 143-155.

Tate, E., S. L. Cutter, M. Berry, 2010: Integrated multihazard mapping. *Env. Planning B: Planning Design*, **37**, 646-663.

U.S. Census Bureau, Statistical Abstract of the United States: 2010 (129th Edition) Washington, DC, 2009; <http://www.census.gov/statab/www/>

Wilks, D., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Elsevier, 648 pp.