

An analysis of methods for identifying social vulnerability to climate change and sea level rise: a case study of Hampton Roads, Virginia

Sarah Stafford¹  · Jeremy Abramowitz¹

Received: 13 April 2016 / Accepted: 6 October 2016 / Published online: 14 October 2016
© Springer Science+Business Media Dordrecht 2016

Abstract This paper examines methods for quantitatively identifying communities that have high social vulnerability to environmental hazards. We first provide an overview of the existing literature on quantifying social vulnerability. We discuss the strengths and weaknesses of the one of the most common methods, composite indexing using principal component analysis (PCA), in more detail. We present a case study of Hampton Roads, Virginia, that compares PCA-based composite indexing to an alternative method using *k*-means clustering to identify socially vulnerable communities. We find that PCA-based indexing is particularly sensitive to changes in geographic boundaries and the number of input variables while clustering is less so. However, both methods exhibit the highest levels of consistency when vulnerability is measured by a relatively small number of well-established quantitative indicators. While the cluster method is more intuitive and easier to interpret than composite indexing, it is not well suited for large analyses across a disparate geographic area or where one needs to explicitly rank observations.

Keywords Resilience · Sea level rise · Principal component analysis · Cluster analysis

1 Introduction

For many reasons, the poor, elderly, and disabled are generally at much greater risk to natural disasters such as flooding and hurricanes than the general public.¹ The term social vulnerability is often used to describe this phenomenon. While there is no consensus as to the exact definition of social vulnerability, for the purposes of this paper we use Blaikie

¹ See, for example, Fothergill and Peek (2004) and Fothergill et al. (1999).

✉ Sarah Stafford
slstaf@wm.edu

¹ College of William and Mary, Williamsburg, VA, USA

et al.'s (1994) definition that social vulnerability refers to the characteristics of an individual or group that influences their capacity to anticipate, cope with, resist and recover from a physical hazard. Since socially vulnerable groups are less resilient to physical hazards, they may face disproportionate losses from either a natural or man-made disaster. Our focus in this paper is on social vulnerability to natural hazards, particularly those that are likely to increase due to climate change and sea level rise. Following Hurricane Katrina, the Stafford Act, which directs how federal resources are used to respond to major disasters, was revised to require that any federal funds used to respond to a disaster must provide equitable treatment on the basis of race, sex, age, disability, English-language proficiency, and economic status.² Thus, as governments formulate or evaluate potential policy responses to impending sea level rise they will need to have an understanding of the social vulnerabilities in their communities to ensure such equitable treatment.

The literature on social vulnerability has its roots in the 1976 article in *Science* by O'Keefe et al., "Taking the Naturalness out of Natural Disasters." In this article the authors argue that impact of disasters is as much, if not more, a consequence of socioeconomic factors, particularly what they term "underdevelopment," than natural factors. This paper spurred a number of other papers and analyses examining socioeconomic vulnerability to physical hazards, primarily using a political ecology framework. Building on this initial literature, Blaikie et al. (1994) developed a "pressure and release model" of social vulnerability that examines the ways in which social, economic, and political processes influence how hazards affect different people. Their model focuses on the root causes and social processes that drive vulnerability and create the potential for unsafe conditions ("pressures") and how to reduce the fundamental vulnerability of the community ("releases"). A separate strand of the literature focuses less on the root cause of social vulnerability and much more on the interaction between the exposure to physical hazards as a result of the geographic location of a community and the community's social vulnerability. This "hazard of place" approach is outlined in Cutter (1996). More recently, authors like Turner et al. (2003) have used the framework of sustainability science to examine vulnerability, conceptualizing it as a coupled human–environment system and focusing attention on how to reduce vulnerability and build more adaptive and resilient communities. This approach considers the human and environmental conditions to be intertwined and interdependent and thus does not consider social vulnerability independent of physical conditions.

Even as the theoretical literature on social vulnerabilities continues to expand and evolve, there is concurrently increasing interest in measuring social vulnerability, however defined. One reason for this emphasis on measurement is that governments at all levels are increasing their attention to planning for and responding to natural hazards, particularly those associated with climate change. One of the key challenges in measuring social vulnerability is the wide range of approaches to social vulnerability and the lack of a consensus as to what exactly social vulnerability is. In this paper, we examine two different approaches to identifying and measuring social vulnerability using a case study of Hampton Roads, Virginia. The two methods examined are composite indexing using principal component analysis (PCA) and *k*-means clustering. The case study approach allows us to illustrate the strengths and weaknesses of these two methods using a concrete example, although the conclusions from our study should apply more globally.

We chose Hampton Roads as our case study because the area is one of the most threatened by sea level rise in the USA, and there are significant ongoing efforts to plan for

² 42 U.S.C. 5151.

and adapt to climate change. Hampton Roads stretches from Virginia's Atlantic coast up the shore of the Chesapeake Bay and into the interior of the state. The area encompasses the cities of Norfolk, Virginia Beach, Hampton, Newport News, and Chesapeake. It is home to several US military installations, including the world's largest naval base, and to one of the largest commercial ports on the East Coast. Relative to other coastal locations, Hampton Roads faces accelerated effects of sea level rise due to significant land subsidence and the prevalence of low-lying flood-prone terrain in the Chesapeake basin (Virginia Institute of Marine Sciences 2013). Several eastern Virginia communities have already seen a significant increase in the severity and frequency of tidal flooding and sea level rise is expected to bring even more tidal flooding and damaging storm surges to the area in coming decades.

The area includes both highly urbanized and very rural communities as well as some of the richest and the poorest neighborhoods in the state. In order to develop effective programs to adapt and respond to the impacts of climate change and sea level rise that meet the equity requirements of the Stafford Act, policymakers must consider the needs of all of the region's residents and stakeholders, particularly those who are least able to provide for themselves in the event of disaster. It is this need which motivates this paper's critical examination and refinement of methods for measuring social vulnerability to climate change and sea level rise. While we are primarily interested in developing a method to quantitatively estimate social vulnerability to help local and state planners determine how to best marshal limited resources to adapt to sea level rise while ensuring that resources are targeted effectively and equitably, our analysis has implications beyond sea level rise and climate change. In particular, our examination of methods for quantifying social vulnerability should also apply to other types of natural and man-made hazards that result in significant damage to physical infrastructure and disruption of communication, transportation, and utility networks.

The remainder of the paper is organized as follows. The next section provides a survey of existing quantitative methods for identifying socially vulnerable populations with a focus on composite indexing using principal component analysis (PCA), one of the most commonly used methods of measuring social vulnerability. We then present an analysis of the strengths and weaknesses of PCA-based indexing and identify some caveats to its use. The third section presents our case study of Hampton Roads which compares PCA-based indexing to an alternative method that uses *k*-means clustering to identify socially vulnerable communities, considering the merits and limitations of the two methods. The paper concludes with a discussion of our findings as well as recommendations for researchers and practitioners looking to measure social vulnerability to climate change and sea level rise.

2 Measuring social vulnerability: SoVI[®] and beyond

In the mid-2000 s, a group of researchers at the University of South Carolina (USC) led by Susan Cutter developed an approach to quantifying social vulnerability to environmental change and natural disasters (Cutter et al. 2003). This approach, known as the social vulnerability index, or SoVI[®], uses principal component analysis to identify common factors underlying an assortment of potential vulnerability indicators. PCA is a statistical technique that takes as its input a matrix of interrelated socioeconomic variables—in this case those considered to measure various dimensions of social vulnerability—and creates a set of principal components that extract the important variation in the underlying input data

while reducing the noise and redundancy in the data. After conducting the PCA, the researcher combines the newly created component variables into a composite index that provides a single value for each observation in the dataset, in this case a social vulnerability score. The utility of a PCA-based index is that it encapsulates a lot of information in an easily consumed form and individual observations can be ranked relative to each other.

More specifically, prior to conducting a PCA, variables in the input matrix are typically standardized to z scores with zero means and unit variances to avoid any confounding effects that might arise from using variables of different magnitudes in the analysis. The PCA is run on the standardized input matrix and generates a new set of orthogonal principal components. The researcher then decides which principal components to integrate into the final index and how to do so. In particular, the researcher must determine the directionality of each retained component, that is, whether higher values of the component increase the index value (positive directionality) or decrease it (negative directionality). The components are then added together either with or without weights. To ensure that the indices are easily comparable to each other, most final indices are standardized to z scores with zero means and unit variances.³

A review of the existing academic literature on the construction of quantitative social vulnerability indicators reveals that since its creation, SoVI[®] in particular, and PCA-based composite indexing more generally, have become one of the most common paradigms for quantitatively identifying social vulnerability to environmental hazards, particularly for studies that focus on natural hazards associated with climate change such as hurricanes, storm surge, flooding, and coastal erosion. Consequently, a fairly robust academic literature using PCA-based indices has developed, as shown in Table 1.⁴ PCA-based indices are also widely used in tools directed at policy practitioners and planners. For example, the state of California has incorporated a version of SoVI[®] into its State Hazard Mitigation Plan (California Emergency Management Agency 2013) and the Pacific Institute, a non-profit research organization focusing on environmental issues uses a 19-variable census tract-level SoVI[®] replica for the state of California (Pacific Institute 2016). At the national level, the National Oceanic and Atmospheric Administration (NOAA) includes the county-level SoVI[®] as one of the available layers in its public sea level rise risk assessment mapping tool (NOAA 2016). Similarly, the nonprofit group Climate Central has included SoVI[®] in its Web-based application for identifying communities at risk from sea level rise (Climate Central 2016).

While SoVI[®] and related PCA-based indices have the advantage of replicability, transferability, and general acceptance in the academic literature, there are some caveats to using them to measure social vulnerability. First, the strength of the PCA method—that it reduces a complex matrix of information to a single number—is also one of its weaknesses, as the index can be difficult to interpret and the methodology for computing the index is not transparent to most users of the index. More specifically, as we discuss below, users are unlikely to understand how the selection of input variables, the use of component selection and weighting criteria, and considerations of scale influence the resulting index.⁵ Additionally, while the PCA process results in a unique set of components, the researcher must

³ Abdi and Williams (2010) provide a more complete description of the PCA method.

⁴ While this list is not exhaustive, it does provide a large cross section of the literature.

⁵ This is certainly not the first paper to point this out. For example, Tate (2012) and Tate (2013) also discuss how these factors, as well as a number of other factors, affect the construction of a composite index. However, our approach to comparing PCA-based composite indexing to a cluster analysis approach focusing on these factors is, to the best of our knowledge, unique.

Table 1 Survey of SoVI® and PCA-based indices

References	Hazard type	Spatial extent	Spatial granularity	# of input variables	Component retention	Component weighting	High-vulnerability criterion
Cutter et al. (2003)	General environmental	USA	County	42	Kaiser (Eigenvalue > 1)	Equal	>1 SD above mean
Boruff et al. (2005)	Coastal erosion	USA	County	39	Eigenvalue > 0.95	Equal	>1 SD above mean
Rygel et al. (2006)	Hurricane (storm surge)	Hampton Roads, VA	Census block group	57	Kaiser	None	>1.5 SD above mean
Burton and Cutter (2008)	Flooding (levee failure)	Sacramento-San Joaquin River Valley, CA	Sub-county	36	Kaiser	Equal	>1 SD above mean
Cutter and Finch (2008)	General environmental	USA in 1960–2000	County	42	Kaiser	Equal	>1 SD above mean
Myers et al. (2008)	Hurricane Katrina, Hurricane Rita	US Gulf Coast	County	24	Unspecified	None	>1 SD above mean
de Oliveira Mendes (2009)	General environmental	Centre Region, Portugal	Municipality	50	Kaiser	Equal	>1 SD above mean
Burton (2010)	Hurricane Katrina	Mississippi Gulf Coast	Census tract	32	Kaiser	Equal	Top 20 percent of tracts
Finch et al. (2010)	Hurricane Katrina	Orleans Parish, LA	Census tract	29	Kaiser	Equal	>0.5 SD above mean
Tate et al. (2010)	General environmental	Charleston, SC	Census block group	32	Kaiser	Unspecified	Unspecified
Wood et al. (2010)	Tsunami	Pacific Northwest USA	Census block	29	Kaiser	Variance weighting	>1 SD above mean
Bjarnadottir et al. (2011)	Hurricane (storm surge)	Miami-Dade County, FL in 2050	Municipality	16	Kaiser	Expert-assigned	None
Schmidlein et al. (2011)	Seismic	Charleston, SC metropolitan area	Sub-county/census tract	26	Kaiser	Equal	None
Solangaarachchi et al. (2012)	Bushfire	Southeastern Australia	Census Collection District	26	Kaiser	Variance	>1.5 SD above mean

Table 1 continued

References	Hazard type	Spatial extent	Spatial granularity	# of input variables	Component retention	Component weighting	High-vulnerability criterion
Wang and Yarnal (2012)	Hurricane	Sarasota, FL	Census block group	64	Kaiser	None	Individual components > 1.5 SD above mean
Armaş and Gavriş (2013)	Earthquake	Bucharest, Romania	Census tract	22	Kaiser	Equal	>1 SD above mean
Cutter et al. (2013)	Flooding	Southeastern USA	Census tract	32 and 9	Unspecified	Unspecified	Unspecified
Holand and Lujala (2013)	General environmental	Norway	Municipality	37	Kaiser	Equal	>1 SD above mean
Martinich et al. (2013)	Sea level rise	US coastal counties	Census tract	26	Unspecified	Equal	>1.5 SD above mean
Toké et al. (2014)	Seismic	Los Angeles, CA	Census block group	20	Unspecified	Unspecified	None

ultimately make subjective decisions about how to aggregate those components into the final index, which creates the potential for two researchers to develop different indices from the same underlying set of input data.

2.1 Complexity and interpretability of PCA-based indices

One of the key challenges facing practitioners who wish to capture the dimensions of social vulnerability quantitatively is that vulnerability depends upon many interacting factors and the specific context of a particular situation or hazard type. Attempting to account for a wide range of distinct factors by including them explicitly in a single composite index introduces complexity into the model, which can make the resulting index difficult to interpret and apply. As shown in Table 1, the number of variables included in PCA-based indices has ranged from a low of 8 to a high of 64 with most studies using between 25 and 40 variables. Technically, PCA can successfully capture subtle variations in the input data, which may encourage researchers to take an inclusive approach to collecting data that capture the many complex dimensions of social vulnerability. Existing analyses of both applied and simulated index construction, however, suggest that a PCA-based index constructed from many factors, most of which by themselves have weak explanatory power in predicting specific outcomes cannot necessarily distill meaningful value from fundamentally noisy data.⁶ Additionally, the larger the number of input variables, the more likely the index will suffer from unexpected correlation between input variables, bidirectional vulnerability signals (where both high and low values of a variable indicate vulnerability), and differences in the true measures of vulnerability across heterogeneous populations. These types of interactions, which we refer to collectively as confounding variable effects, may bias the final index in many different ways.

For example, a variable measuring the proportion of mobile homes in a community may be an important indicator of vulnerability in rural areas, but including a principal component in the final index that is heavily influenced by that variable will result in a lower social vulnerability index for urban areas where mobile homes are not an important concern. Alternatively, including measures such as median age where vulnerability is a concern at both tails of the distribution can result in a noisy signal. While there are solutions to this, such as Fekete's (2012) recommendation to use binary indicators for extreme values of specific variables that signal potential vulnerability on both ends of the scale, many indices continue to use these types of variables. By limiting the input variables to only those that capture or are strongly correlated with the social vulnerability factors of greatest interest based on review of the relevant literature and engagement with local communities, researchers can minimize potential distortions from confounding variable effects. However, as shown by the large number of variables used in most of the indices listed in Table 1, the tendency seems to be to err on the side of inclusion rather than exclusion.

One argument for including a large set of variables in a PCA-based index is "to avoid both pre-principal components analysis data loading and ex post-data adjustment that could impart a bias in the results used for decision making" (Cutter et al. 2013, p. 342). However, subjective considerations of data loading are unavoidable in constructing a PCA-based index as such indices necessarily reflect the relative influence of the variables that were

⁶ For example Sherrieb, Norris and Galea (2010) conduct a correlation analysis of existing social vulnerability composite indices using county-level data from the state of Mississippi find a relatively low correlation of these indices with survey measures of social cohesion (correlation -0.17) and collective efficacy (correlation -0.10).

inputs into the PCA process. As Burton and Cutter (2008) note, composite indices “do not always capture the most important or ideal factors, but rather those that best explain the variation in input data” (p. 141). Thus the contribution of a particular dimension of vulnerability to the overall index depends in part on how many indicators of that dimension are included in the input matrix. For example, in Baum et al. 2008 study, only two out of 19 input variables measure income and thus income-related measures explain only a small share of the total variance in the input data and ultimately have a relatively small impact on the final index. On the other hand, when financial security is measured by the selection of many input variables such as income, home value, rent, and poverty level, financial security is likely to have a larger effect on the final index. Ultimately, choosing to include a variable in an index can alter the results in the same ways that excluding a variable can. Unfortunately, in many situations it can be difficult to come up with an observable measure of social vulnerability against which to externally validate a researcher’s choice of variables. For example, while there might be a well-defined outcome measure for some hazards such as the deaths associated with heat waves, in the case of recurrent flooding and storm surges from sea level rise one is likely interested in a range outcomes including deaths, changes in physical and mental health, property damage, increased cost of home maintenance, and lost wages, some of which might be quite difficult to measure.

In addition to potentially weakening the vulnerability measure, increasing the number of input variables to capture a wide range of vulnerability concerns also reduces the interpretability of the composite index as it becomes very difficult to trace how changes in a single factor influence the overall index. Thus, researchers and practitioners should be aware of all of the possible trade-offs associated with the inclusion of additional measures of social vulnerability and need to convey those tradeoffs to users of the index.

2.2 Component selection and weighting criteria

Another key consideration researchers must make in constructing a PCA-based composite index is which output components will be used in the index and how they will be weighted in the aggregation. As shown in Table 1, most indices we surveyed use the Kaiser selection criterion, in which components from the PCA are incorporated into the final matrix if they have an eigenvalue greater than 1.0 (i.e., the component explains more than $1/n$ th of the total variance of the input matrix, where n is the total number of input variables). Most indices then equally weight the retained components in the final index. Finch et al. (2010) defend equal factor weights as the most logical method for aggregating the retained components into the final index because “the purpose of the SoVI index generally...is to provide an aggregate measure of *all* the factors that contribute to the social vulnerability of the parish, not just one or two individual variables or components” (p. 184). We believe that weighting each factor according to its share of explained variance in the sample is preferable to equal weighting for two primary reasons. First, if one selects and combines the input variables in proportion to perceived overall contribution to vulnerability, variance weighting will preserve this choice while equal weighting will not. Second the Kaiser selection criterion used by most researchers is inconsistent with equal weighting. That is, if the researcher believes that the explained variance is relevant enough to merit its use as a component selection criterion, it should also be relevant enough to merit its consideration in weighting those components. Moreover, our case study, which is discussed in detail in

Sect. 3, shows that variance-weighted indices are more consistent when there are changes in the geographic scope of analysis and in the input variables than equal-weighted indices.⁷

2.3 Scale

Another issue that researchers and practitioners should consider when constructing a vulnerability measure is scale. Scale includes both the level of observation (e.g., census tract vs. county) as well as the overall geographic boundaries of the analysis. For some data sources such as the American Community Survey, data for the smaller geographies (census block groups), are more noisy than data at larger geographies (tracts and counties), due to the sampling method used to collect data. Moreover, some data sources may only have information for larger geographies. With respect to the geographic area spanned by the analysis, for the most part that is driven by the nature of the study. However, trying to capture the important dimensions of vulnerability when the factors that contribute the most strongly to vulnerability vary significantly across observations increases the likelihood of confounding variable effects, and larger geographic areas are more likely to face this problem than smaller and more homogeneous ones. Our case study examines the consistency of various PCA-based indices to variations in geographic boundaries, but not to variation in the granularity of the analysis.⁸ In our case study presented in Sect. 3, we show that PCA-based indices are sufficiently sensitive to the drawing of a different set of boundaries to confound their utility even at relatively small changes in geographic boundaries, such as from Hampton Roads census tracts to Coastal Virginia census tracts.⁹

2.4 Alternatives to PCA-based indexing

Several alternatives to SoVI[®] and PCA-based indices have been offered in the literature on social vulnerability to natural hazards. Generally these alternatives fall into three categories: separate consideration of individual vulnerability measures; simple aggregation of a small set of variables; and measurement of similarity to a reference district determined to be vulnerable. As an example of the first type, Wang and Yarnal (2012) consider individual factors that contribute to social vulnerability separately and map them to identify potential vulnerability hot spots. The method thereby avoids the problem of weighting factors, but at the cost of added complexity in the final measure and reduced transferability to other regions for comparison and validation. Additionally, this method does not allow one to rank locations relative to each other in terms of vulnerability.

As an example of the second category, Chakraborty et al. (2005) construct a composite Social Vulnerability for Evacuation Index (SVEAI), which is a simple additive aggregation of standardized variable values for a set of 10 social indicators thought to represent

⁷ Tate (2013) and Wolters and Kuenzer (2015) discuss a third option, the use of parametric indicators or component weights assigned based on parameters assessed from outside information such as expert opinion or econometric analysis. While expert-derived or parametrically-weighted indices might offer greater validity than other approaches to weighting, they have disadvantage of requiring additional analysis or researcher judgment. Thus we chose to focus on the two weighting options that do not rely on additional information or inputs.

⁸ Tate (2012) shows that the granularity of the analysis does have a significant effect on PCA-based indices.

⁹ Cutter et al. (2013) also investigate the consistency of various PCA-based indices under changes in geographic boundaries. However, they examine applying the same index construction method to distinctly different geographic areas while we consider applying the same method to different overlapping geographic boundaries.

vulnerability dimensions for assessment of hurricane evacuation planning in Hillsborough County, Florida. As in other composite indices of social vulnerability, differential access to resources, as measured by both poverty and specific indicators of communication and transportation access, is an important consideration and serves as a proxy for a great deal of other correlated vulnerability signals. As is the case with PCA-based index, this method does provide a means for ranking observations relative to each other. As an example of the third approach, Chang et al. (2015) present a composite social vulnerability indicator meant to assess the similarity of administrative districts in British Columbia along a several different dimensions of vulnerability. Their Hazard Vulnerability Similarity Index (HVSI) measures the similarity of districts in their analysis to a chosen reference district in order to facilitate knowledge sharing among areas facing similar issues. The cluster analysis method we apply in our case study builds on this similarity approach, using a data mining technique to group similar communities within Hampton Roads.

3 Hampton Roads case study

In this section we present a case study for measuring social vulnerability in Hampton Roads, Virginia. As discussed in the introduction, Hampton Roads is under serious threat from sea level rise and climate change. As localities work on adaptation strategies and mitigation plans, there is pressure to ensure that these plans are informed not only by physical vulnerability, but also by social vulnerability to natural hazards. In this case study we compare a PCA-based approach to quantifying social vulnerability to an alternative method that uses *k*-means clustering to group similar communities. The resulting clusters are then evaluated with respect to predictors of social vulnerability to determine whether the communities in each cluster are socially vulnerable. In our comparison of the two methods, we focus primarily on the consistency of the methods under changes in the set of input variables and geographic scope. In addition we consider the consistency of the PCA-based indices under changes in the weighting method.¹⁰

For the purposes of local planning for climate change and sea level rise, it is important to identify socially vulnerable communities at a reasonably small geographic scale. However, as is the case with most quantitative analyses of demographic data, the level of spatial aggregation often depends on the data available, rather than a methodological decision about the appropriate level at which to measure social vulnerability. This study is no different. In order to use a wide range of variables that have been identified by other studies as important indicators of social vulnerability, the smallest geographical unit that we feel comfortable using are census tracts.¹¹ For this analysis we consider the 1874 census tracts in Virginia that have at least 100 or more inhabitants. Including census tracts across all of Virginia allows us to explore the extent to which using more observations as well as a more diverse set of observations to construct the vulnerability measure affects the consistency of that measure.

¹⁰ Since the cluster analysis approach does not use weights, we do not compare the methods on this dimension.

¹¹ Some of the key variables used in the analysis are available only from the American Community Survey (ACS). Because the ACS samples approximately 1 in 40 households every year, the error bounds for smaller geographies such as block groups are subject to significant sampling error.

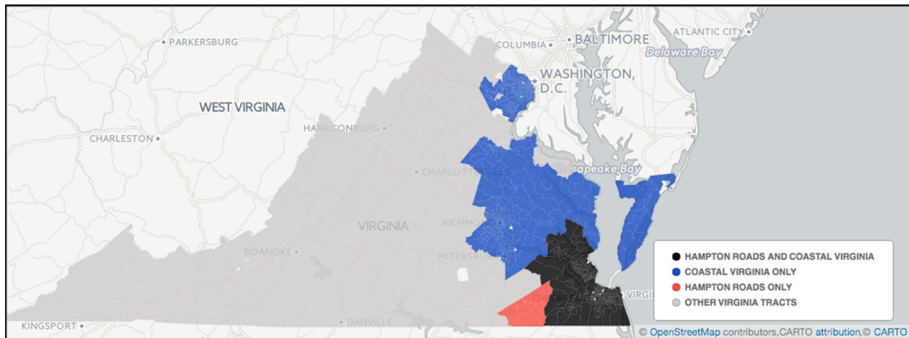


Fig. 1 Map of all Virginia, Coastal Virginia, and Hampton Roads census tracts

3.1 Construction and evaluation of the PCA-based indices

For this case study we constructed eighteen different PCA-based composite social vulnerability indices that vary across three dimensions: the geographic boundaries of the analysis, the set of input variables, and the weighting of the components generated by the PCA. With respect to geographic boundaries, we consider three different sets of census tracts as shown in Fig. 1: all 1874 census tracts in Virginia with at least 100 inhabitants; the 1172 of those census tracts in Virginia’s coastal counties; and the 408 of those census tracts in the Hampton Roads area. We consider three different sets of input variables: The largest set includes 41 variables chosen to replicate the original Cutter et al. (2003) SoVI[®] index, measured at the census tract-level where available and at the county level otherwise; the next set includes only the 29 variables from the first set that can be measured at the census tract-level; and the smallest set includes 13 variables, all census tract-level, considered by us to be the most direct determinants of social vulnerability, including measures accounting for income, race, ethnicity, age, and gender.¹² Table 2 presents the variables used in each set.¹³ Finally, we considered two different methods for aggregating the PCA-derived components: an equal-weighted index which gives the same weight to each of the PCA output components used in the analysis and a variance-weighted index that weights each component according to the total variance it captures from input variables. While the Cutter et al. (2003) set of input variables were purposefully constructed to measure social vulnerability to general environmental hazards and our selection of the limited set of 13 variables was made in the context of sea level rise and climate change, it is likely that similar sets of variables would be appropriate for quantifying social vulnerability to other types of natural and man-made hazards. Thus we expect the insights from our analysis to apply broadly to most PCA-based quantifications of social vulnerability.

Before conducting the PCA, the variables were first standardized to z scores with zero means and unit variances, as is common in the literature, to avoid confounding effects that

¹² The Cutter et al. (2003) study used 42 variables; however, we were unable to find recent data on general local government debt to revenue ratios—thus we only include 41 of the original variables in our analysis.

¹³ We use 2010 as the base year for this study. Variables that were collected from a source other than the 2010 Census were collected as close to 2010 as possible.

Table 2 Variables used in the social vulnerability measures

Variable name	Source	Spatial granularity	Set 41	Set 29	Set 13
Median age	2010 US Census	Tract	X	X	
Per capita income	ACS 2006–2010	Tract	X	X	X
Median value of owner-occupied housing units	ACS 2006–2010	Tract	X	X	
Gross rent	ACS 2006–2010	Tract	X	X	
Physicians per capita	2004 AMA data	County	X		
Pct. voting Obama in 2008	MSNBC	County	X		
Birth rate per capita	VA state health district data	County	X		
Net internal migration per capita	Census population estimate	County	X		
Percent land in farms	2007 census of agriculture	County	X		
Percent population Black or African-American alone	2010 US Census	Tract	X	X	X
Percent population American Indian and Alaska Native alone	2010 US Census	Tract	X	X	X
Percent population Asian alone	2010 US Census	Tract	X	X	
Percent population Hispanic or Latino	2010 US Census	Tract	X	X	X
Percent population under 5 years old	2010 US Census	Tract	X	X	
Percent total population over 65 years old	2010 US Census	Tract	X	X	X
Percent unemployed in civilian force age 16 and over	2010 US Census	Tract	X	X	X
Average household size	2010 US Census	Tract	X	X	
Percent of households earning more than \$75,000/year	ACS 2006–2010	Tract	X	X	
Percent population below poverty line	ACS 2006–2010	Tract	X	X	X
Percent housing units occupied by renters	ACS 2006–2010	Tract	X	X	
Number of farm operators per capita	2007 census of agriculture	County	X		
Percent housing units that are mobile homes	ACS 2007–2011	Tract	X	X	
Percent population age 25 and over with no HS degree	ACS 2006–2010	Tract	X	X	X
Housing units per square mile	2010 US Census	Tract	X	X	
Private housing building permits	2010 Building Permits Survey	County	X		
Manufacturing establishments	2007 economic census	County	X		
Sales, receipts, and value of shipments for all firms	2007 economic census	County	X		
Number of firms	2007 economic census	County	X		
Farm income	2007 economic census	County	X		
Percent population 16 and over in the labor force	ACS 2006–2010	Tract	X	X	
Percent females 16 and over in civilian labor force	ACS 2006–2010	Tract	X	X	X

Table 2 continued

Variable name	Source	Spatial granularity	Set 41	Set 29	Set 13
Percent population employed in extractive industries	ACS 2006–2010	Tract	X	X	
Percent population employed in transport, utility, information	ACS 2006–2010	Tract	X	X	
Percent population employed in services (education, arts, other)	ACS 2006–2010	Tract	X	X	
Percent population in nursing homes	2010 US Census	Tract	X	X	X
Hospital beds per capita	2004 American Hospital Association	County	X		
Percent population change 2000–2010	2010 and 2000 US Census	Tract	X	X	
Percent urban population	2010 US Census	Tract	X	X	
Percent population that is female	2010 US Census	Tract	X	X	X
Percent households: with female householder, no husband	2010 US Census	Tract	X	X	X
Percent households with Social Security income	2010 US Census	Tract	X	X	X

Table 3 Results of the principal component analyses

	Geography	Input set	Retained components	Variance explained by retained components (%)
VA41	All of VA	41	10	74
VA29	All of VA	29	7	70
VA13	All of VA	13	4	69
CV41	Coastal VA	41	10	75
CV29	Coastal VA	29	8	72
CV13	Coastal VA	13	4	72
HR41	Hampton Roads	41	10	77
HR29	Hampton Roads	29	8	75
HR13	Hampton Roads	13	4	74

can occur from using variables of different scales.¹⁴ For each combination of geographic area/input set we conducted a PCA, keeping those principal components with eigenvalues greater than 1 (the Kaiser selection criterion) as most of the studies cited in Table 1 do. The results of the PCA process are presented in Table 3. As expected, more complex input sets are reflected in more retained components. Since each component contributes to the final index, indices with more components are harder to interpret. Also note that while additional retained components do increase the total amount of variance explained, the difference in variance explained by the retained components does not decrease substantially

¹⁴ All variables were standardized relative to the means for all the 1874 census tracts in Virginia. As a sensitivity analysis, we examined whether it made a difference if the variables were standardized with respect to all Virginia census tracts or to the particular geography used for that index (e.g., coastal Virginia or Hampton Roads) and found little difference in the results.

even when the number of variables decreases substantially. The geographic area does not have a significant effect on the number of components, although the retained components do explain more variance when the geographic area is smaller and less heterogeneous.

As a next step, we conducted a Varimax rotation of the components to facilitate interpretation of each component because—as is the case with all PCA-based indices—we must determine the directionality of each retained component, that is whether higher values of the component increase the level of social vulnerability (positive directionality) or decrease the level of social vulnerability (negative directionality). Where the directionality of the component was clearly negative, we scaled the component by a factor of -1 before including it in the composite index so that higher values of the scaled component would increase the overall vulnerability index. Following Cutter et al. (2003), in instances when the effect of the component on vulnerability is ambiguous (as is the case when the different variables that make up the component work in opposite ways or when a variable exhibits bidirectionality), we assume a positive directionality. For the equal-weighted indices, the components are added together without weights while for the variance-weighted indices the components are multiplied by the variance each component captures from the total input matrix before aggregation. To ensure that the indices are easily comparable to each other, we standardize the resulting aggregated values to z scores with zero means and unit variances.

Following Cutter et al. (2013) we compare the indices using the Pearson's R correlation coefficient. Correlations above 0.90 indicate a strong positive correlation between the two indices while correlations below 0.70 indicate substantial inconsistencies.¹⁵ As one might expect the various equal-weighted indices are highly correlated with the analogous variance-weighted indices. Seven of the nine pairs of indices have correlation coefficients above 0.90 while one pair has a coefficient on 0.82 (the 29-variable index for Hampton Roads) and one has a coefficient of 0.75. However, when we compare the indices based on geographic area or input set, overall the correlations are much lower. Table 4 shows the Pearson's R coefficient for pairwise correlations among the various PCA-based indices that we constructed. The top half of the table shows the correlations across equal-weighted indices, while the bottom half shows the correlations across variance-weighted indices. For the equal-weighted indices, most pairwise comparisons result in a correlation coefficient below 0.70. There are two general exceptions—the coastal Virginia and Hampton Roads indices that use the smallest set of input variables (CV13 Eq and HR13 Eq)—are reasonably correlated with all of the overall Virginia indices and the HR13 Eq index is reasonably correlated with the CV41 Eq index and highly correlated with the CV13 Eq index. Overall the variance-weighted indices show higher correlations than the analogous pairs of equal-weighted indices, with the CV13Var and HR13Var indices showing the highest correlations. Thus the 13-variable indices show greater consistency under changes in geographic boundaries than the indices using larger numbers of variables. The 29- and 41-variable indices are much more inconsistent, and in one case the correlation between pairs of indices using 29 variables is actually negative, demonstrating the cumulative effects of confounding variables. In particular, we note that both the 29- and 41-variable sets contain potential bidirectional variables such as age, percent urban, and net internal migration while the 13-variable set does not. Additionally, we note that using variance-weighted principal components provides a more consistent measure across both different geographic scales and different input sets.

¹⁵ Cutter et al. (2013) also report the Cronbach alpha test for pairwise comparisons of indices. For our indices the Pearson's R correlation and the Cronbach alpha test provide the same results.

Table 4 Pearson's *R* correlation for the PCA-based indices

	VA41 Eq	VA29 Eq	VA13 Eq	CV41 Eq	CV29 Eq	CV13 Eq	HR41 Eq	HR29 Eq	HR13 Eq
Pairwise comparison of equal-weighted indices									
VA41 Eq	1								
VA29 Eq	0.69	1							
VA13 Eq	0.54	0.67	1						
CV41 Eq	0.17	0.25	0.27	1					
CV29 Eq	0.48	0.63	0.46	0.42	1				
CV13 Eq	0.72	0.80	0.87	0.44	0.50	1			
HR41 Eq	0.02	0.04	0.45	0.42	0.10	0.35	1		
HR29 Eq	0.30	0.10	0.36	0.35	-0.20	0.33	0.03	1	
HR13 Eq	0.81	0.84	0.80	0.75	0.56	0.95	0.21	0.44	1
VA41Var	VA29Var	VA13Var	CV41Var	CV29Var	CV13Var	HR41Var	HR29Var	HR13Var	
Pairwise comparison of variance-weighted indices									
VA41Var	1								
VA29Var	0.64	1							
VA13Var	0.60	0.79	1						
CV41Var	0.21	0.59	0.59	1					
CV29Var	0.54	0.72	0.63	0.52	1				
CV13Var	0.76	0.85	0.97	0.64	0.60	1			
HR41Var	0.20	0.29	0.54	0.53	0.22	0.57	1		
HR29Var	0.61	0.51	0.75	0.66	0.19	0.76	0.37	1	
HR13Var	0.82	0.88	0.96	0.92	0.62	0.99	0.48	0.77	1

EqVar, index calculated using equal-weight or variance-weighted components
 VA/CV/HR, index calculated for all 1874 Virginia tracts, 1172 Coastal Virginia tracts, or 408 Hampton Roads tracts
 41/29/13, number of variables used in calculating index components

Looking at the correlations between indices is not the only way to compare them. In addition to using a PCA-based index to rank observations relative to each other in terms of vulnerability, many studies also use their index to identify a group of “highly vulnerable” observations. Various papers have implemented different decision rules for determining which observations are considered to be highly vulnerable, as shown in the last column in Table 1.¹⁶ By far the most common method is to use a threshold value calculated using the mean and standard deviation of the index. Often, but not always, the threshold value is 1 standard deviation above the mean, which is the threshold we use for this analysis. Table 5 compares the number and percent of census tracts that are designated as highly vulnerable using this decision rule for the various indices in our case study. We also include a second category of “marginally vulnerable” tracts. These are defined as tracts with an index value between 1 and 0.5 standard deviations above the mean.

The percentage of tracts identified as highly vulnerable ranges from a low of 12 % to a high of 17 %. This relative consistency is expected given the nature of the decision rule and the fact that the index values are standardized with a mean of 0 and a standard deviation of 1. When we include marginally vulnerable tracts, the percentage ranges from 25 to 31 %. For most geography/input set combinations, the equal-weighted and variance-weighted indices identify roughly the same number of highly vulnerable or marginally vulnerable tracts. However, when one considers the overlap of tracts identified as highly or marginally vulnerable across the two weighting schemes (reported in columns three and six of Table 5), it is clear that while there is significant overlap, it is not perfect. Looking at the overlap of highly vulnerable tracts across indices that share the same geography and weighting (rows four, eight, and twelve), the number of consistently identified tracts is even lower. As was the case for the pairwise correlations, the variance-weighted indices appear to be somewhat more consistent under changes to the number of input variables than the equal-weighted indices.

A final comparison looks at the effect of geographic boundaries on the identification of highly vulnerable and marginally vulnerable tracts. For this comparison, we only consider the 401 counties that are in both the Hampton Roads and Coastal Virginia geographies.¹⁷ Using the variance-weighted indices for the three different geographies, we determined how many census tracts are consistently identified as highly vulnerable for the various input sets. When the 41 variable input set is used, only 11 of the 401 tracts or 3 % are consistently identified as highly vulnerable across the three different geographic boundaries, even though typically between 10 and 15 % of tracts are identified as highly vulnerable for a given index. When we compare the indices for the 29 variable input set, 24 of the 401 (6 %) tracts are consistently identified as highly vulnerable and when we compare the indices for the 13 variable input set, 65 of the 401 (16 %) tracts are. Repeating the same exercise but counting tracts that are classified as either highly or marginally vulnerable, for the 41 variable input set, 34 of the 401 tracts or 8 % are consistently identified as highly or marginally vulnerable across the three different geographic boundaries compared to ranges of 25–31 % of tracts for a given index. For the 29 variable input set, 49 of the 401 (12 %) tracts are consistently identified as highly or marginally vulnerable and when we compare the indices for the 13 variable input set, 100 of the 401 (25 %) tracts are.

¹⁶ Some papers categorize observations into bins representing different levels of vulnerability. For those studies, we consider how the most vulnerable group is categorized.

¹⁷ This excludes 7 census tracts in city of Franklin and county of Southampton that are considered part of Hampton Roads, but are not in a coastal county.

Table 5 Number and percent of tracts determined to be highly or marginally vulnerable^a by PCA-based indices

Geography/ input set	Highly vulnerable tracts			Highly or marginally vulnerable tracts		
	Equal- weighted index	Variance- weighted index	Overlap between the two indices	Equal- weighted index	Variance- weighted index	Overlap between the two indices
VA41	245 (13 %)	248 (13 %)	190 (10 %)	512 (27 %)	504 (27 %)	425 (23 %)
VA29	270 (14 %)	255 (14 %)	202 (11 %)	546 (29 %)	586 (31 %)	437 (23 %)
VA13	273 (15 %)	289 (15 %)	255 (14 %)	532 (28 %)	536 (29 %)	492 (26 %)
Overlap across all VA indices	107 (6 %)	127 (7 %)	82 (4 %)	258 (14 %)	267 (14 %)	215 (11 %)
CV41	180 (15 %)	189 (16 %)	106 (9 %)	322 (27 %)	315 (27 %)	211 (18 %)
CV29	194 (17 %)	179 (15 %)	140 (12 %)	356 (30 %)	360 (31 %)	296 (25 %)
CV13	177 (15 %)	175 (14 %)	145 (12 %)	299 (26 %)	291 (25 %)	250 (21 %)
Overlap across all CV indices	27 (2 %)	69 (6 %)	15 (1 %)	91 (8 %)	145 (12 %)	58 (5 %)
HR41	58 (14 %)	59 (14 %)	51 (13 %)	106 (26 %)	102 (25 %)	89 (22 %)
HR29	49 (12 %)	61 (15 %)	30 (7 %)	112 (27 %)	109 (27 %)	77 (19 %)
HR13	60 (15 %)	70 (17 %)	49 (12 %)	111 (27 %)	107 (26 %)	96 (24 %)
Overlap across all HR indices	8 (2 %)	20 (5 %)	6 (1 %)	32 (8 %)	50 (12 %)	29 (7 %)

^a Highly vulnerable tracts have an index value greater than 1 SD above the mean. Marginally vulnerable tracts have an index value between 0.5 and 1 standard deviation above the mean

Overall, our findings indicate that the 13-variable indices are more consistent measures of social vulnerability. Additionally, using the 13-variable index eases the interpretation both of each PCA output component and of the overall index itself as each of the components map to specific input variables. In our case study, the PCA process for the 13-variable index results in an income/race component explaining about 30–40 % of total variance, as well as age, ethnicity, and gender components that each explain about 10–15 % of the total variance. Larger input matrices result in component scores that are difficult to interpret because of the larger number of inputs loading onto each component, particularly for components with lower eigenvalues where meaningful patterns in variables loading onto each component become more difficult to identify. Using smaller numbers of input variables reduces both statistical noise and subjective influence in interpreting components and hence in calculating a PCA-based index. Of course, reducing the number of variables included in the index does affect its ability to measure multi-dimensional social vulnerability. Thus the researcher needs to carefully weigh the trade-offs when deciding which variables to include or exclude from a particular index, as increasing the dimensions of the input matrix is not a costless decision.

3.2 Description and evaluation of the *k*-means clustering approach

Given the limitations associated with using a single index variable—however constructed—to capture such a multi-dimensional concept as social vulnerability, we

examined the potential for using cluster analysis as alternative method for identifying socially vulnerable communities.¹⁸ Like the PCA technique, cluster analysis also takes a large matrix of data and reduces it to a more manageable and arguably more meaningful measure.¹⁹ However, rather than reducing the data matrix to a single ordinal value, cluster analysis groups “like” observations into clusters that share common characteristics. The researcher can then look at the characteristics of each cluster and determine whether tracts in that cluster are socially vulnerable. One of the advantages of this approach is that it identifies tracts that may be vulnerable in only one or two dimensions. It also allows factors to be considered holistically so that a factor that may not contribute to vulnerability in urban locations but would in rural locations can be included in the analysis. Similarly, if the concern is about the health implications of increased storm surge or recurrent flooding one may be interested in different factors than if one is concerned about financial implications of sea level rise. Of course, cluster analysis does have its own limitations. Cluster analysis certainly requires the researcher to make subjective judgments about the relative importance of different factors and how such factors might interact to increase or decrease social vulnerability to a particular hazard. Additionally the clustering process is not perfect—if a researcher uses too few clusters, the clusters may not be similar enough to result in accurate assessment. On the other hand, if the researcher uses too many clusters, the identification process becomes unwieldy. Finally the clustering process is more time intensive than construction of a PCA-based index, both in processing time and the time required for a researcher to evaluate each cluster. Thus for projects with large set of observations to evaluate, such as all counties in the USA, it may not be feasible to conduct.

At the most basic level, the underlying objective of cluster analysis is to group observations so that the observations in each cluster are more similar to other observations in that cluster than to objects in other clusters. Each cluster is represented by a prototype. The exact nature of the prototype depends on the particular clustering algorithm used. For our analysis, we use a *k*-means clustering algorithm where *K* is the number of clusters that are identified in the data and is a parameter in the analysis chosen by the researcher. In this algorithm, *K* observations are randomly selected as initial prototypes. Each observation is assigned to the cluster that has the “closest” prototype where proximity is measured as Euclidean distance between the observation and the prototype. Once observations have been assigned to an initial cluster, the algorithm computes the centroid, or mean of the observations, for that cluster and uses that as the cluster’s new prototype. The observations are reassigned to the closest prototype and the centroid is recomputed. This process is repeated until neither the centroids/prototypes nor the assignments change.

K-means clustering has the advantage that it is relatively straightforward on a conceptual level. Of course there are some drawbacks to using this method. First, randomly choosing the initial prototypes can result in selection of a local, rather than global, minimum sum of squared distances. Thus results are not always replicable when repeated with a different initial selection of prototypes. The typical solution to this, which we do in our analysis, is to perform multiple runs with different initial prototypes and select the one with the smallest sum of squared distances (SSD) (the squared distance between each

¹⁸ Another alternative we initially considered was the data envelopment analysis (DEA) method that Clark et al. (1998) use to combine common factors into a single scalar, but which also allows for observations that have high values on only one component to score highly on the vulnerability index. One could characterize the DEA method as a hybrid of the PCA and cluster method as DEA uses components produced by an initial factor analysis as inputs, rather than individual variables. Thus some of the concerns about PCA-based approaches also apply to the DEA method.

¹⁹ See Mirkin (2013) for a thorough discussion of cluster analysis.

observation and its assigned prototype distance). To determine how many runs to perform, we examined the improvement in the clustering process that resulted from additional runs and determined that after 500 runs, the likelihood of a significant improvement in the clustering process was quite small.²⁰

A second concern is the clustering might not be able to identify sufficiently homogeneous clusters either because too few clusters have been specified by the researcher or because the underlying observations are too heterogeneous to be clustered. If that is the case, clustering can result in mischaracterized observations. To determine the optimal number of clusters we used the η^2 measure to assess the improvement in the clustering process that resulted from additional clusters (Makles 2012). For each of the nine different geographic boundaries/inputs set combinations discussed in the prior section, we began with 20 clusters and computed η^2 for the run with the smallest SSD. We then increased the number of clusters by 5 to 25, completed 500 runs, calculated the η^2 for the best run where $k = 25$. As long as the increase in η^2 exceeded 2 % points we kept increasing the number of clusters by 5 until doing so increased η^2 by less than 2 % points. Thus the number of clusters used for each geography/input set is the highest multiple of 5 that increases η^2 by at least 2 % points. For VA41, VA29, CV41, HR41, and HR29 we use 30 clusters, for CV29 we use 35 clusters, and for VA13, CV13, and HR13 we use 25 clusters.²¹

As was the case for the PCA-based indices, before conducting the cluster analyses we first standardized the variables relative to all of the Virginia census tracts. However, in evaluating a particular cluster's profile to determine whether it was socially vulnerable or not, we examined both the absolute and standardized values of the variables. Thus we know not only the true level of the variable for each prototype but also how that relates to the average for Virginia. For example, consider the three prototypes presented in Table 6. What is shown are the absolute values for each variable. The prototype for Cluster A exhibits very low per capita income and very high poverty and unemployment rates. This area is a densely populated urban community and contains a large number of renters. Both housing values and rents are lower than average. There is low labor force participation overall relative to the mean for Virginia, although this is not a particularly elderly community, and over a third of adults do not have a high school degree. A very high percentage of the community is black and many families have female head-of-households. For all of these reasons, we classify this cluster as a vulnerable cluster. The prototype for Cluster B is a sparsely populated rural area although not many of the citizens in the community are farmers. The community has a moderately low per capita income, with just over 15 % of the community below the poverty line and almost 10 % unemployed. Education levels are low as over a quarter of adults do not have a high school degree. Labor force participation is relatively low, but this is consistent with the fact that almost one-fifth of the population is over 65 and almost 40 % receive social security benefits. While this prototype exhibits many factors that are indicative of social vulnerability, most of the levels are not extreme and thus we classify this cluster as marginally vulnerable. The prototype for Cluster C is a

²⁰ We used the η^2 measure to assess the improvement in fit that resulted from additional runs (Makles, 2012). We conducted a series of cluster analyses using the VA41 dataset, calculating η^2 after each 100 runs up to 1000 and then repeated the series 5 times. After 100 runs, η^2 ranged from 64.1 to 64.6 %. After 500 runs, η^2 ranged from 64.2 to 64.9 % with a maximum increase of 0.8 % points. After 1000 runs, η^2 ranged from 64.6 to 64.9 % and the maximum increase was 0.5 % points. We therefore decided that 500 runs were sufficient given the small probability that doubling the number of runs would result in a significant improvement in fit.

²¹ The results using the $\eta^2 > 2$ % points decision rule are very similar to the result we get if we chose 30 clusters for all 9 geography/input set combinations.

Table 6 Sample clusters for all of Virginia using all 41 variables

Variable	Cluster A	Cluster B	Cluster C
Median age	31	44	32
Per capita Income	\$11,257	\$20,647	\$28,702
Median house value	\$105,023	\$126,992	\$317,379
Median rent	\$541	\$628	\$1382
Physicians per capita	0.004	0.001	0.003
Pct. voting Obama in 2008	68 %	49 %	57 %
Birth rate per capita	0.016	0.009	0.015
Internal migration per capita	0.002	0.001	0.006
Pct. Farmland	0 %	7 %	1 %
Pct. Black	84 %	34 %	18 %
Pct. Native American	0 %	0 %	1 %
Pct. Asian	1 %	0 %	10 %
Pct. Hispanic	3 %	3 %	36 %
Pct. under 5	10 %	5 %	9 %
Pct. over 65	10 %	18 %	6 %
Pct. unemployed	22 %	9 %	6 %
Average household size	2.6	2.4	3.1
Pct. earning more than \$75 K	6 %	20 %	45 %
Pct. in poverty	41 %	16 %	11 %
Pct. renters	69 %	25 %	47 %
Pct. farmers	0 %	3 %	0 %
Pct. mobile homes	1 %	22 %	1 %
Pct. without HS degree	34 %	26 %	21 %
Housing density (per sq. mile)	1.9	0.0	2.4
Building permits (per sq. mile)	3.8	0.1	3.9
Mfg. establishments (per sq. mile)	1.7	0.0	0.9
Firm revenues (per sq. mile)	\$232,122	\$1852	\$243,909
Number of firms (per sq. mile)	121.3	2.9	189.6
Farm earnings (per sq. mile)	\$314	\$5890	\$9337
Pct. pop. in labor force	56 %	57 %	78 %
Pct. females in labor force	56 %	53 %	70 %
Pct. employment in extractive industries	0 %	4 %	0 %
Pct. employment in utilities	5 %	7 %	7 %
Pct. employment in services	46 %	30 %	30 %
Pct. in nursing homes	0.3 %	0.3 %	0.1 %
Hospital beds per capita	0.006	0.001	0.001
Pct. population change	−2 %	1 %	16 %
Pct. Urban	100 %	6 %	100 %
Pct. Female	55 %	50 %	49 %
Pct. female head of household	42 %	14 %	15 %
Pct. with social security income	28 %	38 %	14 %

dense urban area with relatively high levels of internal migration and a significant increase in population between 2000 and 2010. The area is characterized by a large Hispanic population compared to the rest of Virginia. Although per capita income is lower than the state average and only around a fifth of the adults have a high school education, there is a high level of labor force participation and unemployment is lower than average. The area has a lot of new building permits and high levels of economic activity. A relatively small percentage of the population is dependent on social security income. This cluster is classified as not vulnerable.

As shown by the examples above, a cluster approach clearly requires that the researcher use his or her judgment in deciding whether or not a prototype should be classified as socially vulnerable. While some might disagree whether or not a specific prototype represents a socially vulnerable community, the decision rules followed by the researcher can be laid out in the analysis and thus the rationale for classification and any potential biases and subjectivity will be transparent. For our analyses, we reviewed each cluster prototype and made a determination of vulnerability based on the prototype's full characterization, that is, all of the variables used in the cluster analysis. As noted above, we looked at both the absolute value of each variable as well as how that value related to mean level for the rest of Virginia. We then looked at commonalities across the highly and marginally vulnerable clusters to identify the minimum common characteristics across each cluster that could describe why particular clusters were categorized as highly or marginally vulnerable to sea level rise and climate change. We believe these rules could be useful for others looking to classify communities as socially vulnerable but note that context is important in determining which clusters fall into which categories, as what might indicate lack of representation or resources in Virginia today may or may not be the same in other geographic areas or at other times.

Ultimately, a cluster was categorized as highly vulnerable if its prototype met any one of the following conditions:²²

- Median income less than \$20 K, percent living in poverty greater than 20 %, AND either percent unemployment greater than 10 % OR labor force participation less than 66 % OR female labor force participation less than 50 % (lack of resources).
- Median income less than \$20 K, percent living in poverty greater than 20 %, percent of the population without a HS degree is more than 25 %, AND percent black greater than 50 % (lack of resources, lack of representation).
- Percent of the population over 65 is greater than 20 %, the percent of the population in nursing homes is greater than 2 %, the percent of the population receiving social security is greater than 25 % and median income is less than \$25 K (lack of resources, high need).

Marginally Vulnerable clusters are identified as clusters that are not categorized as vulnerable but where the prototype meets one of the following conditions:

- Percent of the population without a HS degree is more than 25 %, the percent of the population receiving social security is greater than 33 %, and the percent of housing units that are mobile homes is greater than 15 % (lack of resources, high need, vulnerable shelter).
- Percent of the population without a HS degree is more than 25 % and there are high populations of under represented races (relative to Virginia as a whole), i.e., the percent

²² Importantly, many clusters meet more than one of the decision rules below.

- black is greater than 50 % OR the percent Hispanic is greater than 15 % or the percent of Native Americans is greater than 1 % (lack of representation, lack of resources).
- The percent of the population that is black is greater than 50 % and the percent of households that are headed by a female is greater than 20 % (lack of representation, high need).
 - The percent of the population in nursing homes is greater than 2 % (high need).
 - The percent of the population that is over 65 is greater than 20 % and the percent of the population receiving social security is greater than 33 % but median income is greater than \$25 K (high need).
 - The percent of the population living in poverty is greater than 20 % (lack of resources).
 - Median income less than \$20 K, less than 66 % labor force participation, and percent of the population without a HS degree is more than 25 % (lack of resources).

In the above decision rules, note that each highly or marginally vulnerable cluster represents populations facing one or more dimensions of social vulnerability that could make adapting and responding to climate change more difficult or which could exacerbate the impacts of sea level rise on health or financial outcomes: lack of resources, lack of political representation or access to government services, high levels of need, or vulnerable shelter. Each of these dimensions is represented by several different variables in our decision rules (and a larger set of variables in the overall dataset). For example, lack of resources is manifest not just in the income-oriented variables (percent in poverty, per capita income, percent unemployed) but also through limits on earning potential (percent without HS degree, percent female labor force participation). Lack of representation and access to government services is manifest in minority populations—for Virginia this means blacks, Hispanics, and Native Americans. High levels of need are measured by the percentage of the population over 65, in nursing homes and on social security. With respect to vulnerable shelter, we use both the percent of renters (as renters have little control over their shelter relative to homeowners) as well as the percent of mobile homes.

Table 7 presents the results of the clustering analysis for each of the geography/input set combinations. While the decisions rules for identifying clusters as vulnerable are based on the levels of poverty, unemployment, etc. in a tract, we note that determining whether a tract is vulnerable is not tied to the statistical properties of the input matrix. Thus, as shown in Table 7, the percentage of tracts that are identified as highly or marginally vulnerable changes for different geographies. For example, Coastal Virginia tends to have a lower percentage of tracts in the two vulnerable categories than Hampton Roads or Virginia overall, which reflects the fact that a high proportion of the Coastal Virginia tracts are in the wealthier and more developed Northern Virginia area.

To assess the performance of the clustering method, we examine the ability of this method to consistently identify vulnerable tracts regardless of the level of geography chosen or the number of variables used in the analysis, just as we did for the PCA-based indexing. As shown in Table 7, for the Hampton Roads geographies, there is a reasonable level of consistency across the different input sets in terms of identifying highly vulnerable tracts. Well over half of the tracts identified as highly vulnerable in any one cluster analysis are also considered highly vulnerable in the other analyses for the same geographies. The Virginia and Coastal Virginia analyses are not as consistent. We see the same patterns with respect to the group of highly or marginally vulnerable tracts. In the Hampton Roads analyses, 105 tracts are consistently identified as either highly or marginally vulnerable, which represents well over three quarters of the tracts identified by any one cluster analysis. However, for Virginia overall only 317 tracts (12 %) are consistently identified as

Table 7 Number and percent of tracts determined to be highly or marginally vulnerable using the cluster method

Geography/input set	Highly vulnerable	Highly or marginally vulnerable	Not vulnerable
VA41	238 (13 %)	512 (27 %)	1362 (73 %)
VA29	178 (9 %)	644 (34 %)	1230 (66 %)
VA13	213 (11 %)	537 (29 %)	1337 (71 %)
Overlap across all VA indices	89 (5 %)	317 (17 %)	1050 (56 %)
CV41	140 (12 %)	220 (19 %)	952 (81 %)
CV29	84 (7 %)	278 (24 %)	894 (76 %)
CV13	81 (7 %)	194 (17 %)	978 (83 %)
Overlap across all CV indices	52 (4 %)	136 (12 %)	855 (73 %)
HR41	41 (10 %)	119 (29 %)	289 (71 %)
HR29	24 (6 %)	124 (30 %)	284 (70 %)
HR13	33 (8 %)	133 (33 %)	275 (67 %)
Overlap across all HR indices	24 (6 %)	105 (26 %)	264 (65 %)

highly or marginally vulnerable which represents less than two-thirds of any one cluster analysis. Similarly, for Coastal Virginia 136 tracts are consistently identified as vulnerable, which is less than half of the tracts identified as vulnerable for the CV29 analysis. Adding together the percentage of tracts consistently identified as vulnerable (highly or marginally) and consistently identified as not vulnerable, for Virginia overall 73 % of tracts (17 % + 56 %) are classified consistently across all three variable sets.²³ For Coastal Virginia, 85 % (12 % + 73 %) are classified consistently and for Hampton Roads, 91 % (26 % + 65 %) are. These results support our expectation that the clustering method is more consistent for smaller and more homogeneous geographies. Looking across the three input variable sets, no clear patterns emerge in Table 7. Obviously the number of variables in the input matrix does affect the determination of which tracts are clustered together as well as the number of clusters, which ultimately affects the identification of tracts as either highly or marginally vulnerable. However, from Table 7, it is not clear that using a particular input variable set is preferred to using one of the others.

Table 8 also examines the total percentage of tracts consistently identified as vulnerable (highly or marginally) and consistently identified as not vulnerable, but does so using pairwise comparison across all of the geographies/input sets. For most comparisons, the percentage of tracts consistently characterized is in the 80 s or low 90 s. Also note that the percentage of tracts that are consistently characterized is slightly higher for the 13-variable and 29-variable pairs than it is for the analogous 41-variable pairs. However, the differences are not dramatic, as one might expect given that the clustering method is not subject to confounding variable effects in the same way that PCA-based indexing is. Although one concern with clustering is that it might not be able to identify sufficiently homogeneous clusters because the underlying observations are too heterogeneous to be clustered, given the consistency of the clustering analysis to different geographies and sets of variables, we feel confident that the clustering method does significantly help a researcher to identify potentially vulnerable populations. Even though clustering is more consistent under

²³ Thus a tract considered marginally vulnerable in one analysis and highly vulnerable in the other is considered to be correctly characterized while a tract considered marginally vulnerable in one analysis and not vulnerable in the other is considered to be mischaracterized.

Table 8 Overlap of cluster analysis vulnerability classifications: percent of tracts consistently categorized as highly/marginally vulnerable or not vulnerable

	VA41	VA29	VA13	EV41	EV29	EV13	HR41	HR29	HR13
VA41	1								
VA29	80 %	1							
VA13	80 %	86 %	1						
EV41	88 %	89 %	87 %	1					
EV29	87 %	92 %	91 %	91 %	1				
EV13	87 %	88 %	90 %	89 %	90 %	1			
HR41	79 %	89 %	89 %	84 %	89 %	89 %	1		
HR29	80 %	90 %	90 %	83 %	90 %	88 %	94 %	1	
HR13	78 %	88 %	90 %	80 %	88 %	87 %	92 %	95 %	1

changes in the input set than PCA-based indices, we continue to recommend that researchers carefully weigh the benefits of including additional variables. Finally we note that the clustering approach may be particularly useful in cross-jurisdictional analyses because it identifies groups that may be facing similar challenges and would facilitate sharing of ideas and resources across jurisdictions.

3.3 Comparison of composite indexing and clustering as identification strategies for social vulnerability

To compare the two methods to each other, we focus on the 401 census tracts that are in both the Hampton Roads and Coastal Virginia geographies and consider how consistently these tracts are characterized by the two methods across the various geographies and input sets. For the PCA-based indices, we consider only the variance-weighted indices as these were shown earlier to be more consistent than the equal-weighted indices. Table 9 shows the number of tracts that are identified as highly or marginally vulnerable by the two methods for each of the nine geography/input sets. Most importantly, it shows the number of tracts for each geography/input set that are consistently identified by the two methods. As was the case in Sect. 3.1, we used the greater than 1 standard deviation threshold for identifying highly vulnerable tracts and between 1 and 0.5 standard deviations as the thresholds for identifying marginally vulnerable tracts for the PCA-based index. In almost all cases, the 1 standard deviation threshold for highly vulnerable tracts results in more tracts identified as highly vulnerable than in the cluster analysis. Considering both highly and marginally vulnerable tracts, the PCA-based index identifies more tracts for Virginia for all three input sets and for two of the three Coastal Virginia input sets (CV29 and CV13). However, for Hampton Roads the PCA-based method identifies fewer tracts. We also provide data on the overlap of tracts identified as highly or marginally vulnerable by the two methods (columns 3 and 6 of Table 9). Overall, the overlaps are not particularly high, showing that the choice of method will affect which tracts are classified as vulnerable or not—or more specifically that the two methods are not interchangeable.

To better understand what may be happening across the two methods, we look more closely at the non-overlap tracts for the HR13 pair of analyses. Note that 40 tracts are identified as highly vulnerable by the PCA-based index and not by cluster analysis, while only 3 were identified as vulnerable by cluster analysis and not identified as vulnerable by the PCA-based index. Of the 3 tracts identified as vulnerable by cluster analysis and not by

Table 9 Comparison of PCA-based indexing to cluster analysis for the 401 Hampton Roads/Coastal Virginia Tracts

Geography/input set	Highly vulnerable tracts			Highly or marginally vulnerable tracts		
	PCA-based index	Cluster analysis	Overlap across the methods	PCA-based index	Cluster analysis	Overlap across the methods
VA41	97 (24 %)	83 (21 %)	59 (15 %)	187 (47 %)	130 (32 %)	78 (19 %)
VA29	87 (22 %)	37 (9 %)	30 (7 %)	174 (43 %)	115 (29 %)	102 (25 %)
VA13	73 (18 %)	37 (9 %)	35 (9 %)	118 (29 %)	123 (31 %)	98 (24 %)
Overlap across all VA indices	56 (14 %)	31 (8 %)	25 (6 %)	100 (25 %)	70 (17 %)	61 (15 %)
CV41	34 (8 %)	65 (16 %)	31 (8 %)	71 (18 %)	79 (20 %)	54 (13 %)
CV29	81 (20 %)	53 (13 %)	49 (12 %)	152 (38 %)	123 (31 %)	74 (18 %)
CV13	78 (19 %)	44 (11 %)	37 (9 %)	129 (32 %)	89 (22 %)	77 (19 %)
Overlap across all CV indices	29 (7 %)	26 (6 %)	24 (6 %)	52 (13 %)	60 (15 %)	41 (10 %)
HR41	53 (13 %)	39 (10 %)	13 (3 %)	96 (24 %)	116 (29 %)	52 (13 %)
HR29	59 (15 %)	22 (5 %)	16 (4 %)	107 (27 %)	121 (30 %)	79 (20 %)
HR13	68 (17 %)	31 (8 %)	28 (7 %)	103 (26 %)	130 (32 %)	95 (24 %)
Overlap across all HR indices	19 (5 %)	22 (5 %)	5 (1 %)	49 (12 %)	102 (25 %)	43 (11 %)

the PCA-based index, all are in clusters with very low median income, high percent living in poverty and either high unemployment or low labor force participation. However, all three of these tracts have factors that worked against them in the PCA-based index—that is they had very low levels of individuals over 65, very low nursing home populations, low percentage of individuals on social security, and low Hispanic and native populations, all of which decrease the vulnerability index without negating the vulnerability of the community.

Of the tracts 40 identified as highly vulnerable using the PCA-based index and not highly vulnerable by cluster analysis all were in clusters that were considered to be marginally vulnerable. One cluster has a high relative percentage of people over 65 and in nursing homes, but with above average income; the second has a high relative percentage of people over 65 and on social security, but only marginally low income; the third has a high percentage of people in nursing homes, but only moderate values for the other variables of interest; the fourth has a high percentage of people below the poverty rate, female heads of household, and blacks but only moderately low income and moderately high unemployment; and the fifth has a high percentage of female heads of household, blacks, and people on social security, but only moderately low income and moderately high unemployment.²⁴

²⁴ Similar to other studies, we do find that it may not be necessary to include race in a PCA-based index because the most important racial variable of interest, percent black, is strongly correlated with other included variables such as income (correlation -0.69), percent in poverty (0.67), percent with no high school degree (0.70), and percent female head of household (0.84). While this is useful for policymakers who may wish to or need to avoid explicitly specifying race as a criterion of analysis for legal or political reasons, to the extent that race may increase vulnerability on its own, we believe it should be included in measures of social vulnerability.

Looking at all tracts identified as vulnerable (highly or marginally), only 8 tracts are identified as vulnerable by the PCA-based index and not by cluster analysis, while 25 were identified as vulnerable by cluster analysis and not by the PCA-based index. Of the 8 identified as vulnerable by the PCA alone, all were marginally vulnerable and all were in clusters whose prototypes did not exceed threshold values for any of the variables. Of the 25 identified as vulnerable by cluster analysis alone, only one was in a highly vulnerable cluster. The other 24 were all in marginally vulnerable clusters where only one or two variables at a high level were necessary for the cluster to be categorized as vulnerable such as a high percentage of individuals in nursing homes or a high black population and a high percentage of female head of households. As discussed before, the PCA-based method does not do a good job of identifying tracts that are vulnerable in only one or two dimensions.

4 Discussion and recommendations for policymakers

This paper presents an alternative to PCA-based indexing using cluster analysis. Based on our case study, we believe that cluster analysis should be considered as a complementary method to PCA-based indexing, that is the cluster analysis approach may be more appropriate in some situations and less appropriate in others. Cluster analysis appears to be best suited for analyses at the state or local level for a number of reasons: the method is more time intensive than PCA-based indexing and thus works best if there are a manageable number of clusters. Since larger and more diverse sets observations will require additional clusters, using cluster analysis may be unwieldy. One of the strengths of cluster analysis is that it appears quite consistent under changes in the set of input variables, so researchers do not need to worry excessively as to whether adding an additional variable will dramatically affect the results of the analysis. Another one of its strengths is that it is quite transparent and the results are easy for lay audiences to interpret, so that it is well suited to local planning efforts that involve stakeholders or efforts where practitioners need to use the vulnerability determinations in very context-specific situations. Finally cluster analysis groups like observations which allows a researcher to easily identify areas that may face similar types of social vulnerability and could potentially share adaptation and mitigation solutions.

PCA-based indices are likely better suited than cluster analyses to multi-state or national analyses where there are many diverse observations. However, they are best used by individuals who understand how the indices are constructed and how to interpret them, including any subsequent rankings. One of the strengths of the PCA-based indices is that they provide a way of ranking communities in terms of vulnerability which cluster analyses do not. In particular, PCA-based indices are more appropriate than cluster analyses for academic research where the ability to provide a continuous measure of vulnerability is important. However, based on the results of our case study, we strongly recommend that researchers who are developing PCA-based indices think carefully about which variables to include in their analysis, limiting the use of bidirectional variables and assessing the potential for other confounding effects, and that researchers using the Kaiser selection criterion use variance weighting in constructing the final index.

A caveat in using both a PCA-based index and the cluster method, or any quantitative method, is that doing so risks creating a false appearance of objectivity by reducing social vulnerability to a quantitative measure. Clearly, some subjective interpretation on the part

of the researcher is still required—while this subjectivity is obvious for the clustering method, the subjectivity of the PCA-based index is less transparent. For both methods, the advantage of using readily available data from the US Census and other sources data to identify vulnerable communities is that it potentially provides a fair and consistent standard for evaluation and does not incur substantial cost in its implementation. However, quantitative analysis using data collected for purposes other than determining vulnerability to physical hazards cannot substitute for a more holistic and qualitative analysis of community need or for an inclusive stakeholder-oriented process. To be effective and equitable, adaptation and mitigation policies must include input from members of affected communities themselves as they experience environmental changes firsthand. For example, McNeeley (2014) provides a case study of drought in northwest Colorado using an analytical framework for the acquisition of ground-up, locally based knowledge useful in constructing and testing the validity of quantitative measures. This “toad’s eye” view of vulnerability emphasizes qualitative methods including interviews, participant observation, and analysis of critical documents informing stakeholder viewpoints, as well as traditional quantitative geophysical observation and analysis. Robust engagement with policy stakeholders builds trust between researchers and the community and allows the researcher to conduct an informed analysis of intersecting local social, cultural, economic, geophysical, and biophysical characteristics with strong internal validity.

Adaptation to climate change and sea level rise is a complex endeavor requiring sensitivity both to the needs of diverse communities and to the level of resources available to policymakers to respond to those needs. Assessing the social vulnerability of communities is an essential part of adaptation from the standpoint of both policy effectiveness and environmental justice. While quantitative approaches to identifying vulnerable communities have inherent limitations, they can provide an important guide for policymakers seeking to maximize the benefit of resources targeted at these efforts. With some modifications and caveats, both PCA-based indices and cluster analyses provide complementary, consistent, and transferable tools that policymakers can use to improve services in Hampton Roads, Virginia, in other communities in the USA, and around the world.

Acknowledgments Funding was provided by Blue Moon Fund (Grant no. 775602A-712685).

References

- Abdi H, Williams LJ (2010) Principal component analysis. *Wiley Interdiscip Rev Comput Stat* 2:433–459
- Armaç I, Gavriş A (2013) Social vulnerability assessment using spatial multi-criteria analysis (SEVI model) and the Social Vulnerability Index (SoVI model)—a case study for Bucharest, Romania. *Nat Hazards Earth Syst Sci* 13:1481–1499
- Baum S, Horton S, Choy DL (2008) Local urban communities and extreme weather events: mapping social vulnerability to flood. *Australas J Reg Stud* 14:251–273
- Bjarnadottir S, Li Y, Stewart MG (2011) Social vulnerability index for coastal communities at risk to hurricane hazard and a changing climate. *Nat Hazards* 59:1055–1075
- Blaikie PM, Cannon T, Davis I, Wisner B (1994) *At risk: natural hazards, people’s vulnerability and disasters*. Routledge, London
- Boruff BJ, Emrich C, Cutter SL (2005) Erosion hazard vulnerability of US coastal counties. *J Coast Res* 21:932–942
- Burton CG (2010) Social vulnerability and hurricane impact modeling. *Nat Hazards Rev* 11:58–68
- Burton C, Cutter SL (2008) Levee failures and social vulnerability in the Sacramento-San Joaquin delta area, California. *Nat Hazards Rev* 9:136–149
- California Emergency Management Agency (2013) State hazard mitigation plan. http://hazardmitigation.ca.gov/plan/state_multi-hazard_mitigation_plan_shmp

- Chakraborty J, Tobin GA, Montz BE (2005) Population evacuation: assessing spatial vulnerability in geophysical risk and social vulnerability to natural hazards. *Nat Hazards Rev* 6:23–33
- Chang SE, Yip JZK, de Jong SLVZ, Chaster R, Lowcock A (2015) Using vulnerability indicators to develop resilience networks: a similarity approach. *Nat Hazards* 78:1827–1841
- Clark G, Moser S, Ratick K, Dow W, Meyer S, Emani W, Jin J, Kasperson R, Schwartz H (1998) Assessing the vulnerability of coastal communities to extreme storms: the case of Revere, MA, USA. *Mitig Adapt Strat Glob Change* 3:59–82
- Climate Central (2016) Suring seas: sea level rise analysis by climate central. <http://sealevel.climatecentral.org/>. Accessed 2016
- Cutter SL (1996) Vulnerability to environmental hazards. *Prog Hum Geogr* 20:529–539
- Cutter SL, Finch C (2008) Temporal and spatial changes in social vulnerability to natural hazards. *Proc Natl Acad Sci* 105:2301–2306
- Cutter SL, Boruff BJ, Shirley WL (2003) Social vulnerability to environmental hazards. *Soc Sci Q* 84:242–261
- Cutter SL, Emrich CT, Morath DP, Dunning CM (2013) Integrating social vulnerability into federal flood risk management planning. *J Flood Risk Manag* 6:332–344
- de Oliveira Mendes JM (2009) Social vulnerability indexes as planning tools: beyond the preparedness paradigm. *J Risk Res* 12:43–58
- Fekete A (2012) Spatial disaster vulnerability and risk assessments: challenges in their quality and acceptance. *Nat Hazards* 61:1161–1178
- Finch C, Emrich C, Cutter SL (2010) Disaster disparities and differential recovery in New Orleans. *Popul Environ* 3:179–202
- Fothergill A, Peek LA (2004) Poverty and disasters in the United States: a review of recent sociological findings. *Nat Hazards* 32:89–110
- Fothergill A, Maestas E, Darlington JD (1999) Race, ethnicity and disasters in the United States: a review of the literature. *Disasters* 23:156–173
- Holand IS, Lujala P (2013) Replicating and adapting an index of social vulnerability to a new context: a comparison study for Norway. *Prof Geogr* 65:312–328
- Makles A (2012) How to get the optimal k -means cluster solution. *Stata J* 12:347–351
- Martinich J, Neumann J, Ludwig L, Jantarasami L (2013) Risks of sea level rise to disadvantaged communities in the United States. *Mitig Adapt Strat Glob Change* 18:169–185
- McNeeley SM (2014) A “toad’s eye” view of drought: regional socio-natural vulnerability and responses in 2002 in Northwest Colorado. *Reg Environ Change* 14:1451–1461
- Mirkin Boris (2013) *Clustering: a data recovery approach*, 2nd edn. CRC Press, Boca Raton
- Myers CA, Slack T, Singelmann J (2008) Social vulnerability and migration in the wake of disaster: the case of Hurricanes Katrina and Rita. *Popul Environ* 29:271–291
- National Oceanic and Atmospheric Administration (2016) Sea level rise and coastal flooding impacts. <http://coast.noaa.gov/slr/>. Accessed 2016
- O’Keefe P, Westgate K, Wisner B (1976) Taking the naturalness out of natural disasters. *Nature* 260:566–567
- Pacific Institute (2016) Mapping social vulnerability to climate change in California. http://www2.pacinst.org/reports/climate_vulnerability_ca/maps/about.html. Accessed 2016
- Rygel L, O’Sullivan D, Yarnal B (2006) A method for constructing a social vulnerability index: an application to hurricane storm surges in a developed country. *Mitig Adapt Strat Glob Change* 11:741–764
- Schmidtlein MC, Shafer JM, Berry M, Cutter SL (2011) Modeled earthquake losses and social vulnerability in Charleston, South Carolina. *Appl Geogr* 31:269–281
- Sherrieb K, Norris FH, Galea S (2010) Measuring capacities for community resilience. *Soc Indic Res* 99:227–247
- Solangaarachchi D, Griffin AL, Doherty MD (2012) Social vulnerability in the context of bushfire risk at the urban-bush interface in Sydney: a case study of the Blue Mountains and Ku-ring-gai local council areas. *Nat Hazards* 64:1873–1898
- Tate E (2012) Social vulnerability indices: a comparative assessment using uncertainty and sensitivity analysis. *Nat Hazards* 63:325–347
- Tate E (2013) Uncertainty analysis for a social vulnerability index. *Ann As Am Geogr* 103:526–543
- Tate E, Cutter SL, Berry M (2010) Integrated multihazard mapping. *Environ Plan B Plann Des* 37:646–663
- Toké NA, Boone CG, Arrowsmith JR (2014) Fault zone regulation, seismic hazard, and social vulnerability in Los Angeles, California: hazard or urban amenity? *Earth’s Future* 2:440–457

- Turner BL, Kasperson RE, Matson PA, McCarthy JJ, Corell RW, Christensen L, Eckley N, Kasperson JX, Luers A, Martello ML, Polsky C, Pulsipher A, Schiller A (2003) A framework for vulnerability analysis in sustainability science. *Proc Natl Acad Sci* 100:8074–8079
- Virginia Institute of Marine Sciences (2013) Report to the Governor, Senate Document No. 3: Recurrent Flooding Study for Tidewater Virginia
- Wang C, Yarnal B (2012) The vulnerability of the elderly to hurricane hazards in Sarasota, Florida. *Nat Hazards* 63:349–373
- Wolters ML, Kuenzer C (2015) Vulnerability assessments of coastal river deltas—categorization and review. *J Coast Conserv* 19:345–368
- Wood NJ, Burton CG, Cutter SL (2010) Community variations in social vulnerability to Cascadia-related tsunamis in the U.S. Pacific Northwest. *Nat Hazards* 52:369–389