Extreme Heat Vulnerability of the Population in Georgia, USA

BY

BAKARI BAKER
B.S., Clark Atlanta University, 1999
MPH, University of Illinois at Chicago, Chicago, 2010

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Public Health Sciences
in the Graduate College of the
University of Illinois at Chicago, 2019

Chicago, Illinois

Defense Committee:

Michael Cailas, Chair and Advisor
Sage Kim, Public Health Risk Analysis
Apostolis Sambanis, Environmental Public Health Preparedness
John Canar, Public Health Geographic Information Systems
Kristin Osiecki, University of Minnesota Rochester
Roland Thorpe Jr., John Hopkins University

This thesis is dedicated to my fiancé/wife, Kharel; my mother and father, Larry and Sarah Baker, for all the support and motivation to complete my PhD. This is also dedicated to my daughters Nailah and Safiya who sacrificed time with their father, so I could concentrate on the completion of my dissertation.

ACKNOWLEGDMENTS

I would like to thank my thesis committee (Dr. Michael Cailas, Dr. Sage Kim, Dr. Roland Thorpe, Dr. Apostolis Sambinas, Dr. Kristen Osiecki, Dr. John Canar) for the support and assistance. They provided guidance in all areas that helped me accomplish my research goals. I would also like to acknowledge Arash Jalali, who dedicated his time to assist me with all of the contributions he made toward helping me complete the mapping for this thesis.

I would also like to that the Georgia Department of Public Health for assisting me with my data collection.

BB

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF TABLES (Continued)

TABLE PAGE

LIST OF TABLES (Continued)

LIST OF FIGURES

LIST OF FIGURES (CONTINUED)

FIGURES PAGE

LIST OF FIGURES (CONTINUED)

FIGURES PAGE

LIST OF ABBREVIATIONS

A.A.Morb	Age Adjusted Morbidity
AC	Air Condition / Air-Conditioning
ADMHI	Average Daily Max Heat Index
AT	Apparent Temperature
ATSDR	Agency for Toxic Substances and Disease Registry
CAT I	Category One
CAT II	Category Two
CAT III	Category Three
CCES	Center for Climate and Energy Solutions
CDC	Center for Disease Control and Prevention
CEVI	Combined Exposure Vulnerability Index
CF	Classification Failure
CI	Crime Index
CM	Coincidence Matrix
DT	Decision Tree
DLLPI	Disaster Loss Period 1
EH	Extreme Heat
EHE	Extreme Heat Event
EHEE	Extreme Heat Exposure Exceedance
EHEV	Extreme Heat Event Vulnerability
EHRI	Extreme Heat Risk Index
EHVI	Extreme Heat Vulnerability Index

LIST OF ABBREVIATIONS (CONTINUED)

ER	Emergency Room
FC	Failure Class
FIPS	Federal Information Processing Standards
GA	Georgia (U.S. State)
GDPH	Georgia Department of Public Health
HVI	Heat Vulnerability Index
ICD	International Classification of Diseases
IPCC	Intergovernmental Panel on Climate Change
LC	Land Cover
LULC	Low Urban Land Cover
LSR	Land Surface Resilience
LST	Land Surface Temperature
LSV	Land Surface Vulnerability
MBR	Morbidity Rate
MDHSS	Missouri Department of Health and Senior Services
MLC	Material Land Cover
MODIS	Moderate Resolution Imaging Spectroradiometer
NLADS	North America Land Assimilation System
NLCD	National Land Cover Database
NOAA	National Oceanic and Atmospheric Administration
NWS	National Weather Service
NVDI	Normalized Difference Vegetation Index

LIST OF ABBREVIATIONS (CONTINUED)

OCP	Overall Classification Performance
OECD	Organisation for Economic Co-operation and Development
OOR	Overall Overestimation Rate
OUR	Overall Underestimation Rate
PAM	Performance Assessment Matrix
PAO	Performance Assessment Optimization
P.A.Morb	Population Adjusted Morbidity
PCA	Principle Component Analysis
PFR	Percentage Fractional Ranks
PHV	Predictive Heat Vulnerability Model
PPM	Predictive Performance Matrix
PR	Percentile Rank
P.T.ULC	Urban Land Cover
PVI	Prevalent Vulnerability Index
SES	Socio-economic Status
SoV	Social Vulnerability
SoVP	Social Vulnerability Potential
SVI	Social Vulnerability Index
Tmax	Maximum Temperature
UE	Underestimation Error
UHI	Urban Heat Island
UIC-SPH	University of Illinois at Chicago School of Public Health

LIST OF ABBREVIATIONS (CONTINUED)

ULC	Urban Land Cover
UNISDR	United Nations International Strategy for Disaster Reduction
U.S. EPA	United States Environmental Protection Agency
VARI	Input Variables

SUMMARY

"Heat waves can be deadly. Extreme heat was the leading cause of weather-related deaths in the United States from 2000 through 2009." (CDC 2016). The latest data from the National Weather Service Office of Climate, Water, and Weather service reveals that Heat was the leading cause of weather fatalities from 2006 – 2015 and from 1986 – 2015 (NOAA 2015). The U.S. Natural Hazard Statistics provide statistical information on fatalities, injuries and damages caused by weather related hazards. (Figure 1, Appendix A) These statistics are compiled by the Office of Services and the National Climatic Data Center from information contained in *Storm Data*, a report comprising data from NWS forecast offices in the 50 states, Puerto Rico, Guam and the Virgin Islands (NOAA 2015).

We argue and demonstrate that there is a need to assess the performance of indices used to identify vulnerable segments of the population to the adverse effects of extreme heat. The current research establishes a methodology to validate the performance of these indices. In addition, the various variables describing land cover are introduced and with the proposed assessment methodology it becomes feasible to assess their contribution in developing reliable indices. Our research and analysis propose to determine the extent to which extreme heat has affected and will affect the population in Georgia, USA.

The newly created extreme heat variable will be analyzed along with newly developed land cover variables, the commonly accepted socioeconomic variables and the heat related morbidity rates in GA from 2000 to 2014. These variables will be analyzed via newly developed approaches to determine whether the addition of the land cover variables is significant and provide more accurate identification of vulnerable populations in Georgia, USA.

1. INTRODUCTION AND PROBLEM STATEMENT

Extreme heat (EH) causes the most disaster related deaths annually when compared to other extreme weather events (i.e. floods and earthquakes). As with other natural disasters, research shows that the impact of the heat waves affects some people disproportionally, with some geographical areas reporting more heat-related mortality (Ghiasi 2011). Moreover, vulnerability to EH is different than other natural disasters because there are unique vulnerability factors specific to EH such as age; medical conditions; living conditions; social isolation and type of employment. (Cooley 2010). Different factors such as health and environmental factors, have been linked to increased susceptibility to adverse effects from excessive heat. To help identify vulnerable population to natural disasters, researchers have used a social vulnerability index (SVI) (Cutter et.al. 2003). Vulnerability modeling intends to assist public health officials by identifying areas where there are spatial clusters of vulnerable people and justify the involvement of stakeholders in preventive and mitigation plans. Georgia was chosen in part because, Georgia is of one the most EH vulnerable states due to its large population and the number of heat wave days per year (SA 2017). In addition, the state has not conducting a detailed vulnerability assessment and not implementing adaptation strategies to improve extreme heat resilience (SA 2017).

Resilience is a community's ability to return to normal after a natural disaster and is also information vital to a public health official's ability to identify population areas that are most susceptible to the negative effects of Extreme Heat Events (EHEs). According to Bene 2013, there is a need to better understand resilience at all societal levels (individual, households, communities and societies). One necessary step in this process is to better measure resilience. Without the ability to measure and/or to monitor resilience, policy makers and societies more broadly will not be in position to identify and support interventions that have more effect on people's ability to respond and to accommodate adverse events. Putting this resilience measurement into practice is therefore a priority, but it is not an easy task and many challenges lie ahead (Bene 2013). To the best of our knowledge, determining the interrelation of resilience and vulnerability has not been readily explored.

The state of Georgia has had several EHEs in the past. The hottest summer on record in Atlanta, Georgia was in 1980, with an average daily temperature of 82.6^{o}F. Sixteen daily high temperature records were broken in 1980 (GATE 2016). The second hottest summer on record in Atlanta, Georgia was 2016, with an average daily temperature of 82.4^{o}F. There were 71 days with maximum temperature days above 90^{o}F (NWS 2016). The rate of increase of more frequent and longer heat waves has been higher than the national average in Atlanta, Ga from 1962 to 2010 (Winquit et. al. 2015).

An evaluation of the usefulness of the land cover variables will be completed. To the best of our knowledge, our newly created land cover variables are used for the first time as input variables for determining the accuracy of extreme heat vulnerability. There is a lack of methodology to validate the readily used social vulnerability indices. (Sambanis 2016). We will complete our analysis with an age-adjusted representative population of Georgia within a readily identified segment of the widely accepted vulnerable population to determine which counties in Georgia are the most vulnerable to EH.

2. BACKGROUND

A. Effects of Excessive Heat on the Human Body

Heat waves are one of the natural disasters that cause major health impacts without prominent destruction and property loss similar to other disasters. The human body has the ability to adjust to different situations, including increased heat, but during an EHE, the body's thermoregulatory mechanism can get overwhelmed resulting in a quick rise in the body's temperature reaching dangerous levels (Pooler 2009). Heat illness can result from exposure to a hot, humid environment for a prolonged period of time (non-exertional) and often happens to the elderly and people with chronic illnesses, or from intense strenuous physical activity in hot weather (exertional), such as with school athletes and outdoor workers. (Mayo Clinic 2014) Certain psychiatric medications such as Lithium, Haldol and Thorazine can cause dehydration, lower blood pressure and cause sun sensitivity (GDHS 2007).

> *Heat directly influences the body by interfering with the cellular processes. Blood flow is redistributed to the periphery. With the loss of fluids and electrolytes in sweat and the redistribution of blood volume to the periphery, a tremendous burden is placed on the heart, which ultimately may fail to maintain an adequate cardiac output, resulting in cardiovascular collapse, multiorgan failure, and ultimately death. (Helman,2014)*

Levels of Heat Illness: Heat stress and heat cramps happen after exposure to excessive heat during a heat wave in the summertime or after exercising in a warm environment. Initially the person will feel discomfort and physiologic strain and will have decreased exercise performance. When the body's core temperature rises, heat stress will turn into heat exhaustion or stroke (Jardine 2007). Heat cramps are muscle pains and spasms due to heavy exertion and can be a first signal that the body is having trouble with the heat (EPA 2016).

Heat exhaustion is characterized by an elevated core body temperature between 100.4°F and 104°F, and mild dehydration (Jardine 2007). Symptoms include profuse sweating, intense discomfort, confusion, thirst, nausea, and vomiting and the person suffering from heat exhaustion will feel faint and might collapse. Once cooled and rehydrated, patients with heat exhaustion usually make a full recovery (Jardine 2007) (Becker et. al. 2014). The main difference between heat exhaustion and heat stroke is the absence of neurological symptoms in heat exhaustion cases (Jardine 2007).

On the other hand, heat stroke, the most severe heat-related illness, occurs with severe dehydration, and is identified when a body's core temperature is greater than 104°F with exposure to heat, and neurological symptoms including delirium and confusion that can progress rapidly to coma, and seizure with poor outcome (Jardine 2007) (Becker et. al. 2011). Classic heat stroke develops slowly over days usually in older persons and the chronically ill (Bauchana et.al. 2002). Exertional heat stroke however happens more rapidly and usually occurs in young healthy persons (Howe et. al. 2007).

Heat stroke can cause hypotension, irreversible myocardial impairment, liver abnormalities, and renal failure (Jardine 2007). The outcome of heat stroke varies; mild heat stroke patients usually recover fully but moderate-to-severe heat stroke patients might sustain some sequels (Jardine 2007). From the heat illness classification, only heat stroke is considered a medical emergency (Bauchana et.al. 2002). Recognition of the early signs of heat illness is important to prevent heat stroke and its potential consequences. Some health conditions are worsened by heat stress, including neuropsychiatric disorders (Kim et.al. 2007). Hemorrhagic shock and encephalopathy syndrome is a serious consequence of heat strokes that affects infants and can cause a severe neurological outcome (Jardine 2007).

Hyperthermia mortality risk increases from sustained periods of high temperature (heat waves) rather than from individual days (Hajat et. al. 2006). When a heat wave occurs at an unexpected time, for example, early in the summer months, the impact is usually greater due to lack of preparedness. The longer the duration of the heat wave, the more the detrimental impact on the population is likely due to possible power outages leaving health/emergency systems overwhelmed (Atkins 2013). EHEs are silent killers compared to other weather hazards because of the lack of physical property destruction associated with other disasters such as hurricanes and tornadoes. Although extreme heat is the leading killer among all natural disasters, there is usually no permanent reminder for people to be proactive about possible future events (Changnon, S. A et. al. 1996) (NWS 2016).

One example of the effects of an EHE was the July 1995 heat wave in Chicago, Illinois, where heat caused more than 700 deaths (Klinenberg 2001). Heat persisted for a week and the temperature at night was in the low-to-mid 80s and did not allow for any relief. The city suffered power outages, transportation slowed, and thousands of people developed heat-related illnesses that caused the hospitals to be overwhelmed. The city had to provide refrigerated trucks to store bodies after the morgue at the Cook County Medical Examiner's Office reached its capacity (UCP Press. 2002). Other more recent events that caused high heat-related mortality rates happened in 2003 in Europe affecting France, England, the Netherlands, Portugal, and Spain, among others; (Robine et. al. 2007) (Kovats et. al. 2006) (Garssen et. al. 2005) (Simón 2005) in Melbourne in 2009; (Atkins 2013) and Moscow in 2010 (Wolf et. al. 2013) (Revich et. al. 2012) (Shaposhnikov 2014). The summer of 2015 was one of the deadliest summers in the recent years where more than 1,100 people died in Pakistan when the heat reached 112.6°F (deadliest on record for the country), and more than 2,200 people died in India due to heat waves (Rice 2015).

Some, like certain officials who wanted to downgrade the heat event that happened in Chicago in 1995, might argue that people who die during heat waves are very sick and might probably have died without the added stress of extreme heat (UCP Press. 2002). Studies that reviewed the literature of health outcomes during EHEs after the 2003 heat waves found there was excess mortality that was not explained by the forward shift (harvesting) effect with no significant decrease in mortality in the following weeks after the heat wave (Robine et. al. 2003) (Martiello et. al. 2010).

In the United States, in 2005, there were about 6,200 hospitalization cases related to heat with an average stay of 3.2 days and an average cost-of-stay of $6,200 per case (Merrill et. al. 2008). Wu et al. 2014 examined data from the 2009 and 2010 Nationwide Emergency Department Sample and concluded that approximately 4,100 emergency department visits annually can be the result of heat stroke, with the majority happening in the summer time. Heat illness is preventable and with proper planning the magnitude of mortality, morbidity, and monetary loss can be dramatically reduced.

Chen et. al (2016) recently completed a time-series analysis of heat-waves and emergency department visits in Atlanta, GA. They assessed the increase in the risk of emergency department visits during heat wave days compared to non-heat wave days for the months of May through September from 1993 to 2012. They complied the ICD-9 data for emergency room visits for the following health outcomes from 1993 to 2012: Fluid and electrolyte imbalance; Acute renal failure; Hypertension; Ischemic heart disease; Dysrhythmia; Congestive heart failure; Ischemic stroke; Pneumonia; Chronic obstructive pulmonary disease; Asthma/wheeze; Diabetes mellitus; and Intestinal infection. They found the strongest evidence of significant heat related morbidity with acute renal failure, ischemic strokes, and intestinal infections. They state that in

Atlanta, GA "prolonged heat exposure can confer added adverse health impacts beyond the risk due to higher daily temperature, particularly for renal diseases, cardiovascular diseases and intestinal infection." (Chen et. al. 2017).

B. <u>Climate Change</u>

There is no doubt that the earth's temperature is rising. Climate change experts agree that the earth's temperature has risen 1^{o}C in the last 100 years. One of the effects of this temperature increase is more extreme weather including heat waves (EPA CC 2016). There has been an increase in the number of hot days in the U.S since 1948. (Figure 2, Appendix A) According to the U.S. EPA, "Climate models project that if global emissions of greenhouse gases continue to grow, summertime temperatures in the United States that ranked among the hottest 5% in 1950-1979 will occur at least 70% of the time by 2035-2064." (Melillo et. al. 2014). The increase in extreme heat events is a result of climate change.

C. <u>Urban Heat Island</u>

The Atlanta metropolitan area is at more risk to extreme heat conditions due to the heat island effect. According to the U.S. EPA, "many urban and suburban areas experience elevated temperatures compared to their outlying rural surroundings; this difference in temperature is what constitutes an urban heat island. The annual mean air temperature of a city with one million or more people can be 1.8 to 5.4°F (1 to 3°C) warmer than its surroundings." (EPA 2008).

Elevated temperatures from heat islands increases the rate of compromised health, especially in sensitive and more vulnerable populations, such as children, older adults, and those with existing health conditions (EPA 2008). According to Climate Central, the Atlanta metropolitan area is an urban heat island and will have a significant increase in the number of days with a heat index above 105^{o}F in the near future (CCUS 2016). Rural areas usually have lower surface temperatures due to the trees, vegetation and open land that dominate its landscape. Urban areas usually have more impervious and paved surfaces as well as significantly less green cover (EPA 2008). The lack of green cover in the urban Atlanta metropolitan area will be an additional vulnerability measure in this thesis.

D. Extreme Heat Event Vulnerability

At an empirical level the relevant literature corroborates our major supposition that the Extreme Heat Event Vulnerability (EHEV) indices have a categorical nature. Most of the research in this field aims to generate the spatial distribution of regions classified in terms of vulnerability potential (Cutter and Finch, 2008); e.g., 5 levels with the "most vulnerable" counties being those that have a standard deviation score above 1 (Cutter et al., 2003). Most published studies used Geographic Information Systems (GIS) packages, such as ESRI ArcGIS, that have a standard feature known as class ranges and breaks (Bunting et al., 2014). These features define the amount of data that falls into each class and the appearance of the map. There are two main components in a GIS classification scheme: the number of classes into which the data is to be organized and the method by which classes are constituted. The number of classes is dependent on the objective of the analysis. The main rules by which the data are assigned to a class, are: Manual, Equal Interval, Quantile, Natural Breaks (Jenks), standard deviations, and geometric intervals (De Smith et al., 2007). A common theme of all these rules is the establishment of classes with distinct characteristics in terms of vulnerability (e.g., high, medium, low).

Binta et. al. published an article titled *Climate change vulnerability assessment in Georgia* (Binta et. al 2015). Their research resulted in well substantiated information regarding climate change and vulnerable populations in Georgia. They built on existing assessment techniques to develop a climate change vulnerability assessment which combined climate, social, land cover, and hydroclimatic events in the state of Georgia (IPCC 2007).

According to Binta et. al. (2015), “vulnerability to climate change is the degree to which a system is adversely affected by climate related stimuli and its inability to cope with them”. (IPCC 2007) It is typically characterized as some function of exposure, sensitivity, and adaptive capacity. The Intergovernmental Panel on Climate Change (IPCC) states that climatic variations measure exposure of the system; sensitivity is the effect of variations on the system; and adaptive capacity is the ability of a system to adjust to climate related stimuli. The physical causes, that is, exposure and their effects are explicitly defined, and the social context is captured in terms of sensitivity and adaptive capacity.” (Binta et. al 2015). According to Binta, Shepherd and Johnson, their study “quantifies vulnerability to climate change through a holistic approach by integrating biophysical and social vulnerability with geographic vulnerability” (Binta et. al 2015).

Binta et. al. (2015) performed a Principal Component analysis (PCA) of their variables similar to the Social Vulnerability Index (SVI) method specified by Cutter et al. (2003). Binta et. al. (2015) combined climatic, social, land cover, and hydrological components together into a unified vulnerability assessment. Climate change vulnerability was measured as a departure of decadal mean temperature and precipitation from baseline temperature and precipitation and extreme hydroclimatic hazards indicated by flood, heat wave and drought events. Sensitivity and adaptive capacity were also measured by conceptualizations and methods derived from socioeconomic variables. The location of vulnerable populations also accounted for impervious surface and flood susceptibility areas (Binta et. al. 2015).

According to Binta et. al. (2015), “the variables were standardized into percentage values. Principle component analysis (PCA) was performed with Varimax rotation to identify the variables that provide maximum loading for each of the principal components. The dominant variables in PCA determine the directionality of each principal component. Each principal component score was weighted by its percentage variance such that the components with higher variance contribute more towards overall sensitivity. Each of the weighted principal component scores was summed to construct the overall social vulnerability score. High social vulnerability score indicates high sensitivity and low adaptive capacity and vice versa. The social vulnerability scores are rescaled to 0-4 scale.” (Jolliffe 2014).

According to Sambanis (2016): “Principal Component Analysis (PCA) is a multivariate statistical technique that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components (Jolliffe, 2014) (Fekete 2012). The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to (i.e., uncorrelated with) the preceding components. The principal components are orthogonal because they are the eigenvectors of the covariance (or correlation) matrix, which is symmetric (Fukunaga, 2013). In addition, expert judgment, is currently a critical element in the subjective interpretation of the components generated by these prevailing social vulnerability index methodologies (Fekete, 2012). (Sambanis 2016) The goal of using principal component analysis is to aggregate the original Census variables (e.g., nxp matrix) into a few groups, the principle components (i.e., the characteristic p vectors).” (Logan 2012). In the social vulnerability research field, the scores of the principle components are used to derive the composite index; these scores are the coordinates of the input individually observations expressed in the new orthogonal system of axes defined by the observations of possibly correlated variables characteristic vectors or eigenvectors. If all the PCs are retained then the scores, represent the original input observations in the new uncorrelated dimensionality. As we pointed above, deriving a composite index by aggregating the scores of the selected PCs is likely to pose questions since each uncorrelated component encapsulates unique characteristics of the original space. Summing up these unique characteristics implies that each component is an alternative generalized expression of the same ultimate entity, which in this case could be the vulnerability potential. This supposition is correct only if all the selected correlated variable PCs are contributing to vulnerability in an “equal” manner (Cutter et al., 2003). The relationship of the selected PCs to vulnerability is based on the subjective interpretation and labelling of the components (Fekete, 2012). This subjectivity is evident as well from the introduction of various multipliers (+/- 1) and the use of absolute score values when the sign was ambiguous in terms of contribution to vulnerability (Cutter et al., 2003) The composite social vulnerability indices derived from PCs scores are likely to have better distributional properties since they are the results of linear combinations of the original variables. In addition, many statistical packages (e.g., SPSS or SAS) yield standardized scores having zero mean and a unit variance; standardization of the original input dataset is not required if the correlation matrix is used for PC derivation.

3. MAJOR OBJECTIVES AND RESEARCH FOCI

The overall objective of this thesis is to determine a reliable methodology for identifying the geographic areas that are at a higher risk to extreme heat events within the state of Georgia. To accomplish this overall objective a performance assessment (PA) approach and a benchmark of performance was developed and applied. Owing to this PA approach, the introduction of land cover as an exploratory variable for deriving better EHEV indices is assessed. The proposed PA provides the tools to develop reliable indices which can be used by public health professionals to prepare and implement effective EH mitigation plans. This dissertation is organized into three major research foci, which aim to answer the following questions:

Focus-1: Is extreme heat adversely affecting the people in the state of Georgia?

Focus-2: What is the reliability of the existing Extreme Heat Event Vulnerability (EHEV) indices derivation approaches which are readily used in EHE vulnerability literature.

Focus-3: Is there a more accurate approach for identifying vulnerable populations in GA that can lead to lower morbidity and mortality rates during future extreme heat conditions?

To address these foci, the following research approach will be applied:

Focus-1: Georgia has been undergoing rapid growth and landscape change during the 2000-2014 research timeframe. According to the U.S. Census, the population of Georgia increased from approximately 8.1 million (8,186,453) in 2000 (US Census 2000) to nearly 10 million (9,687,653) in 2010 (US Census 2010). The largest metropolitan area in Georgia is the Atlanta-metropolitan area. The 2016-2017 Census revealed that the Atlanta-metro area had the third largest population increase in the nation (US Census 2017). I will include changes in land cover to determine the role it has in the health effects of extreme heat to the population of Georgia.

Focus-2: The focus of the research analysis is to determine if there is a more accurate approach for identifying vulnerable populations in GA that could be applied to lower morbidity and mortality rates during future extreme heat conditions. I will assess the commonly used methodologies for deriving vulnerability and identify a framework that can be used by public health officials to predict the geographic areas in Georgia are more vulnerable to extreme heat conditions. This will be done by applying and comparing the principle component, ranking and decision tree methods to demographic variables, historical heat, morbidity/mortality and changes in land cover. This first half of the data will be compared to the second half of the data to identify and validate a better approach to analyzing for vulnerability.

Focus-3: We built upon the methodology developed in the *Climate change vulnerability assessment in Georgia* (Binta et. al 2015) article and introduce alternate modeling and assessment techniques which will result in a more reliable approach to identify portions of the population that are more vulnerable to extreme heat in Georgia. This premise is essential to establish a validation methodology, which can be applied to all social vulnerability indices regardless of the derivation approach and assess the relevance of the index in terms of manifested events. I will review the existing methods that are used to identify vulnerable populations and compare them to my newly developed method by completing a vulnerability analysis for the state of Georgia, USA. I will also be performing a retrospective analysis of the heat related hospital visits and deaths in the state of Georgia from 2000 – 2014 to determine the population who experience the highest level of vulnerability to determine the population with the highest future vulnerability to extreme heat. I will analyze the morbidity and mortality rates by county and compare the results to the corresponding daily air temperature, humidity and heat indexes for the same timeframe.

The climate information was obtained from the Center for Disease Control and Prevention’s North America Land Assimilation System (NLDAS) (NLADS 2017) and the National Weather Service (NWS 2017). The morbidity and mortality data were obtained from the Georgia Department of Public Health. This information can also be used to identify potential areas for intervention and further research.

4. LITERATURE REVIEW

A. Evaluation of Georgia, USA

Most of the related literature applied a variation of the methodology developed by Reid (2009). Reid mapped 10 vulnerability factors for heat -related morbidity/mortality in the US in geographic space and identify potential areas for intervention and further research. The ten factors that Reid uses are:

- Percent of the population below the poverty line
- Percent of the population with less than a high school diploma
- Percent of the population of a race other than white
- Percent of the population living alone
- Percent of the population greater than 65 years of age
- Percent of the population greater than 65 years of age living alone
- Percent of the census tract area not covered in vegetation
- Percent of the population ever diagnosed with diabetes
- Percent households without central AC
- Percent households without any AC

The methodologies used to determine extreme heat social vulnerability id highlighted further in Table XXV, Appendix B.

The variation in Reid's methodology had to be modified at the local level in Georgia because the state has no info on the percentage of households with air conditioning. Maier (2003) used a Reid modified heat vulnerability index to perform their research. This article could not account for AC in the state of Georgia, so they substituted this variable with the percent greenspace. Maier and Reid also did not use greenspace as a primary indicator of vulnerability.

Socio-economic status (SES) and demographic variable were used as primary drivers to determine social vulnerability. Binta (2015) also used a Reid modified heat vulnerability index without AC info. The vulnerability factors used in the article included socioeconomic and social vulnerability.

The state of Georgia was chosen because it is located in the more temperate southern region of the United States; the urban and suburban population in GA is increasing rapidly; the number of extreme heat days is increasing annually; and there is a large population that is vulnerable to extreme heat. Georgia's population has increased from nearly 8.3 million in 2000 to nearly 10.4 million in 2017 (US Census-GA 2017). According to the State of Georgia, Georgia's population is expected to grow to approximately 14.7 million by 2030 (SOG 2017). "Georgia currently averages about 20 dangerous heat days a year. By 2050, it is projected to see more than 90 such days per year." (SA 2017). According to the state report card, Georgia is of one the most vulnerable states due to its large vulnerable population and the number of heat wave days per year (SA 2017). The state report card gave the state of Georgia a "C-" grade due in part to the state not conducting a detailed vulnerability assessment and not implementing adaptation strategies to improve extreme heat resilience (SA 2017).

Georgia contains several rapidly growing major metropolitan areas: Albany, Atlanta, Macon and Savannah. (Figure-3, Appendix A;) The most populous of these metropolitan areas is the city of Atlanta (Metro Atlanta) which had a population of 4.1 million in 2000 and a population increase to 5.8 million in 2016 (AFF 2017). Metropolitan areas at more risk to extreme heat conditions due to the heat island effect. According to the U.S. EPA, "The term heat island describes built up area that are hotter than nearby rural areas. The annual mean air temperature of a city with 1 million people or more can be 1.8 – 5.4^{o}F (1 – 3^{o}F) warmer than its surroundings. In the evening, the difference can be as high as 22°F (12°C).” (EPA 2017). According to Climate Central, these four metropolitan areas will have a significant increase in the number of days with a heat index above 105°F (Figures 4-7, Appendix A;).

The metropolitan statistical area of Atlanta, GA has a population of approximately 5.8 million and consists of 29 counties (Figure-8, Appendix-A). The overwhelming racial majority in the state of Georgia is comprised of Whites and African-American. Georgia’s population increased in all the demographic categories between 2010 and 2017, except person < 18 years. Because race is always a social vulnerability determining variable, it important to note the significant increase in Georgia’s African-American population. According to the 2010 US Census data, African-Americans were approximately 30% of GA’s population. Hispanics were counted along with African-Americans in this 30% in the 2010 US Census. The African-American population grew to be 34% of GA’s population, not including Hispanics, in 2017. The nearly 2-million-person population growth in GA is largely due to people from the northern part of the United States migrating to the southern part of the United States. The increase in GA’s urban population and landscape can also be greatly attributed to this migration. An increase in urban landscape and urban population results in more intense urban heat islands during extreme heat events and aids to the needs to include changes in land cover when analyzing extreme heat vulnerability. The demographic characteristics of GA are further highlighted in Table I and Table II, Appendix B)

B. <u>Vulnerability</u>

According to Binta, et.al. 2015, “vulnerability to climate change is the degree to which a system is adversely affected by climate related stimuli and its inability to cope with them” (IPCC, 2007). It is typically characterized as some function of exposure, sensitivity, and adaptive capacity (equation listed below). Climatic variations measure exposure of the system; sensitivity is the effect of variations on the system; and adaptive capacity is the ability of a system to adjust to climate related stimuli (IPCC, 2007). The physical causes, that is, exposure and their effects are explicitly defined, and the social context is captured in terms of sensitivity and adaptive capacity (IPCC, 2007). The Intergovernmental Panel on Climate Change (IPCC) Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (SREX) report (IPCC, 2012) provides a slightly different approach to vulnerability such that exposure (referred to as the location of people, livelihoods and assets) and vulnerability are determinants of disaster likelihood. (Binta et. al 2015)

Vulnerability = f(Exposure, Sensitivity, and Adaptive Capacity)

According to Sigh 2014, "Social vulnerability is determined by various factors such as physical, social, economic, and environmental factors or processes, which increase the susceptibility of a community to the impact of hazards. Poverty, occupation, caste, ethnicity, exclusion, marginalization and inequities in material consumption of a society or community also enhance social vulnerability." (Singh 2014)

The four themes used to create the Center for Disease Control and Prevention (CDC) Agency for Toxic Substances and Disease Registry (ATSDR) social vulnerability index are socioeconomic, household composition, minority status/language and housing/transportation (CDC/ATSDR 2011). The CDC is not the only Federal Agency tool that investigates social vulnerability. According to the United States Environmental Protection Agency (U.S. EPA), certain demographic groups, such as those with lower educational attainment, children, the elderly and those with low socio-economic status (SES), appear to be more susceptible to a given exposure to particulate matter (U.S. EPA, 2009b).

The U.S. EPA has created the EJSCREEN environmental justice mapping and screening tool. This interactive mapping tool includes six demographic indicators, twelve environmental justice indicators and a method to combine environmental and demographic indicators into EJ indexes. The demographic indicators are minority population; low income population; linguistically isolated population; population with less than a high school education; population under 5 years of age and population over 64 years of age.

According to the EPA, "One reason for EJSCREEN to focus on potentially susceptible demographic groups is that a large body of research has documented health disparities between demographic groups in the United States, such as differences in mortality and morbidity associated with factors that include race/ethnicity, income and educational attainment." (EPA 2015). The EPA also states, "A growing body of research has shown that demographic factors are associated with susceptibility – certain groups are more impacted by a given level of exposure to certain pollutants. Various groups have shown increased susceptibility to certain pollutants, but further evidence is still emerging in this area and data are limited.

Bakhish (2015) divides the social vulnerability indicators into three categories: sensitivity, the built environment and heat exposure variables. The socio-demographic vulnerability indicators Binta et. al. (2015) focus on are age groups greater than 65 and less than 5; poverty; racial/ethnic minorities; occupation; urban/rural population; female head of household; inmate population; non-English speaking; unemployment; renter population; and dwelling in mobile homes. They also include the physician to population ratio; education level; per capita income; and irrigated land. (Binta et. al. 2015) Binta et. al. (2015) found climate vulnerability to be the highest in some metropolitan Atlanta counties. Binta et. al. (2015) also found the rural Black belt region of Georgia to be especially vulnerable to extreme heat. The rural Black belt region of Georgia consists of counties with poverty greater than 20%. The Black Belt counties stretch from southern Virginia down through the Carolinas, Georgia, Alabama, Mississippi, and over to east Texas (Figure-9, Appendix A;). These counties have higher than average percentages of African-American residents (McDaniel et. al. 2003).

While there is significant overlap in the vulnerability indicators used by government agencies such as the CDC and the U.S. EPA, as well as readily accepted vulnerability indicators used in research, they are all highly based on socio-demographics. These socio-demographic variables can be justified to exemplify those studied within those socio-economic groups as being more vulnerable to every natural disaster, including extreme heat events. For example, the economic wealth of the African-American population GA has historically been low enough that researchers have determined that their wealth is a factor that contribute to their higher morbidity and mortality rates. If these groups are always used to determine vulnerability, then why complete any further social vulnerability analysis? There is a need to update these analyses with the inclusion of a neutral variable that equally affects all the socio-demographic variables. We will introduce land cover as a significant neutral variable that can be used to more accurately identify vulnerable populations in the state of Georgia.

C. Vulnerability vs. Social Vulnerability

Disaster vulnerability is more related to the structural aspects of a population. For instance, people who live on the coast are more vulnerable to flooding, versus people who live inland; while social vulnerability is how the level of vulnerability is affected by socioeconomic factors. For instance, inner-city populations usually have higher asthma rates than rural areas, due to socioeconomic factors such as population density and higher pollution rates. According to Jungman et. al. (2015) “Vulnerability is dynamic and varying across temporal and spatial scales, and may depend on economic, social, geographic, demographic, cultural, institutional, governance, and environmental factors." According to Cutter et.al. (2009) "Social vulnerability is partially the product of social inequalities—those social factors that influence or shape the susceptibility of various groups to harm and that also govern their ability to respond. However, it also includes place inequalities—those characteristics of communities and the built environment, such as the level of urbanization, growth rates, and economic vitality, that contribute to the social vulnerability of places." The Commonly used extreme heat social vulnerability index factors are further discussed in Table XVI Appendix B.

While considering the commonly used extreme heat social vulnerability variables, our analysis includes socioeconomic variables, heat exposure, and newly developed land cover variables. The socioeconomic variables used to derive the EH indicators for this research are:

- median income;
- percent of the population below the poverty level;
- percent of the population 65 years and above in age;
- percent of the population 5 years and below in age;
- percent female head of household with children;
- percent unemployed age 16 years and over;
- percent African-American; percent Non-White;
- percent no vehicle;
- percent no high school education;
- median home value and percent on public assistance;
- percent living in mobile homes.

The social vulnerability indicators chosen for this research are based on the aforementioned sources.

D. <u>Resilience</u>

The World Conference on Disaster Reduction adopted the Framework for Action 2000-2015: Building Resilience of Nations and Communities to Disasters to which highlighted the need for building the resilience of nations and communities to disasters (WCDR 2005). This conference came to a consensus that resilience is “The capacity of a system, community or society potentially exposed to hazards to adapt, by resisting or changing in order to reach and maintain an acceptable level of functioning and structure. This is determined by the degree to which the social system is capable of organizing itself to increase this capacity for learning from past disasters for better future protection and to improve risk reduction measures.” UNISDR. Geneva 2004 (WCDR 2005). There are several definitions of resilience. The United Nations International Strategy for Disaster Reduction (UNISDR) defines resilience as “The ability of a system, community or society exposed to hazards to resist, absorb, accommodate to and recover from the effects of a hazard in a timely and efficient manner” (UNISDR, 2009, p. 24) The Intergovernmental Panel on Climate Change (IPCC) defines resilience as “The ability of a social or ecological system to absorb disturbances while retaining the same basic structure and ways of functioning, the capacity for self-organization, and the capacity to adapt to stress and change” (IPCC 2018). Resilience is a community ability to return to its normal activities after a disaster. Resilience is a multi-dimensional measure. Human / social resilience refers to the (i) absorptive, (ii) adaptive, and (iii) transformative capacities that individuals, households, groups, communities and societies develop to reduce the risk of long-term detrimental impacts induced by specific advert events (shock/stress) (Béné et al. 2012). For the purpose of this research, resilience is defined as a peoples' (individuals, households and communities) ability to adapt, handle and recover from adverse climate changes in a timely and efficient manner.

There are also several definitions of disaster resilience. Disaster resilience was defined at the World Conference on Disaster Reduction as the ability of countries, communities and households to manage change, by maintaining or transforming living standards in the face of shocks or stresses - such as earthquakes, drought or violent conflict - without compromising their long-term prospects (DID 2018).

Disaster resilience is also defined as the ability of individuals, communities, organizations and states to adapt to and recover from hazards, shocks or stresses without compromising long-term prospects for development. According to the Hyogo Framework for Action (UNISDR, 2005), disaster resilience is determined by the degree to which individuals, communities and public and private organizations are capable of organizing themselves to learn from past disasters and reduce their risks to future ones, at international, regional, national and local levels (GSDRC 2018). The Organisation for Economic Co-operation and Development (OECD) determined that disaster resilience is part of the broader concept of *resilience* – 'the ability of individuals, communities and states and their institutions to absorb and recover from shocks, whilst positively adapting and transforming their structures and means for living in the face of long-term changes and uncertainty." (OECD 2013).

According to Bene, there is a need to better understand resilience at all societal levels (individual, households, communities, societies). One necessary step in this process is to better measure resilience. Without being able to measure and/or to monitor resilience, policy makers and societies more broadly will not be in position to identify and support interventions that have more effect on people's ability to respond and to accommodate adverse events. Putting this resilience measurement into practice is therefore a priority (Bene 2013).

Extreme heat is a natural disaster, but resilience to extreme is difficult to measure. Heat cramps, heat exhaustion, and heat stroke are all sudden conditions. The effects of extreme heat are normally acute, unless the extreme heat is sustained for a concentrated period of time, resulting in a disaster.

Table XVII, Appendix B, summarizes information from several studies, covering a span of more than ten years, in terms of the major social vulnerability (SoV) index derivation techniques and the eventual and practical usage of the derived indices (Sambanis 2016). These studies are just a small sample since many of the quoted authors published other, similar, studies that are not listed in Table XVII to avoid repetition (e.g., Cutter et. al., 1996, 2000, 2007, etc.). As seen in Table XVII, use of the derived SoV index as an implicit classification tool to visualize vulnerability is noticeable, providing, thus, sufficient empirical evidence to substantiate the major supposition postulated in this study. This supposition is further substantiated if SoV published reports from national and international agencies and organizations are considered (e.g., McCarthy, 2001; Briguglio, 2003; and Parry et al., 2007). Measurement is the foundation of any scientific inquiry and, to the best of our knowledge, in the field of SoV research; the underlying measurement scale of SoV indices has not been discussed. This examination is required since, as this research will reveal, SoV indices are not always in the same scale due to differences in the derivation approaches. Ultimately, regardless of the apparent scale of measurement, the indices are converted into a cartographic scale of a few only ordered and discreet classes to represent the levels of vulnerability in a conceptually understandable manner (see Table XVII).

5. EXTREME HEAT DATA AND METHODS

According to the U.S. EPA, there are two viable methods of analyzing extreme heat event (EHE) related morbidity. The more conservative method counts only outcomes on EHE days where the attribution information (e.g., primary diagnosis, cause of death) lists excessive weather-related heat exposure or a condition unequivocally associated with excessive heat exposure, such as heat stroke. (EPA 2006) The second method is based on increases in outcomes during EHE periods being compared to long-term averages (EPA 2006).

We have also chosen the most conservative quantification method of using the outcome counts on EHE days where the ICD-10 Codes list excessive heat-related exposure as the cause for death for morbidity analysis for the state of Georgia from 2000 – 2014.

The heat-related ER visits for 2000 – 2014 for Georgia residents were queried and supplied by the Georgia Department of Public Health (GDPH) (GDPH 2017). The query was compiled based on the ICD-9 Codes in Table III of Appendix-B (ICD 2018):

The diagnosis code list is for any occurrence of extreme heat related adverse health effects (not just principal diagnosis). The GDPH searched the diagnosis code list for any occurrence of codes. The heat-related death data for Georgia residents was compiled based on the ICD-9 Diagnosis Codes in Table IV of Appendix B (ICD 2018).

According to some projection models, EHEs are estimated to become longer, more frequent, and more severe (Karl et. al. 2009) (NRC 2010). As a result of this shift, experts expect a dramatic increase in health problems and mortality due to EHEs over time (CDC 2013). Extreme heat has an adverse effect on the human body and can cause heat stress, worsen the symptoms of an existing illness, and in extreme cases can cause permanent health effects and even death (Anderson et. al. 2011) (Helman et. al. 2014). In the United States alone, extreme heat was the cause of more than 7,800 deaths in the period between 1999 and 2009 (CDC 2013). EH was the leading cause of weather-related mortality between 2000 and 2009 (Kochanek et. al. 2012) (CEC 2013). Extreme heat is also likely to affect infrastructure and services causing, for example, power outages and breakdown in public transportation and support services, thus, increasing the frequency of detrimental effects on the community (Atkins 2013). It is estimated that the death toll in the United States would increase by an additional 1,907 cases if the average temperatures increased by 5°F (Bobb et. al. 2014). Using data collected at weather stations as well as global climate models, researchers from Georgia State University projected that most of the Southeastern USA region will have apparent temperatures similar to that of present-day southern Florida, which has a tropical climate. Apparent temperature (AT) in the summer is often referred to as the heat index, what humans perceive the temperature to be based on a combination of humidity and the actual air temperature. Higher apparent temperatures and more extreme heat days could lend themselves to more heat-related illnesses, and potentially more deaths. The research suggests the summer atmosphere may also be more conducive to extremely high ozone concentrations, a hazard to individuals with lung diseases such as asthma, as well as the elderly. Higher summer temperatures are favorable for the growth of mosquitos capable of transmitting viruses, such as dengue, resulting in the potential for transmitting vector related illnesses at rates similar to tropical areas (Diem 2017).

The use of apparent temperate heat data is the optimal heat index to use because it accounts for humidity as well as temperature. The heat index data obtained for the state of Georgia was downloaded from the CDC Wonder website accounts for humidity and is given in apparent temperature (CDC 2017).

My analysis of the climate heat will be derived using the "Average Daily Maximum Air Temperature" from the Center for Disease Control's North America Land Data Assimilation System (NLDAS) Daily Air Temperature and Heat Index (1979-2011) Request. The "Average Daily Maximum Air Temperature" accounts for humidity and the recorded heat temperatures throughout each day (CDC 2017) (Wonder 2019). The heat index temperature set is also referred to as the Apparent Temperature (AT). "Assessing the Performance of a Vulnerability Index During Oppressive Heat Across Georgia, United States" by Grundstien et.al (2003) accounted for the environmental factor of heat by using the maximum and minimum daily apparent temperature (AT) from the National Climatic Data Center. AT is an assessment of what exposed body surfaces feel like in cold, windy, warm and humid conditions (NCEI 2019). According to the National Centers for Environmental Information, apparent temperature is calculated using the following regression equation:

$$AT = -2.7 + 1.04*T + 2.0*e -0.65*v$$

Where AT and T (air temperature) are °C; e is vapor pressure in kPa; and v is 10-meter wind speed in m/sec (NCEI 2019).

The data used to determine Apparent Temperature for the counties in the Atlanta Metropolitan area from 2012 to 2014 was obtained from the University of Georgia's College of Agricultural and Environmental Science; University of Georgia Network (CAES 2019). The weather stations used to obtain the AT are listed in Table VI, Appendix B. The Maximum Air Temperature and the Average Daily Humidity for each weather station for each county that comprises the Atlanta Metropolitan area was used to calculate the AT from 2012 to 2014. According to the National Oceanic and Atmospheric Administration (NOAA) the reported Heat Index is the AT is temperate temperatures. "Heat index combines the effects of heat and humidity. When heat and humidity combine to reduce the amount of evaporation of sweat from the body, outdoor exercise become dangerous even for those in good shape." (NCEI 2019)." The Heat Index, also referred to as apparent temperature, is an estimate of the temperature (in °F) that would similarly affect the body at normal humidity (about 20 percent). For example, if the actual temperature is 100°F with 40 percent relative humidity, the heat index is 110°F meaning the apparent temperature feels like 110°F to the body." (IAState 2019). Heat Index was calculated for the Atlanta Metropolitan Area from 2012 to 2014 using the following NOAA equation (NWS 2019):

$$HI = -42.379 + 2.04901523*T + 10.14333127*RH - .22475541*T*RH - .00683783*T*T - .05481717*RH*RH + .00122874*T*T*RH + .00085282*T*RH*RH - .00000199*T*T*RH*RH$$

Where:

RH = Relative Humidity (%) and

T = Temperature (°F).

These are the corresponding counties that encompass the Atlanta metropolitan area: Barrow, Bartow, Butts, Carroll, Cherokee, Clayton, Cobb, Coweta, Dawson, DeKalb, Douglas, Fayette, Forsyth, Fulton, Gwinnett, Haralson, Heard, Henry, Jasper, Lamar, Meriwether, Morgan, Newton, Paulding Pickens, Pike, Rockdale, Spalding, and Walton.

SoV indices are likely to be at measurement scale that entails classification. The obvious case are indices based on the percentile rank transformation of the original correlated variables. The individual transformed variables as well as their sum (i.e., composite index) are at an ordinal scale of measurement which, by definition, is a classification scale of points from lowest to highest (Abramson and Abramson, 2008; see also Velleman and Wilkinson, 1993 for issues with the traditional scales). This scale provides ordering information regarding where the n^{th} point of interest lies in relation to the others; however, it does not provide information regarding the magnitude of the difference between points. In the context of SoV studies, this is likely to be the preferable scale to express vulnerability since ordering is the only meaningful property of a construct that signifies a potential state (e.g., one county is much more vulnerable than another is). Quantitative statements of differences and comparisons, such as “one county is 3 times more vulnerable than another is”, are rendered meaningless in the context of a construct that signifies a potential. The SoV indices derived by rescaling of the original correlated variables (e.g., Z-scores) and summation has all the characteristics of an interval scale, however, this approach poses interesting challenges since rescaling, which is a linear transformation, will not change the underlying distribution of the original variables.

Performing a principle component analysis (PCA) provides new data composite variables (i.e. principle components) that can be used to determine a population’s vulnerability. However, PCA does not capture the qualitatively differentiating nature of vulnerability of communities in geographic areas and do not provide a practical and reliable planning tool. My research considers social vulnerability to extreme heat as a classification issue. I will follow the classification modeling and performance assessment techniques introduced by Sambanis (2016), which are likely to provide a more accurate analysis of the attributes influencing vulnerability as well as establish a more reliable approach to identify potentially high-risk areas. Georgia, USA decadal data from 2000 and 2010 was used to perform this analysis. The 2000 and 2010 Georgia, US Census variables that we selected in this thesis are listed in Table XVIII, Appendix B.

The data was analyzed in the same manner as Sambanis (2016), “with the use of Microsoft Excel, the Census variables were filtered to utilize only compatible variables available between the two sets of Censuses. In addition, since the Census data is from different time periods, the Census tract relationships were joined in Microsoft Access by selecting the appropriate crosswalk files for a specific Census year from the Longitudinal Tract Data Base (Logan et al., 2012) (Sambanis 2016). Then, by exporting the common Federal Information Processing Standard (FIPS) Census tract codes and aggregating the data by weight, a common set of records was derived for the two periods." (Sambanis 2016).

I used the decision tree (DT) approach introduced by Sambanis (2016). for exploring the potentials of analyzing social vulnerability within a classification framework. According to Sambanis (2016), "Decision tree induction is a well-known and effective classification technique extensively used in several domains. Its major field of application is the data mining and analytics fields where it is used to explore data structures and induce the tree and its rules that will be used to make predictions. In the context of SV studies, the prediction from a classification model could be a vulnerability category (i.e., severity class) based on actual instances of losses which are placed in categories or classes." (Larose 2014) Also, in accordance with Sambanis (2016), "the percentile rank (PR), PCA, and DT approaches will be implemented with the use of the computer program known as IBM® SPSS® Modeler 16.0." (Larose, 2014) (Sambanis 2016).

A widely performed method in the social vulnerability analysis is to subjectively label each one of the derived principle components based on the magnitude of the coefficients of each component vector (e.g., the coefficients of each characteristic vector). Another common practice is to change the sign of the coefficients (e.g., by multiplying with -1) to accommodate the interpretation of the component vector in terms of its contribution to social vulnerability, which is a simple additive model of the principle component scores. Sambanis (2016) introduced a predictive performance matrix (PPM) technique that I will use to optimize the selection of the number of principle components as well as their direction (i.e., sign). According to Sambanis (2016). “the performance comparison is achieved by using a matrix x matrix confusion (or error) matrix. For a given geographic scale, which defines the area of interest, in this case tract, we define the following classification performance metrics: The numbers of correctly classified areas which occur when their instance class (i.e., Target classification of losses) matches the predicted class (i.e., the diagonal elements of the confusion matrix) provide an overall classification (i.e., the diagonal elements of the confusion matrix) provide an overall classification performance measure. The sum of these matching classes divided by the total number of areas, N, yields the Overall Classification Performance (OCP) rate.” (Sambanis 2016).

6. LAND COVER

Land cover/green spaces are a key part of the standard formulation to determine a population's vulnerability to adverse health effects due to extreme heat. According to my research, there has been limited research on the effect of land cover change and its effects on extreme heat vulnerability and resilience as a primary indicator.

Binta et. al. (2015) state: "vulnerability to climate change is the degree to which a system is adversely affected by climate related stimuli and its inability to cope with them" (IPCC, 2007). It is typically characterized as some function of exposure, sensitivity, and adaptive capacity. Climatic variations measure exposure of the system; sensitivity is the effect of variations on the system; and adaptive capacity is the ability of a system to adjust to climate related stimuli (IPCC,2007). According to Sigh 2014, "Social vulnerability is determined by various factors such as physical, social, economic, and environmental factors or processes, which increase the susceptibility of a community to the impact of hazards. Poverty, occupation, caste, ethnicity, exclusion, marginalization and inequities in material consumption of a society or community also enhance social vulnerability." (Singh 2014).

Reid et.al. (2009) used land cover as a variable in her larger equation when mapping the determinants of heat vulnerability in the U.S. Reid used widely accepted demographic variables to determine vulnerability (% population below poverty line; % population with less than high school diploma; % population of race other than white; % population living alone; % population > 65 year of age; and % population $\geq$ 65 living alone.), land cover (% census tract are not covered by vegetation), diabetes prevalence and % households without air conditioning. (Reid et. al. 2009) They created a cumulative heat vulnerability index, which included all of these variables and performed a factor analysis of the vulnerability variables.

Nayak (2017) also used land cover (% land with high building intensity areas and % land that consist of open undeveloped areas) along with other widely used social demographic variable to develop a heat vulnerability index (HVI) for New York City. The results of the HVI were then mapped across NYC to show its special vulnerability (Nyak et. al. 2017).

Weber (2015) created a vulnerability indicator by using Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature data (LST) to map the exposure, social sensitivity, and vulnerability of urban populations in Philadelphia. Weber defined vulnerability as: Exposure + Sensitivity – Adaptive Capacity. Weber determined exposure to be “the extent to which the population of Philadelphia is being exposed to heat events and how the frequency and intensity of these events are changing over time were evaluated using air temperature data from air weather stations in the study area and satellite-derived LST. Normalized Difference Vegetation Index (NVDI) was also examined both for its potential link to shade-providing trees, both of which can create localized cooler temperatures.” (Weber et. al. 2015).

Reid et. al. (2009) also states: “The published literature on mapping heat vulnerability is scant. Vescovi et al. (2005) geographically overlaid climate variables with socioeconomic variables in southern Quebec to estimate current vulnerable populations and then estimated future population vulnerability using climate and population projections. Overall, that study projected that the population at risk will increase. Harlan et al. (2006) investigated physical attributes of the environment, socioeconomic characteristics, and an outdoor human thermal comfort index in Phoenix and found that neighborhoods with the highest temperatures and the least amount of open space and vegetation were also the most socioeconomically disadvantaged. A publication mapped many heat vulnerability variables by county for the state of California (Climate Change Public Health Impacts Assessment and Response Collaborative 2007).

However, they did not make an index or analyze the collocations of their vulnerability variables. All three studies attempted to situate vulnerability in space, but at different spatial scales and with different variables." (Reid et. al. 2009).

As a part of my research, I compared the change in land cover in the state of Georgia from 2001 to 2011. I used this comparison as a determinant for extreme heat vulnerability and resilience. The land cover data was retrieved from the National Land Cover Database (NLCD) and then analyzed via ESRI's ArcGIS Pro Tabulation Area tool. The description of NLCD Class Descriptions that were used to perform a special analysis of the change in land cover in Georgia can be found in Appendix A (Land Cover: Table XIV).

I used the variables 23 and 24 from Table XIV to map the change in land cover for urban land cover (ULC). I used the following variables from Table XIV to determine the total natural land cover (NLC): 11+41+42+43+52+71+81+82+90+95. My research has found that there is a direct correlation between land cover, vulnerability and resilience, notwithstanding the social or widely accepted vulnerability demographic variables.

The difference in land cover was assessed by comparing the following NLCD Class Descriptions:

- Urban land cover (ULC & P.T.ULC): variables 23+24
- Low urban land cover (LULC): variables 21+22
- Total Natural Land Cover (NLC): variables 11+41+42+43+52+71+81+82+90+95
- Material land cover (MLC): variable 31

Material land cover has been included because their areas have a high probability of being inhabited. The likelihood of someone experiencing extreme heat health effects are low. As described above, variable-31 is “Barren Land (Rock/Sand/Clay) - Barren areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, sand dunes, strip mines, gravel pits and other accumulations of earthen material. Generally, vegetation accounts for less than 15% of total cover” (NLCD 2011).

Figures 10 & 11, Appendix-A depicts the spatial distribution of the 2001 and 2011NLCD for the State of Georgia. The color hues represent NLCD classification. (Multi-Resolution Land Cover Character Consortium, 2016).

7. METHODS

My research focused on May to September, because these months normally have the highest chance of excessive heat-related exposure for the state of Georgia. Most heat exposure analyses used the mean temperature to determine extreme heat exposure (Binta et. al. 2015) (EPA 2019). In this project we are proposing to use the monthly cumulative apparent temperature for the months of May to September from 2000 to 2014, this measure is likely to provide a more accurate representation of the exposure conditions due to extreme heat. This metric was calculated by introducing a threshold above which the temperatures are likely to pose a threat to the public. This threshold was the mean of the Average Daily Max Heat Index (m.ADMHI) for each month. The m.ADMHI estimate was subtracted from the daily values, which were above the threshold, and the extreme heat exposure exceedance (EHEE) was calculated. For each year (i) during which the extreme heat event occurred we have the following daily (j) exceedances:

$$EHEE_{imj} = ADMHI_{imj} - m.ADMHI_{im}$$

For the current study, we are introducing the concept of the monthly cumulative EHEE indicator (EHEEI) which is the sum of the daily EHEE above the threshold:

$$c.EHEEI_{im} = \sum_{j=1}^{30} EHEE_{imj}$$

Where:

i = Year, 2000, 2001,

m = month, 1,2 ,3,4 5 (May to Sept)

j = days 1 to 30 or 31

The c.EHEEI provides a measure of the heat exposure which is likely to cause the unwanted outcomes such as EH morbidity and mortality. This exposure metric is in accordance with the pathology of EH related health outcomes since most cases require an exposure duration to manifest themselves. A similar concept has been used in the IH and chemical exposure literature (EPA 2007) (Sexton et. al. 2007). According to Kim et. al. 2018: "we explored cumulative exposure, climate justice, and flood risk with specific reference to community resilience, vulnerability, and social justice characteristics at the county-level within the U.S. Mississippi River basin from 1990 to 2009. Using a basic conceptual model of spatial resilience to climate risks, temporal lag effect of community capacity, urban and rural spatial classification, integrative cumulative exposure, and spatial clustering of risk, they examined spatial climate risk outcomes and the role of community resilience in reducing such risks." (Kim et. al. 2018). The U.S. EPA has established a "Framework for Cumulative Risk Assessment" which defines cumulative risk as "the combined risks from aggregate exposures to multiple agents or stressors, where agents or stressors may include chemicals, as well as biological or physical agents or the absence of a necessity such as habitat. Cumulative risk assessment, then, is an analysis, characterization and possible quantification of the combined risks to health or the environment from multiple agents or stressors." (EPA 2008).

The mean of the morbidity for every county was adjusted for age and then separated into three age categories: Category I – less than 5 years in age; Category II – 5 to 65 years in age; Category III – greater than 65 years in age. We chose to analyze for the 65 and older segment of the vulnerable population in Georgia. This was done because the 65 and over population (CAT III) is a segment in every variable analyzed. For example, there is a 65 and over segment in the African-American, White and percent below poverty level population. The 65 and older population is not only present within every variable, this age group usually has higher vulnerability to adverse health effects during extreme heat conditions. The 65 and older population who experience adverse effects to extreme heat is larger than the 5 and under population in most cases. It is for these reasons that the data morbidity analysis was age-adjusted. We calculated the mean of the morbidity for each county to assess the importance of the land cover derivatives. The mean of the morbidity was calculated in two periods over a five years span. The first period is 2000 – 2004. Period I was selected, 2000-2004 in order to have a better representation (less missing values per month and counties) within each age category and especially, Category III (>65).

The calibration period of 5 years 2000 to 2004 (Period 1) was selected to minimize missing values and secure a spatial distribution of morbidity, which will include the majority of counties. For validation purposes, 2007 to 2011 data was used for the same reasons.

Two basic structures are being used:

Population adjusted morbidity (P.A.Morb) data and age adjusted Morbidity (A.A.Morb); both rates are reported per 100,000.

13061 Clay and 13239 Quitman counties do not have morbidity data for period 1. We will assume that morbidity for these counties from 2007 to 2011 is also zero.

To identify counties that are likely to have the highest risk for Extreme Heat Event Vulnerability (EHEV) based on the input variables and the additive rank model, a basic prerequisite is the use of easily accessible databases and statistical techniques which will not require the involvement of experts.

A concomitant objective is to assess the use of the proposed Predictive Performance Matrix (PPM) as an approach for identifying an optimum EHEV index, which is not readily used in EHE vulnerability analysis literature.

A third objective is to evaluate the usefulness of the three land cover variables (MLC, NLC, and T.ULC) and the c.EHEE that, to the best of our knowledge, are used for the first time as input variables for EHEVI derivation purpose.

The variables are:

- BASIC:
  - KEY_FIPS, MONTH, YEAR, AGECAT (for CAT modeling), Geography
- ENVIRONMENTAL:
  - c.EHEE, P.LULC, P.ULC, P.MLC, P.NLC, P.T.ULC
- SOCIO-ECONOMIC:
  - Total population, 01_MEDIAN_INCOME, 03_PCT_NO_VEHICLE; 04_PCT_PUB_ASSIST; 05_PCT_FEMALEHOUSE_CHILDREN; 06_PCT_5_UNDER; 07_PCT_65_OVER; 09_PCT_AA; 10_PCT_NO_HSEDU; 11_P_BELOW_POVERTY; 12_PCT_UNEMPLOYED_16_OVER; 13_PCT_MOB_HOME; 16_PCT_TOT_POP_WHITE; 02_MEDIAN_HOME_VALUE

Target: Population adjusted morbidity, per year, per month for period 1 (P.A. MORBIDITY_P1); and age-adjusted morbidity rates (A.A. MORBIDITY) to assess prediction of specific categories.

For the calibration data set, i.e., Period 1, we will apply the performance metrics developed by Sambanis (2016) described in the Data and Methods section. In the current project, the major advantage of these metrics will be fully exploited by using them as an "optimization" process to select a rank-based additive model with the best predictive performance in terms of identifying areas at risk for high levels of morbidity due EHE. The following terms are being utilized in accordance with Sambinas (2016):

- FC: Failure Class
- CF: Classification Failure
- OCP: Overall Classification Performance
- OOR: Overall Overestimation Rate
- OUR: Overall Underestimation Rate

8. RESULTS

The trend line in Figure 12 Appendix A, shows that the Average Daily Max Heat Index (ADMHI) have an upward path for the entire state of Georgia. This historical evidence provides justification to believe that heat levels will continue to rise and that the upward trend will continue. Figure 12 also shows an upward trend in the heat levels for the entire state of Georgia from 2000 to 2011. According to Figure 13 Appendix A, the mean of the Average Daily Max Heat Index for the Atlanta Metropolitan Area indicated that the average apparent temperature is above 90 degrees for the summer months. Figure 12 Appendix A shows that the ADMHI from May to September, from 2000 to 2011, has been above the average apparent temperature of 90 degrees. The ADMHI has also been above that 90-degree threshold more than 50% of the years analyzed for the month of September.

Figure 13 Appendix A shows that the mean of the Average Daily Max Heat Index indicates that the average apparent temperature is above 90 degrees more during the summer months for the all of Georgia. There is also a significant amount of days of extreme heat exposure exceedance (EHEE) well above 95 degrees. Table VII, Appendix-B shows that the mean c.EHEE May to September, from 2000 to 2011, has had an increase in higher than normal extreme heat occurances. The increase in the mean c.EHEE are likely due to climate change.

The mean mortality rate for each category for Period I is displayed in the Table VIII, Appendix-B. The mean morbidity rate for Category I (five and under) and Category III (65 and older) are significantly higher than the morbidity of Category II (Between 5 and 65). The higher morbidity rates are further highlighted in Figures 15 &16, Appendix-A. Figure 22 shows a positive correlation between the morbidity for each age category and the cumulative EHEE from

2000 to 2004. The morbidity rates for all of the categories increases with c.EHEE, however, the morbidity rates for Category I & III increases more with c.EHEE than Category II.

Age-adjusted morbidity increased as the median income of this population decreases. The highest counts of morbidity exist when the population's median income is less than $50K annually and the most prevalent when their income is less than $35K per year (Figure 17, Appendix A). Age-adjusted morbidity is more prevalent in populations with a median income of less than $35K when urban land cover becomes a factor. Adjusting for urban land cover also increased the morbidity of the population across the entire socio-economic spectrum.

Figure 10, Appendix A; shows the land cover of Georgia in 2001. The Atlanta-metro area is the most heavily developed along with Augusta, LaGrange, Columbus, Macon, Savannah and Albany. A comparison of the urban developed landscape from 2001 (Figure 10, Appendix A) to 2011 (Figure-11, Appendix A) shows that there has been an increase in the special distribution of the developed land cover in these same areas. A comparison of the population's land cover related EH vulnerability and resilience for the same period results in a direct correlation to the increase in developed urban land cover. As the developed land cover increases, the vulnerability to adverse health effects due to extreme heat of the population of Georgia, especially in the more developed areas, increases. The heat resilience of the same population decreases at the same time. By using LC as a surrogate variable, we can state that the vulnerability of population in Georgia is increasing and the resilience of the population to extreme heat is decreasing as the landscape becomes more developed. The counties with the highest land cover related social vulnerability (LCSV) and lowest land cover resilience are listed in Table XIII, Appendix B.

The selected land cover characteristic is the change in land cover between 2001 and 2011 (percent defLSV and percent LSR are used). The difference from in the total urban land cover from 2001 to 2011 was calculated for each county as follows:

D.LSV = LSV2011 - LSV2001 (for each county)

D.LSR = LSR2011 - LSR2001 (for each county)

As previously stated, according to the U.S. EPA, “many urban and suburban areas experience elevated temperatures compared to their outlying rural surroundings; this difference in temperature is what constitutes an urban heat island. The annual mean air temperature of a city with one million or more people can be 1.8 to 5.4°F (1 to 3°C) warmer than its surroundings.” (EPA 2008).

The change is likely due to signify potential structural changes in EHE Vulnerability/Resilience since land cover (or land use) is closely related to the “*heat island*” phenomenon in heavily urbanized areas (i.e., increase in impervious surfaces such as asphalt cover and cement). A significant property of D.LSV and D.LSR is that these variables are closely related to infrastructure changes, which subsequently are closely related to EHE events (e.g., reduction or increase of green space).

There is a direct correlation between heat LC vulnerability and LC resilience in Georgia. The resilience of the population decreases as the vulnerability of the same population increases. The statistical analysis of the LCV and the LCR is displayed in Figures 18-20, Appendix-A. LCV and LCR have a significant reciprocal correlation to further support the evidence that as vulnerability increases, resilience decreases. For land cover, in essence: R=1/V. Further graphical evidence is provided below to clarify the reciprocal relationship between resilience and vulnerability (Figures 20 & 21, Appendix A;). For example, when vulnerability is 3 the resilience is -6. Figures 23 & 24, Appendix A are graphical representations of Georgia's population adjusted morbidity by year and by month, by period, from 2000 – 2004, respectively. These graphs show an increasing morbidity trend from 2000 – 2004. The monthly graph shows an increasing morbidity trend for the population of GA between May and September (2000 – 2004), with the largest increase coming in July. Table XIX shows GA's population adjusted morbidity (2000 – 2004) by demographic variable.

Table XX, Appendix B summarizes the optimization process and the selection of an additive model based on ranks that is likely to identify and classify regions (counties) susceptible to EHE. Table XX and the metrics (FC, CF, etc..) provide the means to tailor the additive model selection process in terms of operational objectives, for example, optimize the model in terms of the classification failure (CF) metric to avoid underestimation.

A. Principle Component Analysis (PCA) Analysis

A detailed picture of performance is achieved with the used of the predictive assessment matrix (PAM) and performance assessment optimization (PAO) which was first introduced by Sambanis (2016) in an emergency management context and for the first time, herein, in an EHEV context. An example will be resented in the final section comparing the optimum solutions for the three basic methodologies (i.e., Rank, PCA, and DT). From Table XX, we can state that in most cases the addition of the land cover variables improves performance especially in terms of CF. Table XX presents the PCA model selection process based on the proposed metrics and the identification of an optimum solution for the areas that are likely to exhibit detrimental EHE outcomes. Based on similar projects (e.g., Heba, 2015) the above solution with an 41% overall classification performance (OCP) is likely to be in line with the predictive performance of the PCA methodology which, in most cases, outperforms the methodology which is based on ranks. The proposed PA and optimization approach (based on the proposed metrics) provides the empirical evidence to substantiate this statement.

A further improvement is to perform an orthogonal VARIMAX rotation of the PCs (Jolliffe (2002). Detailed results are presented in the Appendix; an overall assessment of the PCA solution is the extraction column, which indicates the proportion of each variable's variance that can be explained by the derived principal components; this column indicates that all the variables are well represented in the new space (before rotation). The component matrix provides the loadings, which are the correlations between the variable and the PC and the VARIMAX rotation, in some cases, improves the interpretability of the resulting PCs.

The new orientation signifies the importance of the land cover variables since in the new dimensional representation (after rotation) they have significant loadings on the second PC. The 1st components is likely to represent the S/E status and race variables whereas the 3rd provides a clear representation of the "sensitivity" variables related to age and exposure (i.e., c.EHEE).

The results and discussion offer the opportunity to clarify the major constituent that achieves the objective of the proposed approach; specifically, the derivation of an EHEV predictive model with an optimum classification capacity in terms of identifying regions with a high EHE morbidity risk. In this context, the component scores (and dimensionality reduction) are the major constituents of this approach (and not the identification of latent continua). The proposed metrics enhance the ability to accomplish this objective.

B. <u>Decision Tree (DT) Analysis</u>

The DT approach for identifying counties that are likely to have the highest risk for EHEV.

According to Sambinas (2016):

“D*ecision tree (DT) algorithms are supervised learning algorithms which recursively partition the input data based on its attributes, until some stopping limit is reached (Larose, 2014). This recursive partitioning gives rise to a tree-like formation. DT are popular tools for classification and prediction that are gaining popularity in many fields.*” A basic premise that makes the DT approach attractive is that “*DT methods are exploratory (not inferential) and non-parametric since they do not require assumptions about the data distribution, scale, and model*” (Sambanis 2016).

In addition, these methods can easily deal with missing data which is a common characteristic of real-world data sets values as well as categorical attributes.

A decision tree starts from the root node and contains internal nodes and leaf (terminal) nodes, all internal nodes have two or more child nodes. The root and internal nodes contain splits, which are the building blocks of the tree formation. The split at each node is described by a decision that depends on one selected feature of an attribute A (e.g., Income > $40,000). The feature for A is selected among all possible ones, and the split is selected among all possible splits, with the objective to minimizing the heterogeneity of the resulting subsamples forwarded to the child nodes. The aim is that the final partitions (terminal leaves of the tree) are homogeneous with respect to the classes.” (Sambanis 2016). The C5.0 algorithm was used to derive the index: “*The C5.0 algorithm which is the most recent version of the ID3.0/C4.5 algorithms developed by Quinlan (1986 and 1993); the improvements are documented by Pang and Gong (2009). This algorithm will be used to explore the predictive performance of this approach for classifying vulnerability.”* (Sambinas, 2016).

Table XXII, Appendix B presents the results from DT (C5.0 algorithm) modeling of the P.A. Morbidity as a target variable with 4 and 5 classes. The low classification failure (CF) rates, confirm the superior performance of this approach for identifying high risk EHE regions; these results are in line with all the previous applications of this DT approach developed by the Sambinas (2016). (Sambanis 2016).

The DT (C5.0 algorithm) in combination with the proposed PAO (performance assessment and optimization) approach provide the means to identify high EHE areas with an improved predictive performance (OCP above 70%). In the previous sections, it is evident that OCP for the commonly applied methods based on ranks and PCA is, in most cases, within a 30% to 40% range. Given that both the DT modeling and PAO approach are not explored in the EHE vulnerability literature (to the best of our knowledge the present project is the only application), the current findings signify the potentials that this approach has in this field.

Focusing on Cat III morbidity for Period 1, encapsulates the DT modeling performance difference and clearly delineates the superiority of this classifier especially in terms of classification failure (CF). From a public health point of view, the CF is a critical metric since a high rate is likely to result in misclassifying areas, i.e., high EHE morbidity risk areas that are classified as low, and preparedness measures to protect the public during EHE public health underserved areas.

The two PA matrices (see Figure 28, Appendix A) further corroborate the superiority of the DT classifier for deriving EHEVI and identifying many GA counties at risk based on the Period 1 EHE since the two critical metrics (OUR and CF), from a public health point of view, are at a very low range. The following graph demarcates the input variables and their importance for the derived DT classification that yields an 88.1% OCP.

9. DISCUSSION

Based on the results of the analysis the top counties in GA with highest land cover urban growth from 2001 to 2011 are listed in Table IX, Appendix B. The Counties in GA with zero or negative urban growth from 2001 to 2011 is also listed in Table XX Appendix B.

The GA Counties with the highest morbidity rates for the 65 years of age and older population (CAT III) with low urban land cover was analyzed. The results are located in Table XI, Appendix B. According to the analysis, Webster county has had the highest morbidity rate for any single month between 2000 and 2004. The highest reported morbidity rate for Webster county was 283 per 100,000 people. Webster county is a small county of approximately 2300 people that covers 210 square miles. Agriculture and forestry are the main industries in Webster County (AFF 2000-1). According to the 2000 US Censes, Webster county was approximately 50% White, 47% African-American and 3% other races. The population of 65 and older was 14%. The median household income was $27,992 and approximately 19% of the population lived below the poverty line (Binta et. al. 2015).

When comparing cumulative extreme heat events in rural counties between 2000 and 2004, there are some counties that frequently have historically had the highest morbidity rates for the 65 and older population:

- Jenkins County, GA (2000 to 2004)
- Atkinson County, GA (2002 & 2003)
- Wilcox County, GA (2000 & 2002)
- Stewart County, GA (2003 & 2004)
- Talbot County, GA (2001, 2003 and 2004)

Jenkins and Johnson counties were also identified as counties with negative urban growth between 2000 and 2010 (Table – X: Counties in GA with zero or negative urban growth).

As previously covered, the rural Black belt region of Georgia consists of counties with poverty greater than 20% and these counties have higher than average percentages of African-American residents (AFF 2000-1). The Black Belt account for 60% of the rural counties with the highest morbidity for this period: These counites are: Webster; Atkinson; Marion; Lanier; Clinch; Talbot; Stewart; Berrien and Screven counties. The non-Black Belt rural counites with the highest morbidity rates during extreme heat months are Jenkins; Schley; Long; Miller; Monroe and Wilcox counites.

The demographics and urban landscape of theses counties did not change much throughout the 2000 to 2004 analysis period. The lack of development and population changes yields similar morbidity results and a repeat of the same counties with higher morbidity results in rural the rural areas of Georgia. For instance, there is no official cooling center in Webster County, GA. My analysis has determined that the residents in these counties, especially those who are 65 and older, are more vulnerable to higher morbidity rates during extreme heat events.

The GA counties with the lowest morbidity rates for the 65 years of age and older population (CAT III) with high urban land cover was analyzed. The results are located in Table XII, Appendix B. When comparing cumulative extreme heat events in urban counties between 2000 and 2004, there are some counties that frequently have lower morbidity rates for the 65 and older population:

- Bibb County, GA (2000 – 2004)
- Clayton County, GA (2001 – 2004)

- Muscogee County, GA (2002 – 2004)
- Gwinnett County, GA (2002 -2003)
- Fulton County, GA (2000 and 2004)

Clayton, Fulton, Gwinnett and Cobb counties are part of the Atlanta-metropolitan area. Richmond county is part of the Augusta metropolitan area. Bibb county is part of the Macon metropolitan area. Muscogee county is part of the Columbus metropolitan area.

As previously discussed, most of the vulnerability analysis include demographic values such as race in their analysis. Race is an important factor, but it is not always an accurate predictor of vulnerability. Richmond County had one of the lowest morbidity rates for the 65 and older population in all the counties in Georgia in 2003. According to the 2000 US. Census, Richmond County was 49.8% African-American and 45.6% White. 21.1% of the households in Richmond County are occupied by someone 65 and older (AFF 2000-3). The median household income was $33,086 and 19.6% of the population lives below the poverty level (CAMO 2015).

The development of these urban areas results in lower morbidity rates. The urbanization of these metropolitan areas includes the development of new housing units (i.e. condos and apartments) with central AC. New homes are also being built and rehabilitated with central air conditioning units. The urban landscape is also being developed in a manner that will reduce the existing urban heat island effects of these metropolitan areas. For example, the city of Atlanta has prioritized 10 impact areas which includes developing land use policies and programs designed to protect greenscapes and their current tree canopy (Binta et. al 2015).

10. VALIDATION

In this dissertation, validation of a SoV index is an activity verifying at what extend the members in a derived vulnerability class are members as well of classes that sustained the hazard realization at a comparable scale. A member in this case is the community/location of interest (i.e., geographic unit), subsequently, validation is the activity to verify whether the vulnerability ranking of the community has a relevance to a past or simulated reality in which the hazard is manifested. In Figure 33, Appendix A, this validation approach is presented in a simplified schematic.

The SoV indices that assign a vulnerability status to loactions, are formulated in the first contextual sphere. This elusive status becomes a reality, (described by a quantitative scale of damages, e.g., property damages in dollars), only if the hazard is materialied and the EHEV classification is verified by a comperable DL classification. Otherwise, it is likely that the resulting comparison will be problematic since the measurement scale of the SoV index, seemingly an interval scale, pertains more to categorical characteristics. Associating categorical, in nature, variables to ratio scale loss variables (e.g., number of displaced households) will likely yield problematic results. This issue has been identified in previous studies, e.g., "*The results showed that both in 2000 and 2010, there were not discernible correlations between vulnerability (including socioeconomic, built-environmental and social vulnerability) and disaster losses at the 5 % level of significance.*" (Zhou et al., 2014; see also Cutter et al., 2003, and Flanagan 2011). Having this issue in mind, the authors are proposing practical characteristics of a disaster vulnerability index. The qualitative characteristics of a vulnerability index have been identified in the disaster reduction literature (de León, 2006) as well as in the climate change literature (Wolf et al., 2011). A modified version of these characteristics in the context of the proposed validation methodology is as follows:

1. *Relevance, reliability, and validity:* the index needs to identify and rank vulnerable locations that realized in a comparable ranking scale, harm, losses, and damages. Ranking must be, as much as possible, equivalent. This characteristic will be used to define performance criteria.
2. *Stability:* the index needs to be able to identify and rank vulnerable locations that "*are stable for some time (or show repeating patterns in dynamics) in order to allow vulnerability reducing measures*" (Wolf et al., 2011).
3. *Feasibility:* the input data sets used to derive the index "*need to be cheap, reliable, recent, routine and at a sufficient spatial resolution*" (Wolf et al., 2011).
4. *Transferability:* the index derivation methodology should be applicable to other regions regardless of the geographic unit.

The above characteristics address practical qualities of SoV indices, at a more fundamental level the authors believe that methodological simplicity must be, as well, a paramount characteristic of EHEV indices since "We may assume the superiority ceteris paribus [other things being equal] of the demonstration which derives from fewer postulates or hypotheses." (Aristotle 1960). The lack of the above characteristics as well as "*A myriad of detailed theoretical definitions of concepts have been proposed but have not helped to clarify the confusion. In fact, by adding more trees to the proverbial forest, these definitions enhance confusion.*" (Wolf, 2011) implies the application of these indices as a decision support component for hazard mitigation planning, although, vulnerability is a sought-after quality in the risk assessment phase of these plans.

Implicitly, validation can only be performed in the realization sphere where vulnerability ceases to exist as a potential state of a community and harm, losses, and damages become an unwanted reality (see Figure 33 Appendix A). For extreme heat events, a basic prerequisite for validating the index is the existance of a realization which will define a comperable scale of harm and damages. Such a realization are the morbidity and mortality rates (Bakhsh, 2015).

A. <u>Proposed Validation Methodology and Metrics</u>

For a specific geographic target area, having *n* locations of interest, $i = 1, 2, \ldots, n$, the SoV index can be regarded as a classifier of unrealized instances. The variable(s) quantifying a hazard realization (HR) event (e.g., morbidity and mortility), are easily converted to comperable classes of realized instances with a data analytic technique known as discritization, or binning. For practical and theoretical reasons, the *n* locations will be classified into comperable, qualitatively, categories for both instances; $c = 1, 2, \ldots, m$. By comparing the locations falling within each category we can derive the so called $m \times m$ confusion (contigency, error, or coincidence) matrix (Lewis and Brown, 2001, Chen et al., 1996, and Bhardwaj and Pal, 2012). For this study we will use the term perfomance assessment (PA) matrix. A major advantage of this approach is that this matrix can be used to define specific classification performance metrics appropriate for Extreme Heat Event Vulnearbility (EHEV) research and hazard mitigation planning. An example of the error confusion matrix that we will use can be located in Figure 34, Appendix A.

We are proposing that the total number of correctly classified members are correctly classified members are those that occur when the predicted (i.e., $C(SoV)_{ij}$) vulnerability class matches the hazard realization (i.e., $C(HR)_{ij}$) class. For example, 5 counties are classified to belong to the lowest SoV group ($i = 1$). Based on the HR classification of a past disaster in terms of property damages, the same counties are classified to belong to the hardest hit group (j = 1). Computationally, the sum of the diagonal cells will provide a metric in terms of overall performance. The sum of these matching classes divided by the total number of areas, n, yields the *Overall Classification Performance* (OCP) rate. This metric is similar to the overall classification accuracy in the remote sensing field of research which has many conceptual similarities with the SoV field in terms of location importance (Lewis and Brown, 2001, Foody, 2002).

If the index is used to allocate extreme heat mitigation resources, misclassification can cause overestimation or underestimation allocation problems. The off diagonal elements of the PA matrix provide valuable metrics to assess quantitatively these misclassifications. For this purpose, the following two metrics are introduced:

1. *Underestimation Error* (UE). A EHEV index misclassification of an actual highly vulnerable location (j = 1) into a non-vulnerable class (i = 3) is likely to have serious consequences for hazard mitigation planning leading to a max classification failure (i.e., CF, c_{1m} element in Figure 35). There is a number of ways to quantify this, in the final analysis risk, underestimation error (REF), within the context of hazard mitigation planning, we are proposing an overall underestimation error rate based on the sum of all the lower diagonal elements divided by n.
2. *Overestimation Error* (OE). A EHEV index misclassification of a non-vulnerable location (j = 1) into a highly vulnerable class (i = max) is likely to result in a waste of valuable recourses similar to the consequences of a false alarm or classification (i.e., FC, c_{1m} element in Figure 35). To quantify this overestimation error, we are proposing an overall overestimation error rate based on the sum of all the upper off diagonal elements in Table II Appendix B divided by n.

The progression of the EHEV index validation to a categorical level, for example, with the conversion of the predicted and reference variables into categories and the application of confusion matrices (Figure 35), provides a detailed information framework within which various aspects of a classifier (i.e. EHEV or SoVP index) are assessed. For this dissertation, we will explore the use of these metrics for EHEV index validation as well as an optimization tool (e.g., selecting input variables to derive the index). Application of other PA metrics such as the Positive Predictive Value, Negative Predictive Value, or exploring levels of categorical agreement with the use of the Kappa coefficient (Cohen, 1960, Warrens, 2015) is a topic of a forthcoming publication by the authors.

a. Heat Disaster Realization Data

Heat related morbidity and mortality data was used because these records are strictly diagnosed and coded using the ICD-9 and ICD-10 diagnosis codes (Tables III and IV Appendix B). These records are also readily available via the Georgia Department of Publc Health. The morbidity variable was transformed to Percentage Fractional Ranks (i.e., each rank is divided by the number of records with valid values and multiplied by 100). For this application, a classifier, **c**, such as the one resulting from rescaling of the original variables (e.g., z-scores) could have been used as well. To simplify the notation all losses and harms will be terms and symbolized as disaster loss index (based on morbidity) for a location, i, of interest takes the form:

$$\text{Eq. 1:}\ \ DLI_i = \boldsymbol{c}_i\,(Morb_i)$$

b. Disaster Loss index: Performance Threshold

The first application of the proposed PA methodology will be to assess the use of a "previous" disaster as an indicator of Disaster loss (DL). For this purpose, the morbidity data of period I will be used to assess the association with the DL period II. To be consistent with the proposed PA methodology, we will be using both realized event with the un/realized dichotomy, though, in reality, both events were realized. The first event defines a new vulnerability status (i.e., locations are thus classified) which need to be validated in terms of a second event (i.e., to confirm the classification). Communities hard hit by EHE or other natural disasters undergo major changes that very likely alters their vulnerability (e.g. installation of AC units).

This DL indicator is probably better defined than the EHEV index and although centering on the loss (i.e., morbidity or mortality) it implicitly contains EHEV features. This simple application of the proposed PA methodology will manifest the usefulness as well of past hazard realizations to identify high-risk areas. In addition, it establishes a real-world performance threshold for EHEV indices. This benchmark is seriously needed for establishing the credibility of EHEV indices as a decision support tool for mitigation and preparedness planning "*On average, every euro spent for reduction and preparedness activities saves between four and seven euros that would have been spent in response to the aftermath of disasters.*" (European Civil Protection and Humanitarian and Operations 2019)

At a practical level, such a performance threshold is a reasonable expectation for any EHEV index with real-world aspirations. A vulnerability indicator based on numerous input variables and in some cases, a relatively, sophisticated derivation approach is expected to perform as good as a naïve DL index based on a previous realization event and a handful of variables.

The two DL indices (DLI.PI and DLI.PII) were derived by taking the Percentage Fractional Ranks of the highly skewed original loss variables (i.e., morbidity). Discretization was performed with an equal-depth (frequency) partitioning algorithm, containing approximately the same number of locations (IBM-SPSS Modeler). The number of partitions (i.e., bins) is a characteristic of the PA methodology that merits further examination. A PA matrix needs to validate indices the way they will be used in practice; subsequently the selected categories must correspond to the classification they will illustrate (i.e., maps). Many published maps related to vulnerability studies use 3 to 7 partitions, a practice which, essentially, adopts and uses the index as a categorical variable with a few only classes. For this demonstration case study, we will be using 4 partitions/categories. The population adjusted morbidity as well as category III (above 65) will be used for each period.

The OCP metric in Figure 35 of Appendix A, establishes a 44.0% threshold of performance for the population adjusted morbidity (43.1% for Category III), implying that for each location a simple model based on the period I disaster loss (morbidity) can identify correctly 44.0% of the comparable DL classes during a “future” event. As DLI occurring within one year could have been used, however, for this case study a five year (i.e., Period II) is used to assure representation for most counties. In addition, the DLI data can also be adjusted for the morbidity trend. This naïve indicator, and its validation with the proposed PA methodology, establishes a source of information that can be used to improve the indices.

B. Performance of EHEV based on Percentage Fractional Ranks.

For this dissertation, the Percentage Fractional Ranks (PFR) approach for deriving an EHEV index will be used (Flannigan 2011; Yoon 2012) and it will be based on the 2000 census data. The assessment of the PFR derivation approach is likely to be valid for several rescaling methodologies as well since the derived indices are highly correlated (Yoon, 2012). The EHEV(PFR) index will be validated with the use of the DL.PI index and the proposed PA methodology. Table XXIII, Appendix B presents the validation results of the EHEV(PFR) index always in the context of our selected target location and realizations. As seen, the OCP is relatively low compared to the performance threshold (PT) for the comparable timeframe case (i.e., 2000). From Table XIII in Appendix B, we can see that the performance of this derivation approach is under the threshold (i.e., PT); in two cases (R1 and R3) the performance is numerically the same, however, a closer look at the PA matrix reveals that the performance is not the same. The stability characteristic of the index is validated by using the best EHEV(PFR) and the period II DL realization (i.e., PFR(00) → DLI.PII). As seen, the performance of the index remains at a low level. One approach to enhance the performance of the EHEV(PFR) index is to add the information conveyed by the DLI.PI and use the composite index to assess its coincidence performance in terms of the DLI.PII. In Table XXIII this modified approach is presented as well and, as seen, the classification is improved; however, it remains at a low level compared to the threshold (PT in Table XXIII). For this case study, timeframe, and realizations the PFR approach for deriving a EHEV index does not satisfy the proposed characteristics of reliability and stability. By using the proposed PA methodology, the performance of this index can be further explored, for example, by adding or removing variables. Similar performances are achieved for morbidity suffered by the Category III population (see Table XXIV, Appendix B).

C. Performance of EHEV based on the PCA derivation

For this dissertation, the PCA derivation approach will be applied. Based on the number of selects principle components (PCs), rotation of the characteristic vectors, and multipliers, several EHEV(PCA) indices will be derived. Beyond the performance characteristics mentioned above (i.e., reliability, stability, etc.), another major objective of the current study is to demonstrate the use of the proposed PA metrics as an optimization tool. With the use of these metrics (e.g., OCP, UE, etc.) the PCA solution can be optimized in terms of its ability to classify the region as close as possible to the classification resulting from the period I and period II disaster loss realizations (i.e., morbidity).

In this study, PCA analysis has been performed with the use of IBM-SPSS Modeler Version 18.1. The principle components and the corresponding scores have been derived with the use of the correlation matrix of the original input variables since their measurement scale are in different units. For this application, we will focus on the OCP and the underestimation metrics due to their significance for public health. The selected PC scores used to derive the EHEV indices are standardized (i.e., mean zero and variance of one). In many cases the sign of the PC is changes (i.e., scores are multiplied by -1); this practice does not alter the variance of the corresponding scores to the PC nor its orthogonality with the other eigenvectors (Jolliffe, 2002).

Selection of, k, PCs, to represent the original input variable space, is initially performed by selecting the PCs with corresponding eigenvalues larger than one (i.e., Kaiser criterion; Kaiser, 1960). As we will see, the reduced dimensionality seems to be a major optimization parameter for deriving a reliable and stable EHEV index.

Table XXII, Appendix A; summarizes the performance of the various EHEV(PCA) derived indices. The overall conclusion, for this configuration (i.e., input variables and realizations), is that with the application of the PA methodology an optimum solution is identified exceeding the performance threshold. The common use of the PCA method for deriving EHEV indices is presented as application cases 1 and 2 (AC in Table XXX); 12 input variables are used (the greater than 65 demographic is excluded in most cases). Three PCs are retained having eigenvalues above 1 and explaining more than 75% of the overall variability. In many cases an orthogonal rotation is used as well to derive a simple structure which, in some cases, yields PCs that are easier to interpret; the most well-known is the VARIMAX rotation introduced by Kaiser (Kaiser 1958 and 1959). The VARIMAX rotation of the PCs improves classification performance.

The criterion for PCs selection (i.e., stopping rule) based on eigenvalues above one and the SCREE plot establish two of the many benchmarks for this task (Jackson, 2003). It seems that for EHEV applications these benchmarks have to be carefully examined since more PCs (having eigenvalues less than one) are likely to improve the performance metrics. Conceptually, the only link connecting the **p** dimensional space defined by the PCs and the corresponding hazard realization is the location. The inductive approach for PC selection does not imply a classification based on space characteristics that will be associated with the comparable hazard realization classification.

The PCA methodology, to a certain extent, has the potential to accomplish the task of establishing an appropriate data structure since the derived PC components represent generalized characteristics of the original data space and it is much easier to add and remove components as opposed to individual variables. Nevertheless, for the objectives of this case study we will remove input variables to explore further improvements of performance. To this end (i.e., Table-XXXII application case 10) the proposed PA methodology achieved an “optimum” solution with 10 input variables (with AA, FEMALEHOUSEHOLD, and > 65 were excluded); this data structure exceeds the performance threshold. It is worth noting is that these results, are much better than those obtained with the percentage fractional ranks derivation approach (see Table XXXII Appendix B). The best performance, does not exhaust the exploration options for an “optimum” since this is not the focus of this study.

D. Stability

Application case 11, in Figure 36 explores the stability characteristic of the "optimum" solution by using the DLI.PI (adjusted morbidity for category III during period I). As seen, the performance of the PCA based index is close to the threshold and the metrics are better than those achieved by the PFR derivation approach. The scores of each selected PCA are standardized, thus under the normality assumption, the mean and variance of the PCA based index will be equal to their corresponding sums (i.e., in this case zero and 5). The DL variable can easily be converted to standardized scores to derive the z-scores of period II adjusted morbidity (CAT III). The composite index based on the EHEV and DLI.PI is validated in application case 12. (Table XXV, Appendix B) As seen, this performance has been improved and exceeds the threshold.

E. Proposed methodology for predicting vulnerability classification

In the previous sections, the EHEV indices derived by the PFR and PCA approaches were validated with the proposed performance assessment (PA) methodology and the introduction of a performance threshold (PT). The PCA approach is likely to be the only one capable of surpassing the PT by 4 percentage points. The purpose of the PT is not to sanction a vulnerability model; the PT aims to provide a minimum acceptable level of performance without qualifying if this performance is adequate, for example, in terms of public health protection goals. The adequacy answer is extracted by examining closer the PA matrix. The lower left side triangle of the PA matrix establishes a metric of public health concern since ideally classification at this critical level should not be underestimated. For the PCA "optimum" solution, this critical underestimation is at a 6.7% level (CF at 0.7%). We term this level as critical underestimation error, CUE, and together with the CF provide metrics to assess the performance related to public health concerns.

Given the critical nature of this classification, the above results are not satisfactory. This was the main reason research was conducted at predicting classification with Decision (or classification) Trees (DT). DT are considered to be a popular approach for deriving classification models and in this study their applicability in the EH vulnerability field of research will be demonstrated assisted by the proposed PA methodology and the threshold performance. To the best of our knowledge, only the research by Bakhsh (2015) and Sambanis (2016) applied this classification methodology to SoV research; however, the issue of performance optimization was not addressed. Decision tree induction is a well-known and effective classification technique extensively used in the business sector (i.e., decision support, customer relationship management, and credit scoring tool), e-commerce, image processing, and medicine. Its major field of application is the data mining and analytics fields where it is used to explore data structures and induce the tree and its rules that will be used to make predictions (Cailas, 2014). In the context of SoV studies the prediction from a classification model such as a DT could be vulnerability category (i.e., severity class) based on actual instances of disaster losses which are placed in categories or classes (Hastie 2009). As stated in the previous sections, this dissertation considers vulnerability as a classification issue and, consequently, the use of these techniques should be expected to provide valuable information on variables influencing vulnerability as well as a reliable mechanism to identify potentially high-risk areas.

Decision tree (DT) algorithms are supervised learning algorithms which recursively partition the input data based on its attributes, until some stopping limit is reached (Hastie, et al., 2009, and Larose, 2014). As shown in Figure 43, this recursive partitioning gives rise to a tree-like formation. DT are popular tools for classification and prediction that are gaining popularity in many fields. In this section, we will cover and discuss the modifications and methodological additions that were introduced in order to apply DT classification to the EHEV research field. A basic premise that makes the DT approach attractive is that "*DT methods are exploratory (not inferential) and non-parametric since they do not require assumptions about the data distribution, scale, and model*"; in addition, these methods can easily deal with missing data which is a common characteristic of real-world data sets values as well as categorical attributes (Cailas, 2014).

A decision tree starts from the root node and contains internal nodes and leaf (terminal) nodes, all internal nodes have two or more child nodes. The root and internal nodes contain splits, which are the building blocks of the tree formation. The split at each node is described by a decision that depends on one selected feature of an attribute (e.g., Median Household Income > $40,000). The specific feature is selected among all possible ones, and the split is selected among all possible splits, with the objective to minimizing the heterogeneity of the resulting subsamples forwarded to the child nodes. The aim is that the final partitions (terminal leaves of the tree) are homogeneous with respect to the classes. The schematics of a simple decision tree is in Figure 37, Appendix A (Cailas, M.D. 2014).

The criterion for choosing the best splitting rule varies from algorithm to algorithm and the optimization measures they apply. For this study, the C5.0 algorithm will be used which performs splits that yield the maximum information gain (Quinlan 1986 and 1993). The C5.0 is a current version of the ID3.0/C4.5 algorithms developed by Quinlan (1986 and 1993). The C5.0 is a current version of the ID3.0/C4.5 algorithms developed by Quinlan (1986 and 1993); the improvements are documented by Pang and Gong (2009). This algorithm will be used to explore the predictive performance of DT as a classifier of EHE vulnerability of Georgia with morbidity as a target variable.

To explore the applicability of DT in the vulnerability research field and identify practical advantages and limitations, the C5.0 algorithm will be applied at a basic level without the implementation of algorithm improvement modifications (e.g., boosting, pruning, twoing, etc.) in order to obtain reproducible results. For this purpose, the proposed performance threshold and the PA methodology will be used to:

- validate the DT derived classification in comparison to the performance threshold (PT),
- optimize the resulting DT classification model in terms of the proposed PA metrics critical to public health (i.e., CF and UE), and
- explore the use of a DT classifier for identifying an input variable set which bears a direct association to the target variable (i.e., morbidity).

The latter application is a main feature of the DT classifiers, which is not feasible from the other approaches used in the SoV field (i.e., PCA and Percentage Fractional Ranks). In this study, the use of the DT variable selection feature will be explored for deriving a EHEV index. In the SoV research field a rational variable selection approach which accounts for the target variable, as opposed to the intercorrelations among the input variables, is a much-needed option, e.g., "*Originally, more than 250 variables were collected, but after testing for multicollinearity among the variables, a subset of 85 raw and computed variables was derived.*" (Cutter, 2003). This algorithm will be used to explore the predictive performance of DT as a classifier of EHE vulnerability of Georgia with morbidity as a target variable.

The PA metrics of the DT application for deriving an EHEV classification are presented in Table XXXI, Appendix B. As seen, the DT approach based on the C5.0 algorithm achieves a high level of classification performance with respect to all the proposed performance metrics. An interesting outcome, corroborating the variable selection potentials of DT, is the application cases 7 and 8 (Table XXXI). As seen from these two cases, the same level of performance is achieved with one variable less. For this specific classification setting (i.e., input and target variables), the percent no high school education variable (%NoHSE) does not have a significant contribution to the overall classification model. Discussion of the input variable importance and selection is given in the following section.

This performance (i.e., AC 8) comes at a complexity cost since this level of performance requires 60 decision tree nodes (see Figure 43, Appendix A). Overfitting is an issue with the C5.0 algorithm; however, there are modifications that can reduce this "complexity" which are far beyond the scope of this study (Pandya and Pandya, 2015). As stated, the objective of this thesis is to use the standard options in order to to publish reproducible results. A prominent finding is the comparison of the predictive performances between the traditional SoV approaches (i.e., PCA and PFR) and the classification performance of the DT C5.0 algorithm. Conceptually, the DT approach is expected to have a better performance since the classification model aims to match the target variable (i.e., PI morbidity). This approach bridges the two spheres by deriving a classification model (i.e., input variables defining the vulnerability potential, unrealized sphere) to describe the classification based on the realization event (i.e., PI morbidity). The expected superior comparative performance is demonstrated in Figure 40, these results corroborate the findings of Sambanis (2016) and Bakhsh (2015). The metrics with a public health significance (i.e., CF and CUE) are at low level. By taking into account the results of this study and the results from studies performed by the UIC-SPH team using DT algorithms for classification of the vulnerability potential (Sambanis, 2016 and Bakhsh, 2015), we propose, for a 4-level classification, optimum levels for the public health related metrics less than 5% for CF and less than 10% for CUE. Closer examination of the PA matrix is highly recommended in order to verify that the classification abides with the practical objectives of the SoV study (e.g., low underestimation might be a desirable outcome as well).

F. <u>**Optimization and variable selection**</u>

The predictor importance estimate is used for input variable selection. The C 5.0 algorithm (IBM-SPSS, 2018), this estimate is used to derive the EHEV classification model, providing an estimate of the weight of the input variables with a range of 0 to 1 (Cailas, 2014). In general, for decision tree algorithms the predictor importance is "*a measure of the amount of output (target) variance that is removed when we learn the true value of the predictor*." (Larose 2014). For the optimum EHEV classification model. The predictor importance graph is presented in Figure 41, Appendix A.

A classification model containing less variables is in accord with the principle of parsimony, however, with DT models less variables does not imply a smaller number of splitting nodes and less rules in the rule set. This is seen with the overall decision tree structure of AC 8 in Figure 43, Appendix A, which contains 60 nodes. For comparison, application case 2 with 13 + 3 variables contains 48 nodes. The ten first predictor importance variables for deriving EHEV classification model C5.0 is shown in Figure 42 Appendix A.

Predictor importance by itself will not suffice to identify an acceptable classification model for EHEV. As a minimum, the OCP metric needs to be at an acceptable level as well as the two PH related metrics. In Table XXXII Appendix B the performance assessment metrics are used in combination with predictor importance (not shown) to identify an optimum classification model (i.e., application case, AC, 8).

G. Stability

The "optimum" DT classification model (AC 8 in Table XXXII Appendix B) will be used to validate its performance in terms of predicting the corresponding classes of the period II disaster loss index. The input variable set for this case (i.e., 2000 census data, 10 variables and 1 land cover) is used to predict the DL.PII classification. The results are presented in Table XXXII Appendix B as application case 11. As seen in in Table XXXII Appendix B, the metrics for predicting the DLI.PII classification are above those of the other methods for assessing DLI.PI classification.

The comparative performance of all the methods applied for this study are presented in Table XXXII. As expected, the DT approach surpasses all the other methods commonly applied for deriving EHEV indices. In addition, the DT approach fulfils the index characteristics especially in terms of the reliability and the PH related metrics. The criterion that the input data set "need to be cheap, reliable, recent, routine, and at a sufficient spatial resolution" (Wolf et al., 2013) is feasible as well with DT, though the PFR methodology is probably the best alternative. This criterion is critical for small-scale public health departments which are not able to perform the preliminary data collection and preparation tasks without the support of specialized experts. The decision tree algorithm can be built into a software application so public health departments can take morbidity target data and import local Census data of their area of evaluation; in addition, there is no large database limitations.

A further application of the proposed DT analysis approach is the use of the rule sets that are generated. These rule sets can be applied directly with a database access language (e.g., SQL or by using a simple queries) so that counties falling into a particular category (i.e., hotspot) may be identified without the use of complicated modeling techniques and expensive experts. The

EHEV decision tree classification can be easily duplicated by public health departments at a rule set level which will require morbidity, exposure (i.e., heat), and local Census data of their area of evaluation. In terms of software program requirements, a simple spreadsheet capable of handling tabular form data will suffice (e.g., Google Sheets, Microsoft Excel, Zoho Sheet, etc.). A comparative performance of DT, PCA and PFR approaches for deriving EHEV indices is located in Figure 45, Appendix A.

H. Visualization

The DT optimum solution is visualized in the maps presented in Figures 46 and 47 of Appendix A. To assist visualization of performance, each map represents the classification difference for each county between the DT classes and those derived from the period I disaster loss (DLI.PI). The differences are mapped at five levels, with zero indicating that there is no difference, which implies that the model derived classification coincides with the DLI.PI classification (i.e., morbidity of period I). A visualize comparison of the two figures delineates the differences between the two methods and substantiates at a visual level the superior performance of the DT classification.

11. CONCLUSIONS

We developed a new variable based on apparent temperature to more accurate assess the extent to which extreme heat poses a threat. This cumulative extreme heat exposure exceedance (cEHEE) matrix was calculated by introducing a threshold above which the temperatures are likely to pose a threat to the public. This new cEHEE calculation matrix resulted in the identification of significant amount of days of extreme heat exposure exceedance (EHEE) well above 95 degrees in GA. An evaluation of the usefulness of the land cover variables Urban land cover, Low urban land cover (LULC), Total Natural Land Cover (NLC) and Material land cover (MLC) was performed. To the best of our knowledge, these variables are used for the first time as input variables for EHEVI derivation purpose. We decided to complete our analysis with an age-adjusted representative population of Georgia within a readily identified segment of the widely accepted vulnerable population.

Based on our research analysis, heat-related climate change is adversely affecting the people in the state of Georgia. Extreme heat conditions in Georgia are increasing the morbidity rates of the total population in GA. Analyzing the monthly cumulative apparent temperature for the months of May to September from 2000 to 2014 provides a more accurate representation of the exposure conditions due to extreme heat. The resulting threshold was the mean of the Average Daily Max Heat Index (m.ADMHI) for each month showed a un upwardly trend in the AT for the state of GA from 2000 to 2014. The newly created concept of the monthly cumulative EHEE indicator (EHEEI) showed an increasing trend in the daily EHEE above threshold.

These indicators were compared to the age-adjusted morbidity rates per county from 2000 to 2004. Both age-adjusted morbidity rates for Category I (< 5 years of age) and Category III (> 65 years old) show a similar increasing morbidity trend. The 65 and over population (CAT

III) is a segment in every variable analyzed. For example, there is a 65 and over segment in the African-American, White and percent below poverty level population. The 65 and older population is not only present within every variable, this age group usually has higher vulnerability to adverse health effects during extreme heat conditions. The 65 and older population who experience adverse effects to extreme heat is larger than the 5 and under population in most cases. It is for these reasons that the data morbidity analysis was age-adjusted.

The analyses were compare to the new c.EHEE variable, socio-economic variables and age-adjusted morbidity rates. In most cases the addition of the land cover variables improves performance especially in terms of Classification Failure (CF). Determining the extreme heat vulnerability of a population by utilizing the newly developed land cover and EHEE variables to the historically used socio-economic provides more accurate identification of groups that are currently vulnerable. The use of land cover also allows public health officials to more accurately predict future vulnerable populations. This method can be applied nationally. Thus, for EHEV index derivation applications, we can conclude that the PC selection rules establish minimum requirements and further exploration is needed for validation purposes with the proposed (or similar) PA methodologies. Within this context, the objective then becomes to find a data structure (in the unrealized domain) which will have a geographic distribution of vulnerability categories that will coincide (i.e., coincidence matrix), as much as possible, with the comparable realized DL categories.

We reviewed the existing methods that are used to identify vulnerable populations and compared them to a newly developed method. The principle component analysis, rank, and decision tree methods were applied to the same variables and compared to determine the best method for determining Georgia's most vulnerable population to extreme heat conditions. My research has proven that there are more accurate methods for identifying populations via Principle Component Analysis, Ranking and Decision Tree Methods. The DT modeling performance difference and clearly delineates the superiority of this classifier especially in terms of classification failure (CF). From a public health point of view, the CF is a critical metric since a high rate is likely to result in misclassifying areas, i.e., high EHE morbidity risk areas that are classified as low, and preparedness measures to protect the public during EHE public health underserved areas. Using Land Cover as a primary variable while performing a decision tree analysis is an optimum EHEV approach. We also performed a two-level validation was to ensure the accuracy of our results and validated the claim that the DT approach has the potential to identify classification models with superior performance.

This is historical evidence provides justification to believe that morbidity rates will continue to rise as heat levels will continue to rise. Our research analysis has identified several counites that have been historically more vulnerable to extreme heat conditions and will more than likely continue to be more vulnerable in the future. Actions should be taken to provide more resources and prevention efforts to decrease the potential higher morbidity rates to these more vulnerable counties in the future.

12. LIMITATIONS

A major feature of the proposed methodology to derive EHEV indices based on DT which requires further exploration is the number of classes used to validate the classification model. Complexity of the decision tree structure is another aspect, however, in the EHEV research field the focus is placed on identifying classes of vulnerable regions and not the interpretation of the structure. An interesting aspect which merits further exploration is the potential to use the rule set to identify new variables based on the major split nodes.

CITED LITERATURE

Anderson, B. G., & Bell, M. L. (2011). *Heat waves in the United States: Mortality risk during heat waves and effect modification by heat wave characteristics in 43 US communities*. Environ Health Perspect. 119(2), 210–218

American Fact Finder, U.S. Census Bureau; *DP-1: Profile of General Demographic Characteristics: 2000*: https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=DEC_00_SF1_DP1&prodType=table; Accessed 3/15/19

American Fact Finder, U.S. Census Bureau*; DP-3: Profile of Selected Economic Characteristics: 2000*; https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=DEC_00_SF3_DP3&prodType=table (last accessed 3/15/19)

Aristotle. Posterior Analytics. Topica. Translated by Hugh Tredennick, E. S. Forster. Loeb Classical Library 391. Cambridge, MA: Harvard University Press, 1960. Book 1, Part 25

Atlanta Metro Chamber of Commerce; Executive Office of the President, Office of Management and Budget (OMB) 2013; *29-County Metropolitan Statistical Area (MSA)*: https://dch.georgia.gov/sites/dch.georgia.gov/files/Atlanta%20Service%20Area%20Map.pdf; Accessed November 12, 2017

Atkins, C. (2013). *Feeling the heat: Heatwaves and social vulnerability in Victoria.* (Vol 17). Melbourne, VIC: Victorian Council of Social Service

Bakhsh, Heba (2015); *Mapping Social Vulnerability and Exposure Parameters to Extreme Heat Events in Missouri*; (2015); (Doctoral Dissertation) University of Illinois at Chicago

Becker, J. A., & Stewart, L. K. (2011). *Heat-related illness*. Am Fam Physician, 83(11), 1325–1330

Becker, J. A., & Stewart, L. K. (2011*). Heat-related illness*. Am Fam Physician, 83(11), 1325–1330

Béné, C., Godfrey-Wood, R., Newsham, A. & Davies, M. (2012). *Resilience: New utopia or new tyranny? Reflection about the potentials and limits of the concept of resilience in relation to vulnerability reduction programmes*. IDS working Paper 405, Brighton: Institute of Development Studies, 61 p

Bene, C.; IDS Working Paper 434 (2013). *What Does "Toward a Quantifiable Measure of Resilience*. Institute of Development Studies. Pgs 6-17

KC, B., Shepherd, J. M., & Gaither, C. J. (2015). *Climate change vulnerability assessment in Georgia*. Applied Geography*, 62*, 66-67

Bobb, J. F., Peng, R. D., Bell, M. L., & Dominici, F. (2014). *Heat-related mortality and adaptation to heat in the United States. Environ Health Perspect*, 122(8), 811–816.

Bouchama, A. A., & Knochel, J. P. (2002). *Heat stroke*. *N Engl J Med*, 346(25), 1978–1988

Cailas, M.D. 2014, HPA 565 Seminar 9 and 10: *Data Mining: Classification Methods*. UIC-SPH, Public Health Informatics Program. Delivered each fall and spring semesters, 2004 – 2014

Cole of Agricultural & Environmental Science; University of Georgia; *Weather Stations*; http://weather.uga.edu/ ; Accessed March 24, 2019

City of Atlanta Mayor's Office of Sustainability; *City of Atlanta Climate Action Plan (2015)*: https://atlantaclimateactionplan.files.wordpress.com/2016/02/atlanta-climate-action-plan-07-23-2015.pdf (last accessed 3/15/19)

Changnon, S. A., Kunkel, K. E., & Reinke, B. C. (1996). *Impacts and responses to the 1995 heat wave: A call to action.* Bull Am Meteorol Soc, 77(7), 1497–1506

Center for Climate and Energy Solutions; *Resilience Strategies for Extreme Heat*; November 2017; pages 3-5

Climate Central: *U.S. Faces Dramatic Rise in Extreme Heat, Humidity*; July 13, 2016: http://www.climatecentral.org/news/sizzling-summers-20515#dangerdays; Accessed October 18, 2017

Centers for Disease Control and Prevention; National Center for Environmental Health (2013). *Climate Change and Extreme Heat Events*. Retrieved from: https://www.cdc.gov/climateandhealth/publications.htm; Accessed June 23, 2018

Centers for Disease Control and Prevention (2013); National Center for Environmental Health. *Climate Change and Extreme Heat Events*. Retrieved from: https://www.cdc.gov/climateandhealth/publications.htm; Accessed June 29, 2018

Centers for Disease Control and Prevention; National Center for Environmental Health; *Climate Change and Extreme Heat Events*; https://www.cdc.gov/climateandhealth/pubs/ClimateChangeandExtremeHeatEvents.pdf; Accessed July 10, 2016.

Centers for Disease Control and Prevention; *North America Land Data Assimilation System (NLDAS) Daily Air Temperatures and Heat Index (1979-2011)* Request; http://wonder.cdc.gov/nasa-nldas.html; Accessed September 20, 2017

Centers for Disease Control and Prevention; *North America Land Data Assimilation System (NLDAS) Daily Air Temperatures and Heat Index (1979-2011) Request*; http://wonder.cdc.gov/nasa-nldas.html; Accessed October 20, 2017

Centers for Disease Control and Prevention. (2018). *The social vulnerability index (SVI): Fact sheet*; CDC. Retrieved from: https://svi.cdc.gov/factsheet.html Accessed 3/15/19

Center for Disease Control and Prevention; *CDC Wonder Daily Air Temperature and Heat Index*; http://www.healthdata.gov/dataset/cdc-wonder-daily-air-temperatures-and-heat-index; Accessed March 24, 2019

Chen, Sarnat, Grundstein, Winquist and Chang 2016. *Time-series Analysis of Heat Waves and Emergency Department Visits in Atlanta, 1993 to 2012*. Environmental Health Perspectives (2017); https://ehp.niehs.nih.gov/doi/10.1289/EHP44; Accessed August 21, 2016

Cooley, H., Moore, E., Heberger, M., & Allen, L. (2010). *Social vulnerability to climate change in California*. Oakland, CA: Pacific Institution.

Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003). *Social vulnerability to environmental hazards.* Soc Sci Q., 84(2), 242–261.

De Leon, JCV. (2006). *Vulnerability: A Conceptual and Methodological Review*; United Nations University, Institute for Environment and Human Security; SOURCE No.4/2006.

Department for International Development (UK); *Aid for the British People; Defining Disaster Resilience: A DFID Approach Paper*: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/186874/defining-disaster-resilience-approach-paper.pdf; Accessed March 10, 2018

Diamant-litd.com; County Map of Georgia USA; *Inspirational New Georgia County Map with Cities Priapro of County Map of Georgia*; http://diamant-ltd.com/county-maps-of-georgia/county-map-of-georgia-usa-inspirational-new-georgia-county-map-with-cities-priapro-of-county-map-of-georgia-usa-cute-county-maps-of-georgia/; Accessed 3/24/19

Diem J., Stauber C., Rothenberg R. (2017) *Heat in the southeastern United States: Characteristics, trends, and potential health impact*; https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0177937 ; PLOS ONE 12(5): e0177937.https://doi.org/10.1371/journal.pone.0177937; Accessed February 6, 2018

European Civil Protection and Humaitarian Aid Operatios; *Disaster Preparedness*; https://ec.europa.eu/echo/what/humanitarian-aid/disaster_preparedness_en; Accessed April 20, 2019

Fekete, A. (2012). *Spatial disaster vulnerability and risk assessments: challenges in their quality and acceptance*. Natural hazards, 61(3), 1161-1178

Flanagan, B.E.; Gregory, E.W.; Hallisey, E.J.; Heitgerd, J.L.; and Lewis, B.; "*A Social Vulnerability Index for Disaster Management*," Journal of Homeland Security and Emergency Management (2011): Vol. 8: Iss. 1, Article 3

Garssen, J., Harmsen, C., & de Beer, J. (2005). *The effect of the summer 2003 heat wave on mortality in the Netherlands. Eurosurveillance*, 10(7), 1–7

Georgia Annual Temperatures and Extremes; National Climate Center; http://coolweather.net/statetemperature/georgia_temperature.htm; Accessed July 10, 2016

Georgia Department of Human Services (2007); *August 2007 report*: https://dhs.georgia.gov/august-2007; Accessed July 10, 2016

Georgia Department of Public Health; Public Health Information Portal; *PHIP Data Request* https://dph.georgia.gov/phip-data-request; Data Queried and received from the GDPH on January 27, 2017

Ghiasi, M. (2011). *Social risk factors for heat wave mortality among the elderly*. Retrieved from http://ghiasi.org/2011/12/social-risk-factors-for-heat-wavemortality-among-the-elderly/ ; Accessed July 10, 2016

Governance and Social Development Resource Centre (2018); GSDRC Applied Knowledge Services; *What is disaster resilience;* http://www.gsdrc.org/topic-guides/disaster-resilience/concepts/what-is-disaster-resilience/; Accessed April 2, 2018

Hajat, S.; Armstrong, B.; Baccini M.; Biggeri, A.; Bisanti, L.; Russo, A.; Kosatsky, T. (2006). *Impact of high temperatures on mortality: Is there an added heat wave effect*? *Epidemiology*, 17(6), 632–638

Hastie TJ, Tibshirani, RJ, Friedman JH. *The Elements of Statistical Learning: Data Mining, Inference and Prediction.* Second Edition. Springer; 2009. ISBN 978-0-387-84857-0

Helman, R., & Habal, R. (2014). *Heatstroke*. Retrieved from: http://emedicine.medscape.com/article/166320-overview#showall; Accessed June 23, 2018

Hogg, Robert V., and Elliot A. Tanis. 2014. *Probability and statistical inference*. Upper Saddle River, NJ: Prentice Hall. (9th edition)

Howe, A. S., & Boden, B. P. (2007). *Heat-related illness in athletes*. *Am J Sports Med*, 35(8), 1384–1395

ICD10Data.com; Exposure to forces of nature X30-39; http://www.icd10data.com/ICD10CM/Codes/V00-Y99/X30-X39; Accessed February 3, 2018

IPCC. (2007). *Climate change 2007: synthesis report*. In Core Writing Team, R. K. Pachauri, & A. Reisinger (Eds.), Contribution of Working Groups I, II and III to the fourth assessment report of the Intergovernmental Panel on Climate Change. Geneva, Switzerland: IPCC (104 pp.)

IPCC. (2012). Summary for policymakers. In C. B. Field, V. Barros, & T. F. Stocker (Eds.), *Managing the risks of extreme events and disasters to advance climate change adaptation. A special report of Working Groups I and II of the Intergovernmental Panel on Climate Change* (pp. 3e21). Cambridge, UK, and New York, NY, USA: Cambridge University Press. doi:10.1016/j.apgeog.2015.04.007

Jardine, D. S. (2007). *Heat illness and heat stroke*. Pediatr Rev, 28(7), 249–258. doi:10.1542/pir.28-7-249

Jolliffe, I. (2014). *Principal Component Analysis*. Wiley StatsRef: Statistics Reference Online

Jongmana, B.; Winsemiusb, H.; Aertsa, J., Coughlan de Pereza, E.K.; van Aalstc, M.;, Krone, W.; Warda, P.; *Declining Vulnerability to River Floods and the Global Benefits of Adaptation*; Proceedings of the National Academy of Sciences May 2015, 112 (18) E2271-E2280; DOI: 10.1073/pnas.1414439112

Karl, T. R., Melillo, J. M., & Peterson, T. C. (2009). *Global climate change impacts in the United States.* Cambridge, United Kingdom: Cambridge University Press.

Karstens, C.; Personal Server Page at Iowa State University (2019); https://www.meteor.iastate.edu/~ckarsten/bufkit/apparent_temperature.html; Accessed March 24, 2019

Kim, S. H., Jo, S. N., Myung, H. N., & Jang, J. Y. (2014). T*he effect of pre-existing medical conditions on heat stroke during hot weather in South Korea*. Environ Res, 133, 246–252

Kim, H., Marcouiller, D., Woosnam, K.; *Rescaling social dynamics in climate change: The implications of cumulative exposure, climate justice and community resilience*. Geoforum 96 (2018) 129-140

Klinenberg, E. (2001). *Dying alone: The social production of urban isolation*. Ethnography, 2(4), 501–531. (UCP Press. 2002) UCP Press. (2002). Dying alone in the heat wave, an interview with Eric Klinenberg. Chicago, IL: Univ Chicago Press

Kochanek, K. D., Xu, J., Murphy, S. L., & Minin, A. M. (2012). *National vital statistics reports deaths: Final data for 2009* (Vol 60). Hyattsville, MD: National Center for Health Statistics

Kovats, R. S., & Kristie, L. E. (2006). *Heatwaves and public health in Europe.* Eur J Public Health,16(6), 592–599

Larose, D. T. (2014). *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons

Logan, J.R., Xu, Z. and Stults, B. 2012. *"Interpolating U.S. Decennial Census Tract Data from as Early as 1970 to 2010: A Longitudinal Tract Database"* Professional Geographer, 66,3: 412-420

Mayo Clinic. (2014). *Heatstroke causes, diseases and conditions*. Retrieved from: http://www.mayoclinic.org/diseases-conditions/heat-stroke/basics/causes/con-20032814; Accessed July 10, 2016

Maier, G.; Grunstein, A.; Jang, Woncheol; Li, C.; Naeher, L.P.; Shepherd, M.; *Assessing the Performance of a Vulnerability Index during Oppressive Heat across Georgia, United States*; American Meteorological Society, 2014; Pages 253 – 263

Martiello, M. A., & Giacchi, M. V. (2010). *High temperatures and health outcomes: A review of the literature.* Scand J Public Health, 38(8), 826–837

McDaniel, J., & Casanova, V. (2003). *Pines in lines: tree planting, H2B guest workers, and rural poverty in Alabama.* Southern Rural Sociology, 19, 73e76

Melillo, JM.; Richmond, T.; Yohe, G.; *Climate Change Impacts in the United States: The Third National Climate Assessment*.; (2014); https://nca2014.globalchange.gov/ ; Last Accessed January 15, 2019 U.S. Global Change Research Program

Merrill, C. T., Miller, M., & Steiner, C.; *Hospital stays resulting from excessive heat and cold exposure due to weather conditions in U.S. community hospitals, 2005*. In: Statistical Brief #55. Healthcare Cost and Utilization Project; (2008) Retrieved from http://www.hcup-us.ahrq.gov/reports/statbriefs/sb55.pdf. Accessed August 19, 2016.

Multi-Resolution Land Characteristics Consortium; *National Land Cover Database 2011 (NLCD2011) Legend*: https://www.mrlc.gov/data/legends/national-land-cover-database-2011-nlcd2011-legend; Accessed 3/15/19

National Center for Environmental Information; National Oceanic and Atmospheric Administration; *Apparent Temperature*: https://graphical.weather.gov/definitions/defineApparentT.html; Accessed March 24, 2019

National Research Council. (2010). *Climate stabilization targets: Emissions, concentrations, and impacts over decades to millennia.* Washington, D.C.: The National Academies Press. Retrieved from: https://www.nap.edu/catalog/12877/climate-stabilization-targets-emissions-concentrations-and-impacts-over-decades-to; Accessed March 23, 2018

National Weather Service/National Oceanic and Atmosphere Administration; *North and Central Georgia/Summer 2016 Climate Summary*; http://www.weather.gov/ffc/summerclimatesummary2016 Accessed July 10, 2016

Nayak, S.G.; Shrestha, S.; Kinney; P.I.; Ross, Z.; Sheridan, S.C.; Pantea, C.I.; Hsu, W.H.; Muscatiello, N.; Hwang, S.A.; *Development of a heat vulnerability index for New York State*; The Royal Society for Public Health (2017). https://doi.org/10.1016/j.puhe.2017.09.006 ; Accessed March 3, 2019

NWS Office of Climate W and WS. (2014). NWS Office of Climate, Weather and Water Services; *Weather Fatalities 2018*; Retrieved from: http://www.nws.noaa.gov/om/hazstats.shtml; Accessed August 14, 2016

National Weather Service; National Oceanic and Atmospheric Administration; *Heat Safety Tips and Resources*; http://www.nws.noaa.gov/om/heat/index.shtml#heatindex; Accessed September 23, 2017

National Weather Service; Weather Prediction Center; *The Het Index Equation*; https://www.wpc.ncep.noaa.gov/html/heatindex_equation.shtml; Accessed March 24, 2019

Organisation for Economic Co-operation and Development (2013b). *What Does "Resilience" Mean for Donors? An OECD Factsheet*.

Pooler, C. (2009). *Porth pathophysiology: Concepts of altered health states*. Philadelphia, PA: Lippincott Williams & Wilkins

Preliminary Natural Hazard Statistics: Weather Fatalities; National Oceanic and Atmospheric Administration; Available: http://www.weather.gov/om/hazstats.shtml; Accessed July 10, 2016

Refsgaard J.C., Storm B. *Construction, Calibration and Validation of Hydrological Models. In: Abbott M.B., Refsgaard J.C. (eds) Distributed Hydrological Modelling*. Water Science and Technology Library, vol 22. (1990) Springer, Dordrecht

Reid, C.; O'Neill, S.; Grolund, C.; Brines, S.; Brown, D.; Diez-Roux, A.; Schwartz, J. (2009); *Mapping Community Determinants of Heat Vulnerability*; Environ Health Prespect 117:1730-1736 (2009). Doi: doi:10.1289/ehp.0900683 available via http://dx.doi.org/; Accessed March 3, 2019

Revich, B. A., & Shaposhnikov, D. A. (2012). *Climate change, heat waves, and cold spells as risk factors for increased mortality in some regions of Russia.* Stud Russ Econ Dev, 23(2),195–207

Rice, D. (2015). *Toll tops 1,150 in Pakistan's deadliest heat wave on record.* USA Today. Retrieved from: https://www.usatoday.com/story/weather/2015/06/26/pakistan-heat-wave/29325177/; Accessed August 17, 2016

Robine, J., Cheung, S. L., Roy, S. Le., Van Oyen, H., & Herrmann, F. R. (2007). *Report on excess mortality in Europe during summer 2003.* Statistics (Ber), (February),15

States at Risk; America's Preparedness Report Card 2015: Georgia; Statesatrisk.org: http://reportcard.statesatrisk.org/report-card/georgia/extreme_heat_grade; Accessed October 2, 2017

Sambanis, A. (2016). *Critique of Social Vulnerability Indices and Opportunity for Improvement*; Thesis - University of Illinois at Chicago

Sexton, K.; Hattis, D.; *Assessing Cumulative Health Risks from Exposure to Environmental Mixtures – Three Fundamental Questions. Environmental Health Perspectives*, 2007. 115(5): 825 – 832

Shaposhnikov, D. A., Revich, B. A., Bellander, T., Bedada, G. B., Bottai, M., Kharkova, T. . . . Pershagen, G. (2014). *Mortality related to air pollution with the Moscow heat wave and wildfire of 2010*. Epidemiology, 25(3), 359–364

Simón, F., Lopez-Abente, G., Ballester, E., & Martínez, F. (2005). *Mortality in Spain during the heat waves of summer 2003*. Eurosurveillance, 10(7).

Singh, S. R.; Eghdami, M.; Singh, S. (2014) *The Concept of Social Vulnerability: A Review from Disasters Perspectives*; International Journal of Interdisciplinary and Multidisciplinary Studies, Vol 1, No. 6, 71-82

State of Georgia: Office of Planning and Budget: *Georgia 2030 – Population Projections*: http://www.georgialibraries.org/lib/construction/georgia_population_projections_march_2010.pdf; Accessed October 2, 2017

UCP Press. *Dying alone in the heat wave, an interview with Eric Klinenberg*; Chicago, IL: Univ Chicago Press. (2002)

International Strategy for Disaster Reduction; 2*009 UNISDR Terminology on Disaster Risk Reduction*; United Nations: https://www.unisdr.org/files/7817_UNISDRTerminologyEnglish.pdf; Accessed January 23, 2018

U.S. Census Bureau 2000; *Georgia 2000; Census 2000 profile; US Census 2000; Table DP-1*: https://www.census.gov/prod/2002pubs/c2kprof00-ga.pdf; Accessed September 20, 2017

U.S. Census Bureau 2010; *Georgia 2010: 2010 Census of Population and Housing; Table-1*: https://www2.census.gov/library/publications/decennial/2010/cph-2/cph-2-12.pdf; Accessed September 20, 2017

US Census Bureau; *New Census Bureau Population Estimates Show That Dallas-Forth Worth-Arlington Has Largest Growth in the United States* (2018); Table-7: https://www.census.gov/newsroom/press-releases/2018/popest-metro-county.html; Accessed September 20, 2017

U.S. Census Bureau; *Quickfacts: Georgia*: https://www.census.gov/quickfacts/GA; Accessed September 27, 2017

U.S. Census Bureau; *QuickFacts – Georgia*: https://www.census.gov/quickfacts/fact/table/GA/PST045217; Accessed October 18, 2017

U.S. EPA Office of Atmospheric Programs: *Excessive Heat Events Guidebook*; EPA 430-B-16-001(June 2006); Updated Appendix A (March 2016). https://www.epa.gov/sites/production/files/2016-03/documents/eheguide_final.pdf; Accessed June 25, 2018

U.S. EPA; *Concepts, Methods, and Data Sources for Cumulative Health Risk Assessment of Multiple Chemicals, Exposures and Effects: A Resource Document (Final Report, 2008)*; U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-06/013F, 2007

U.S. EPA; *Climate Change Indicators: U.S. and Global Temperature* https://www.epa.gov/climate-indicators/climate-change-indicators-us-and-global-temperature; Accessed September 10, 2016

U.S. EPA; *Heat Island Effect*: https://www.epa.gov/heat-islands; Accessed October 18, 2017

U.S. EPA; *EJSCREEN Technical Documentation*. Page-18: http://www.epa.gov/sites/production/files/2015-05/documents/ejscreen_technical_document_20150505.pdf; Accessed November 12, 2017

U.S. EPA; *Adapting to Heat*; https://www.epa.gov/heat-islands/adapting-heat; Accessed April 25, 2018

U.S. EPA; *Reducing urban heat islands: Compendium of strategies;* Draft. https://www.epa.gov/heat-islands/heat-island-compendium; pgs.1-14; Accessed March 15, 2019

U.S. EPA; *Reducing urban heat islands: Compendium of strategies*. Draft; 2008. https://www.epa.gov/heat-islands/heat-island-compendium; pg.1. Accessed March 15, 2019

Yoon, DK. *Assessment of social vulnerability to natural disasters: a comparative study*. Nat Hazards, (2012) 63:823–843

Weber, S.; Sadoff, N.; Zell, E.; Sherbinin, A; Applied Geography 63 (2015) 231 – 243.

World Conference on Disaster Reduction (2005): http://www.unisdr.org/2005/wcdr/intergover/official-doc/L-docs/Hyogo-framework-for-action-english.pdf; Accessed January 23, 2018

Winquit, Grundstein, Chang, Hess, and Sarnat. *Warm season temperatures and emergency department visits in Atlanta, Georgia;* Environmental Research 147 (2016) 314 – 323

Wolf, T., & McGregor, G. *The development of a heat wave vulnerability index for London, United Kingdom;* Weather and Climate Extremes, 1, (2013) 59-68

Wu, X., Brady, J. E., Rosenberg, H., & Li, G. *Emergency department visits for heat stroke in the United States, 2009 and 2010*; Inj Epidemiol, (2014) 1(8), 1–8.

APPENDICES

APPENDIX A - FIGURES

FIGURE – 1: Preliminary weather fatalities – 2015

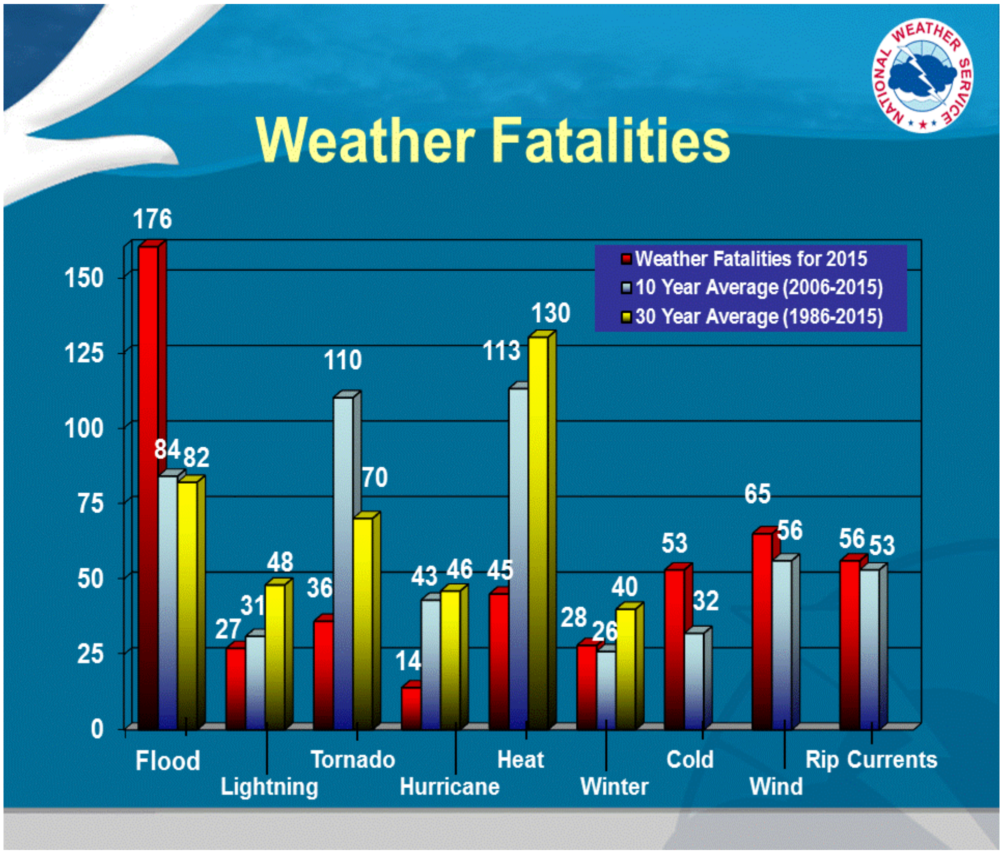
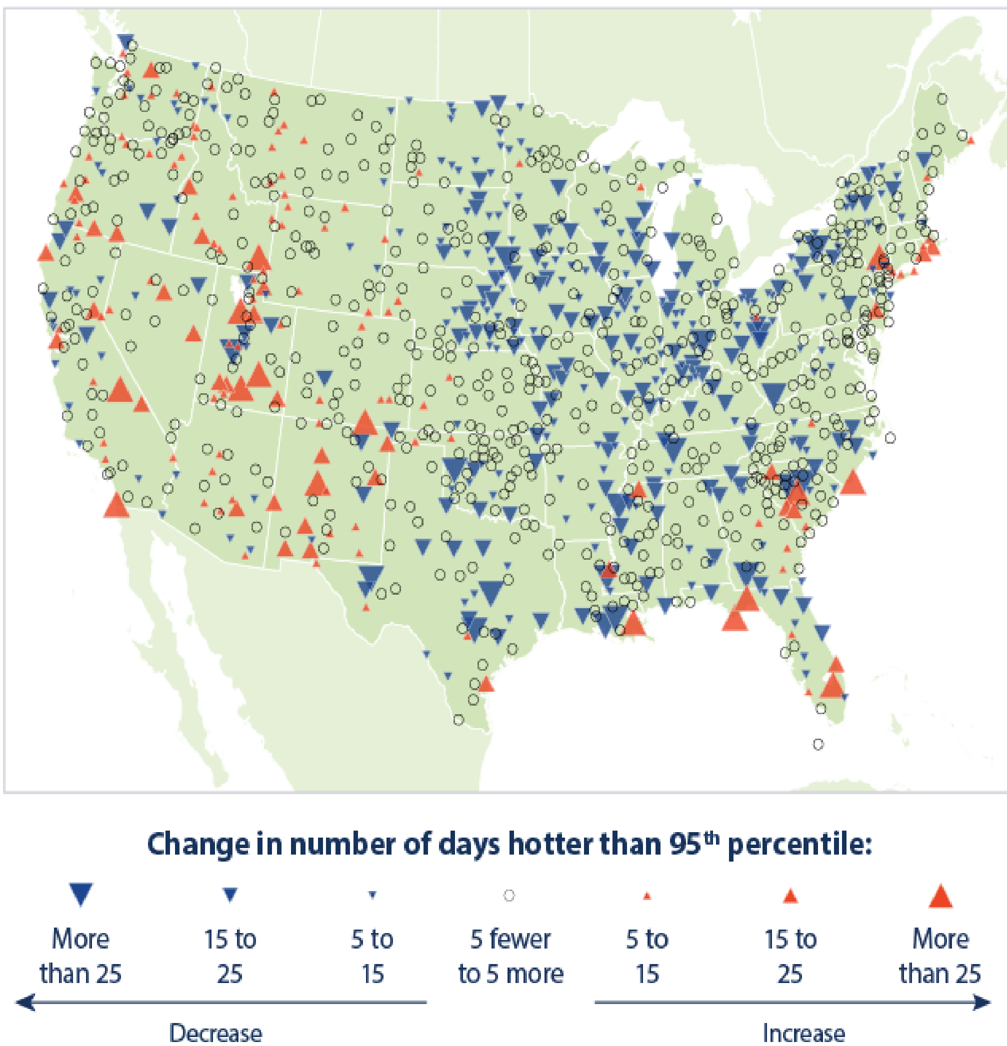


FIGURE – 2: Change in Unusually Hot Temperatures in the Contiguous 48 States, 1948-2014

Change in Unusually Hot Temperatures in the Contiguous 48 States, 1948–2014



Data source: NOAA (National Oceanic and Atmospheric Administration). 2015. National Centers for Environmental Information. Accessed June 2015. www.ncdc.noaa.gov.

For more information, visit U.S. EPA's "Climate Change Indicators in the United States" at www.epa.gov/climatechange/indicators.

FIGURE – 3: Metropolitan map of Georgia

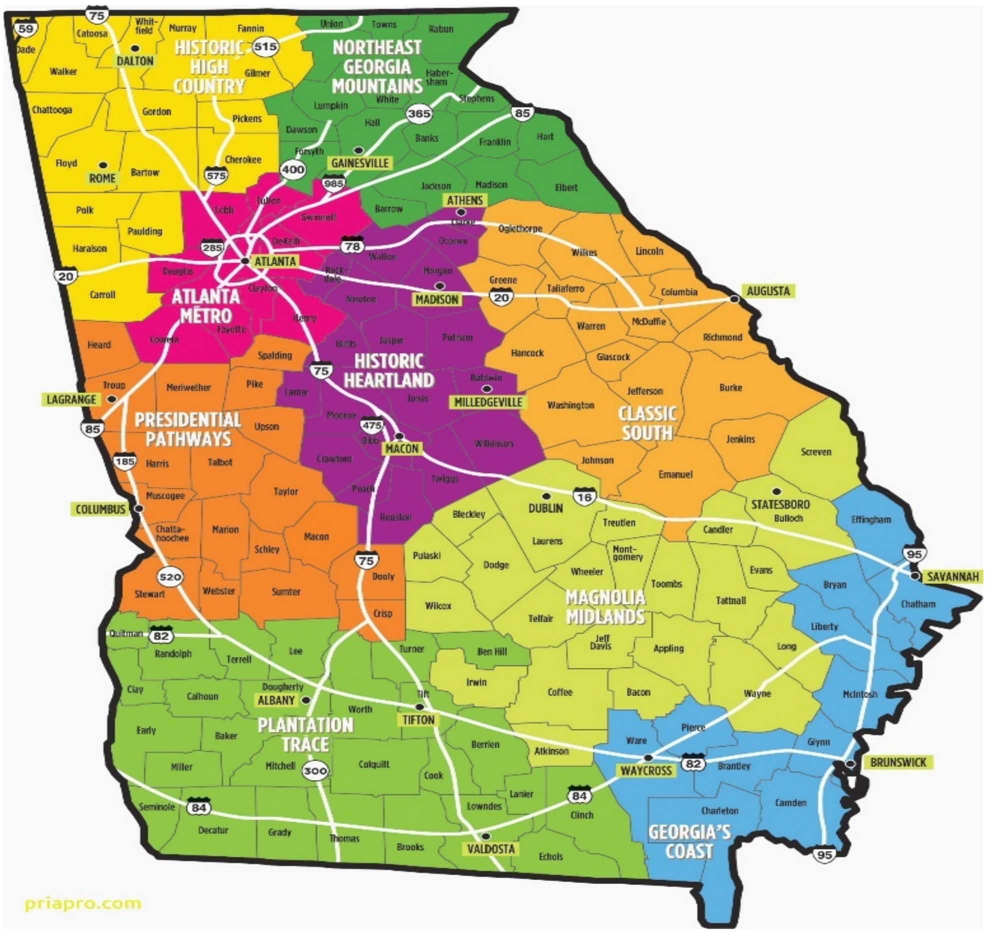


FIGURE – 4: More Danger Days: Heat Index Above 105º (Atlanta Metro)

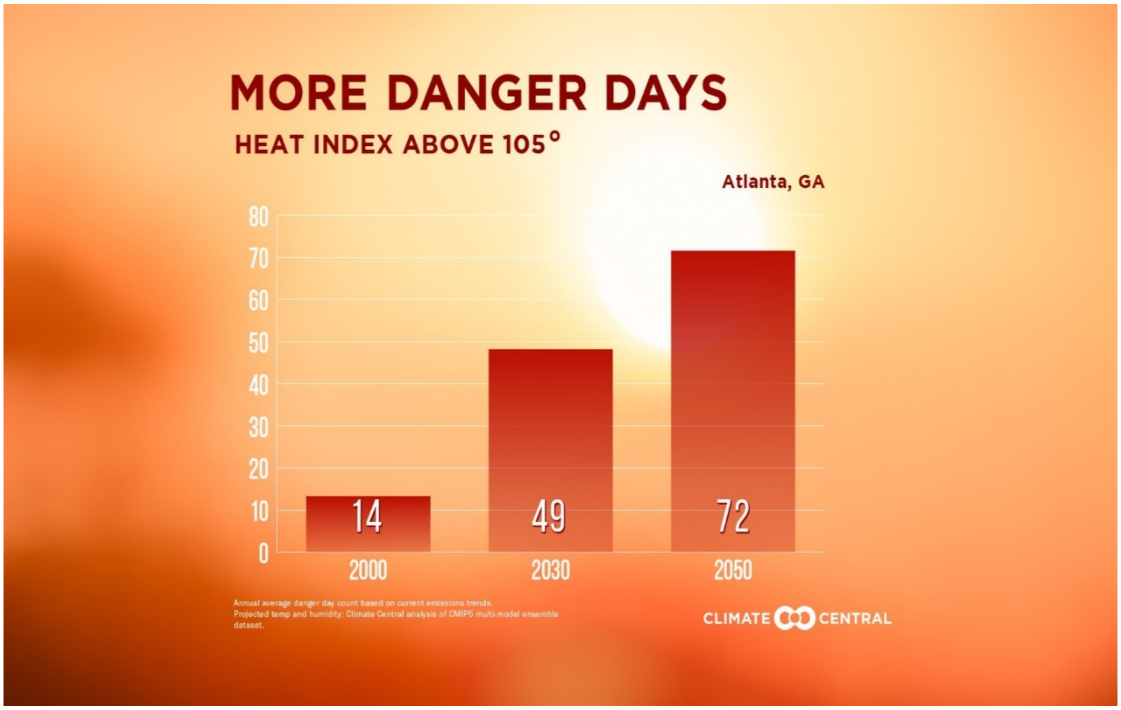


FIGURE – 5: More Danger Days: Heat Index Above 105º (Albany Metro)

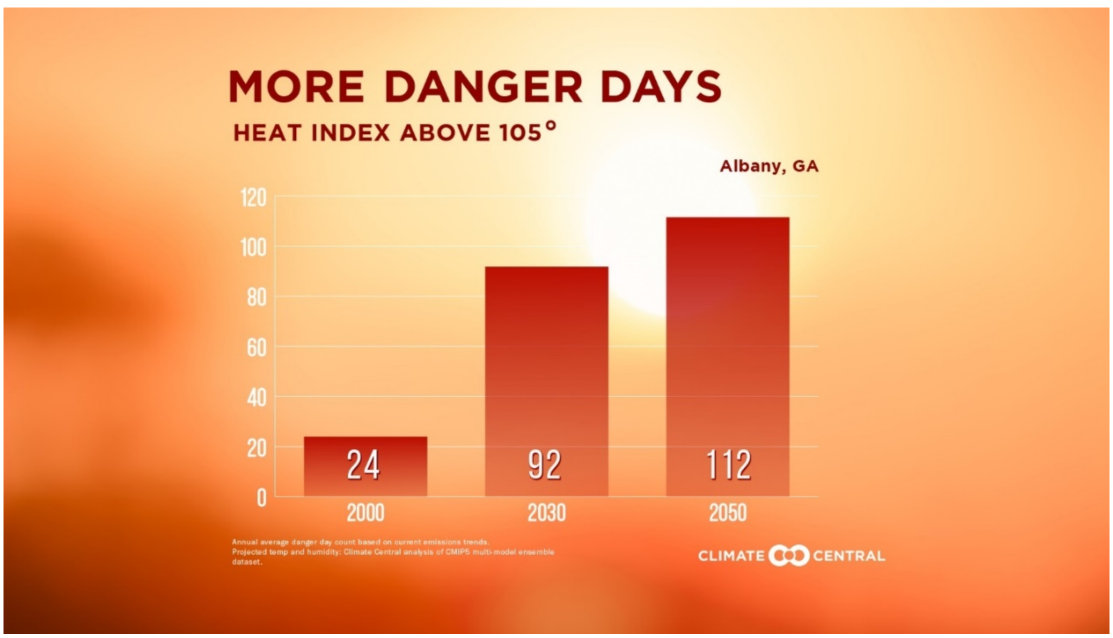


FIGURE – 6: More Danger Days: Heat Index Above 105º (Macon Metro)

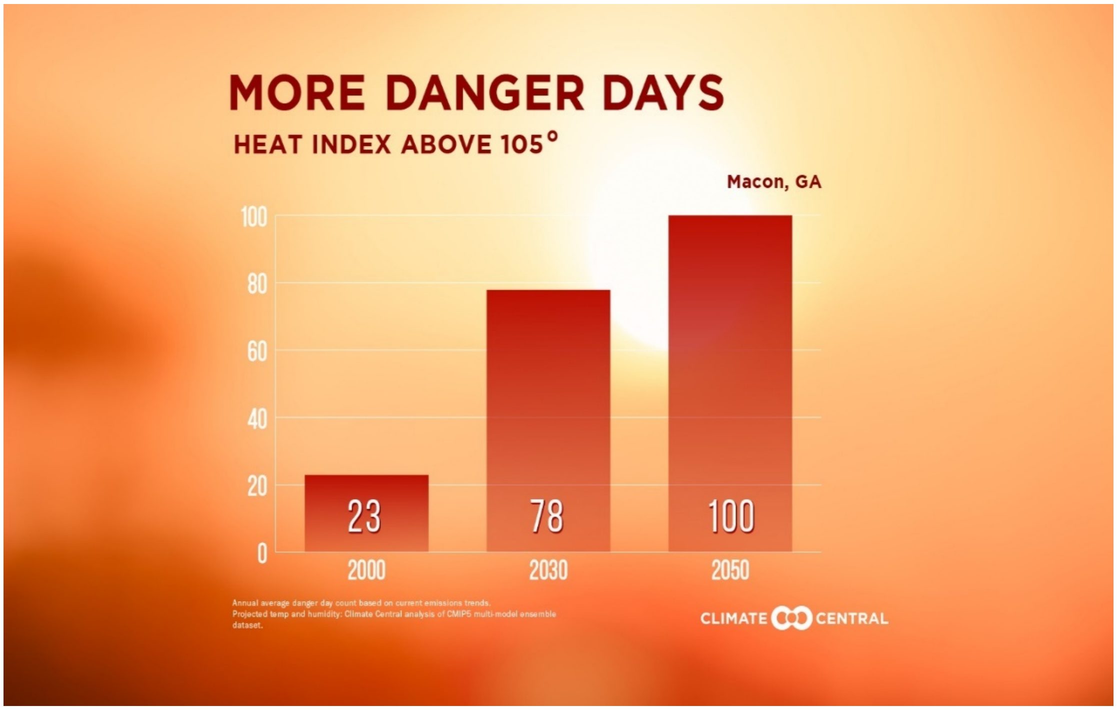


FIGURE – 7: More Danger Days: Heat Index Above 105º (Savannah Metro)

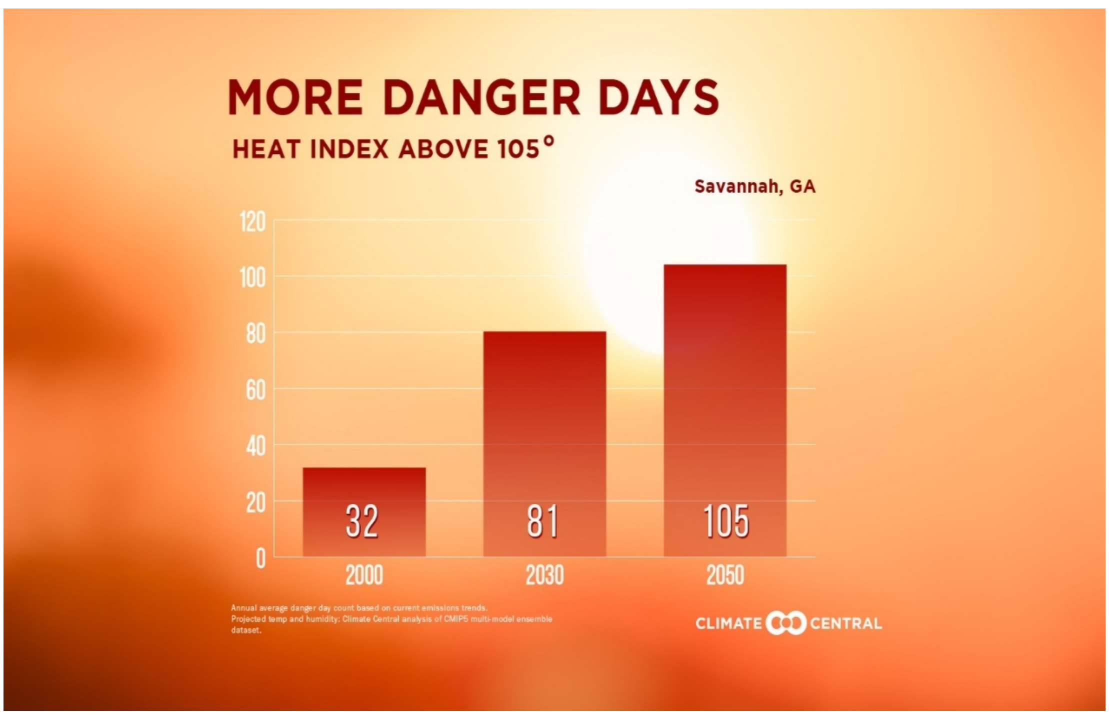


FIGURE – 8: Atlanta metropolitan area counties

FIGURE – 9: GA black belt counties

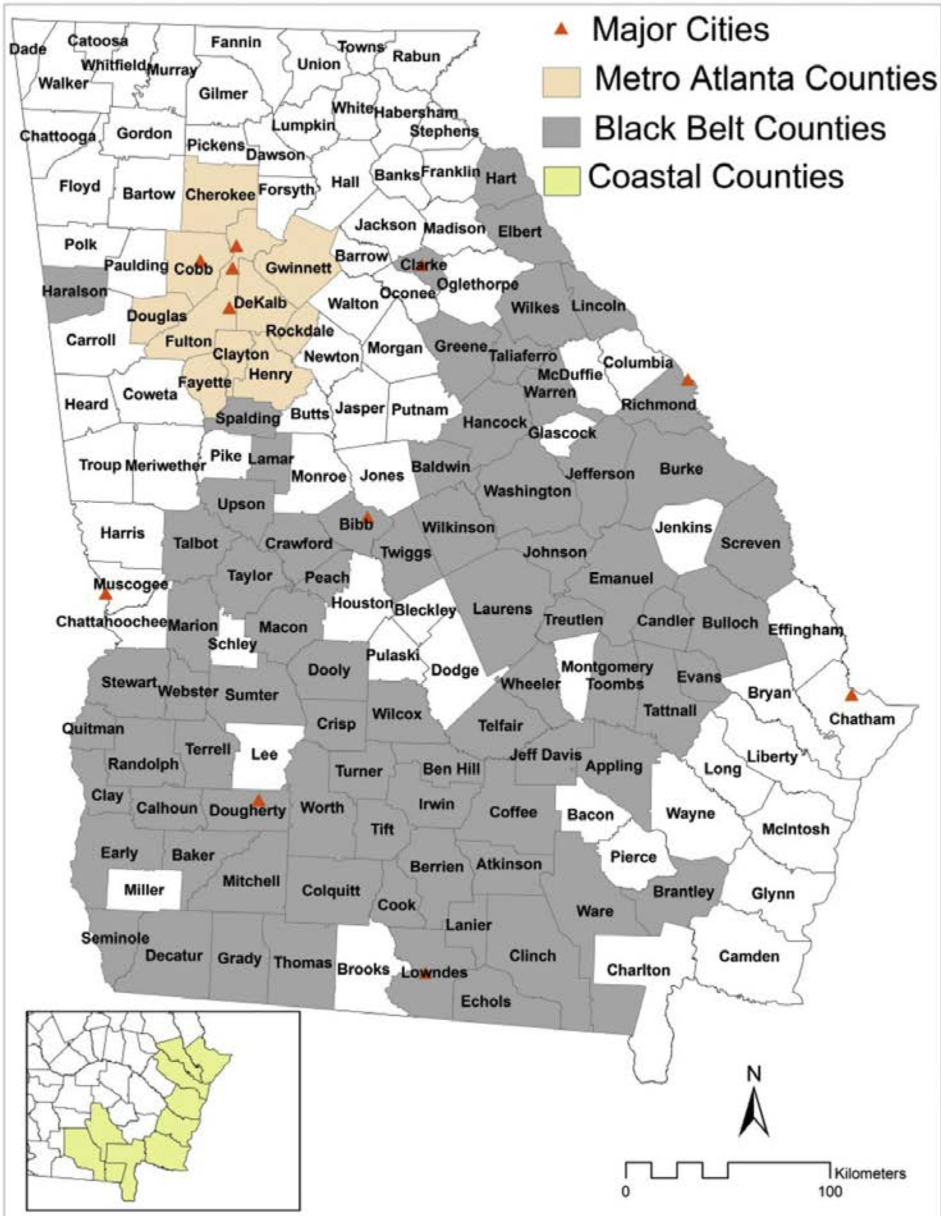


FIGURE – 10: Spatial distribution of the 2001 NLCD

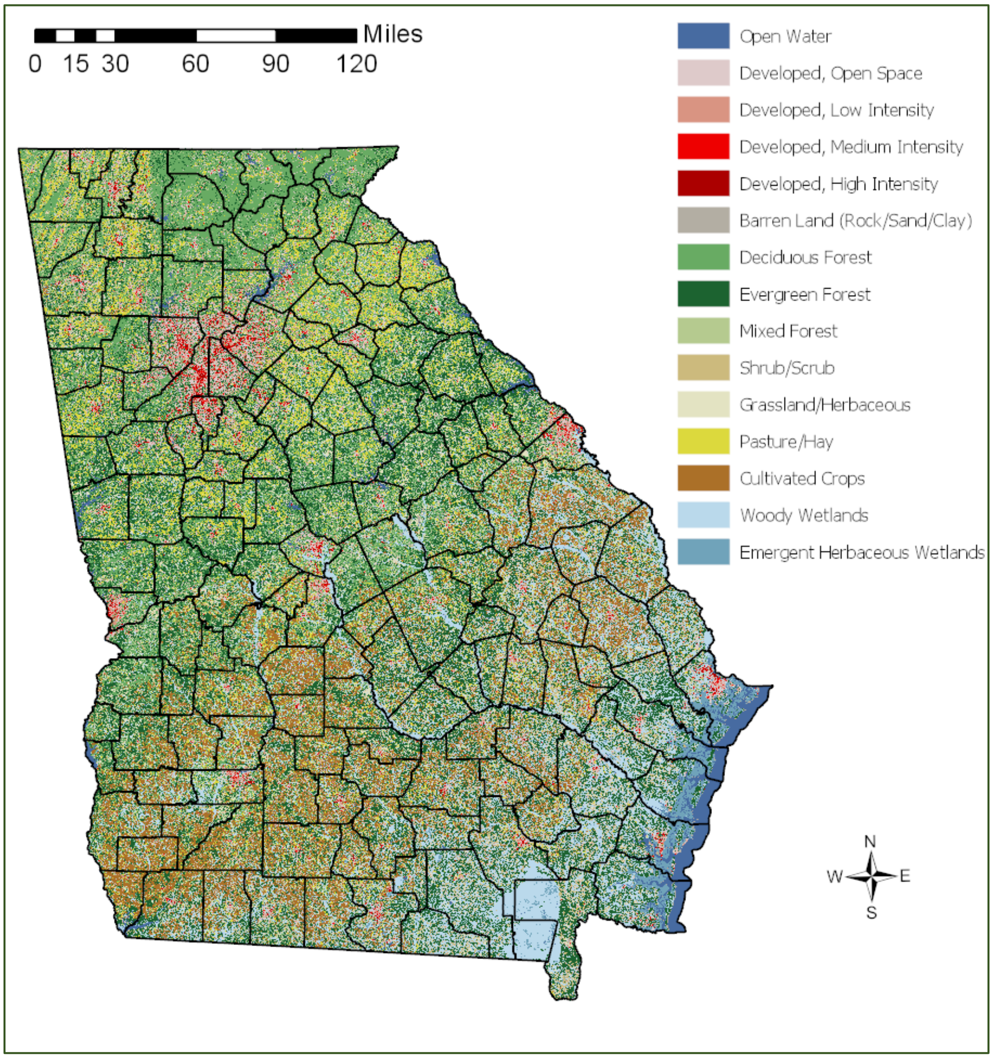


FIGURE – 11: Spatial distribution of the 2011 NLCD

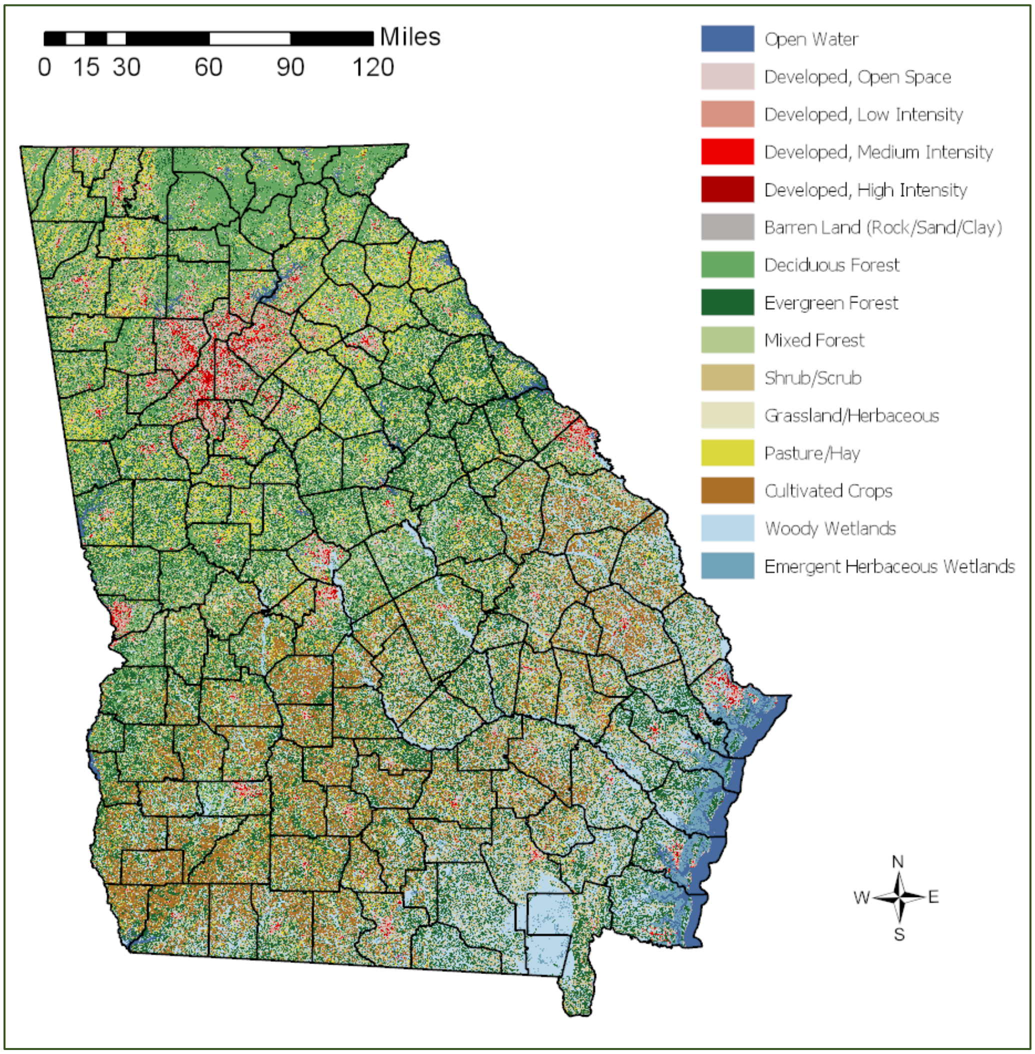


FIGURE – 12: ADMHI for each year (all counties)

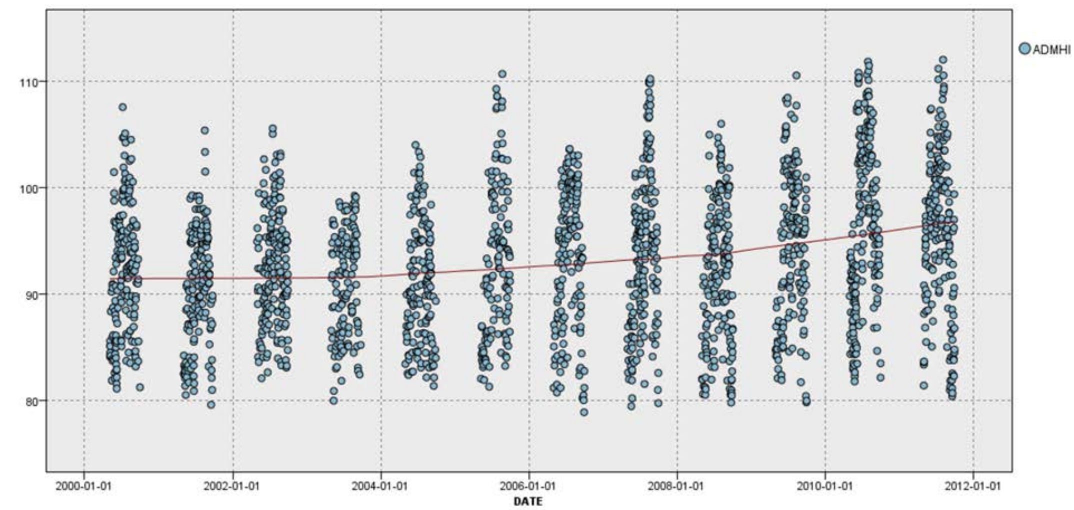


FIGURE – 13: Histogram of the monthly mean ADMHI for the Atlanta-Metropolitan area

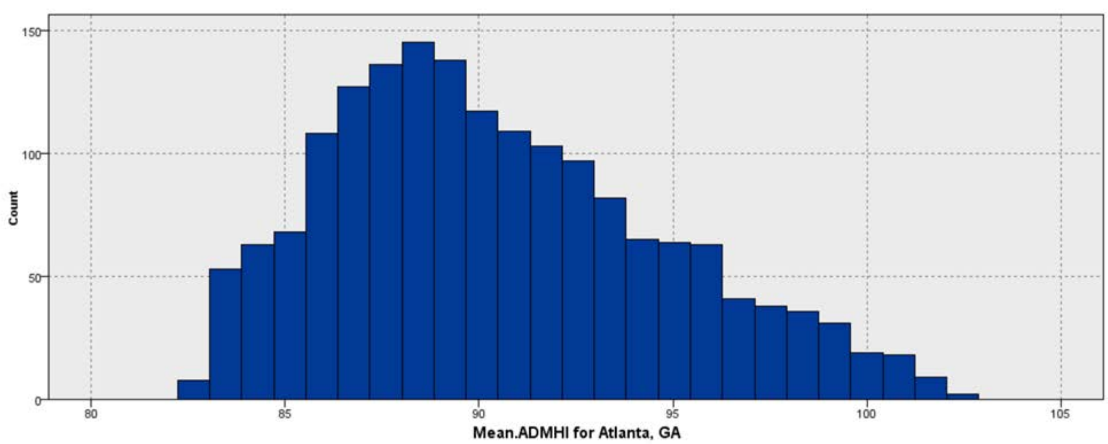


FIGURE – 14: Histogram of the cumulative EHEE (all counties)

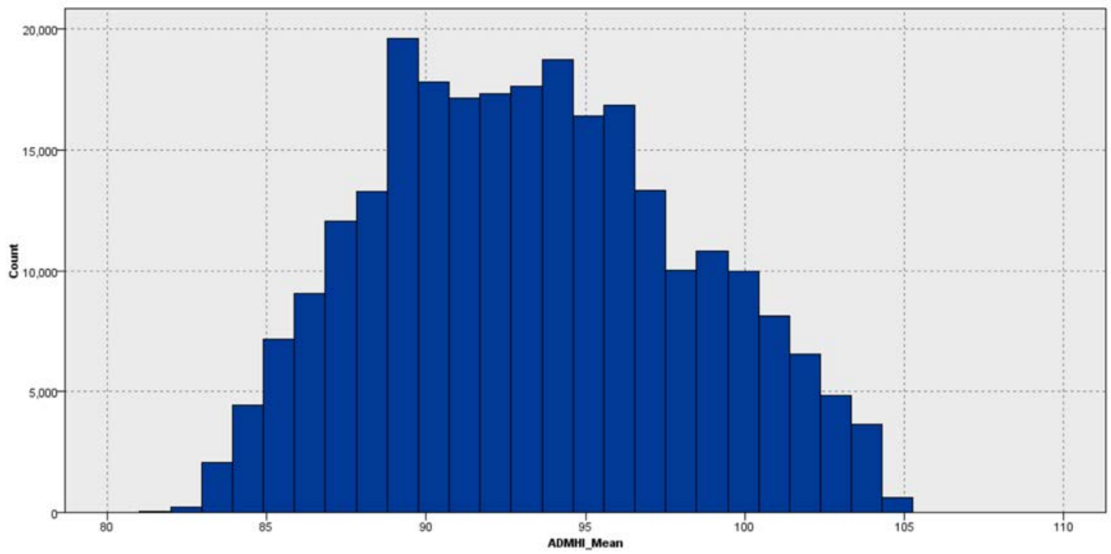


FIGURE – 15: Age-adjusted morbidity for period-I

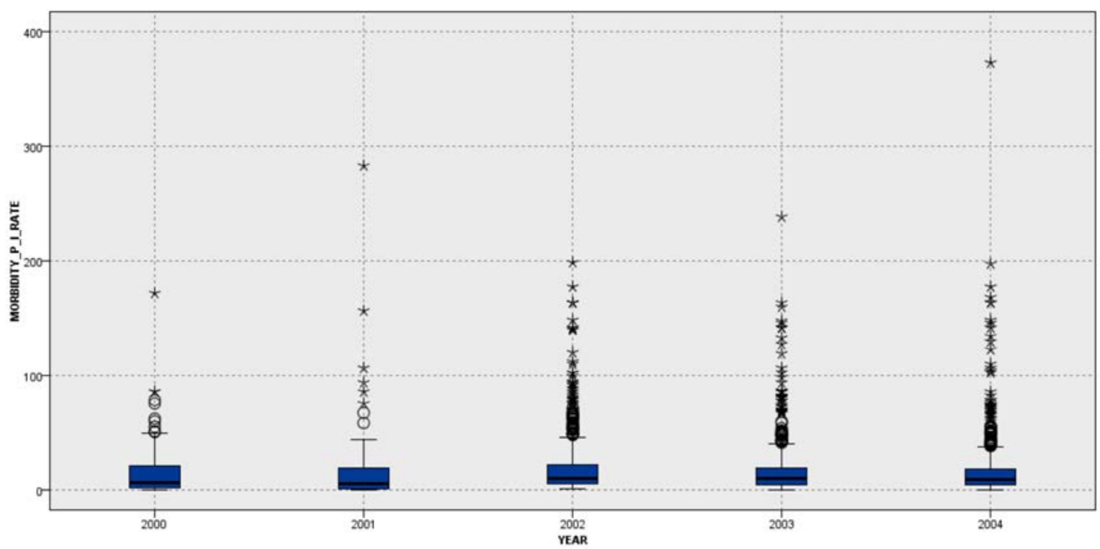


FIGURE – 16: Boxplot per month of category III, age-adjusted morbidity

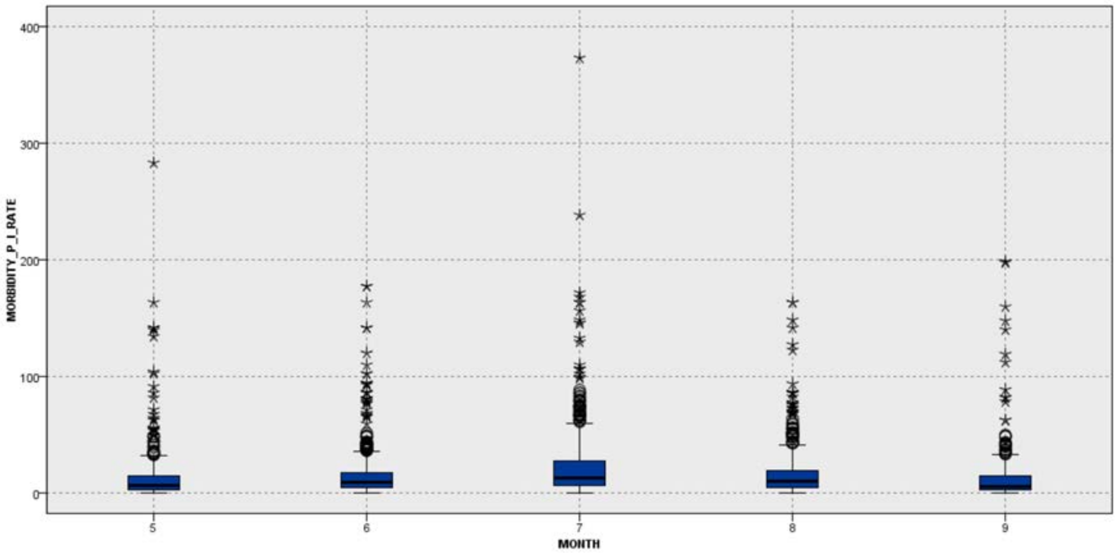


FIGURE – 17: GA Age-adjusted social economic morbidity

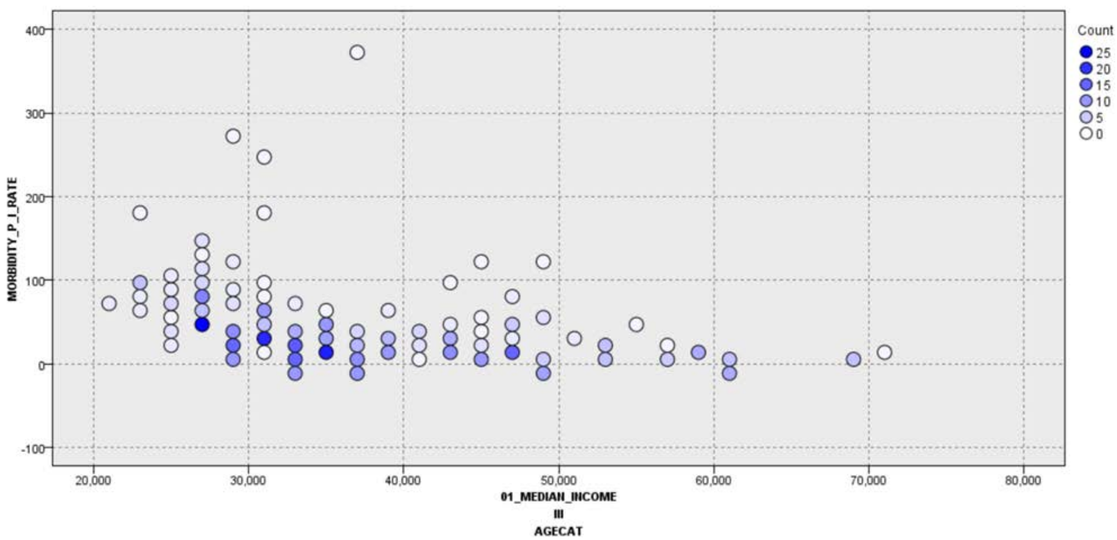
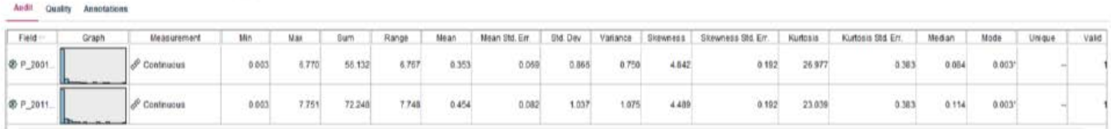


FIGURE – 18: Statistical analysis of 2001 and 2011 LCV



Audit Quality Annotations

Field	Graph	Measurement	Min	Max	Sum	Range	Mean	Mean Std. Err	Std. Dev	Variance	Skewness	Skewness Std. Err.	Kurtosis	Kurtosis Std. Err.	Median	Mode	Unique	Valid
P_2001...		Continuous	0.003	6.770	56.132	6.767	0.353	0.069	0.866	0.750	4.842	0.192	26.977	0.383	0.084	0.003'	--	1
P_2011...		Continuous	0.003	7.751	72.248	7.748	0.454	0.082	1.037	1.075	4.489	0.192	23.039	0.383	0.114	0.003'	--	1

FIGURE – 19: Statistical analysis of 2001 and 2011 LCR

Field	Graph	Measurement	Min	Max	Sum	Range	Mean	Mean Std. Err.	Std. Dev	Variance	Skewness	Skewness Std. Err.	Kurtosis	Kurtosis Std. Err.	Median	Mode	Unique	Vali
P_2001...		Continuous	34.757	96.994	13902.955	62.237	87.440	0.898	11.325	128.259	-2.595	0.192	7.379	0.383	91.719	34.757*	--	
P_2011...		Continuous	31.352	96.953	13744.180	65.600	86.441	1.014	12.792	163.637	-2.504	0.192	6.732	0.383	91.594	31.352*	--	

FIGURE – 20: Statistical analysis of LCV and LCR

Field	Graph	Measurement	Min	Max	Sum	Range	Mean	Mean Std. Err	Std. Dev	Variance	Skewness	Skewness Std. Err.	Kurtosis	Kurtosis Std. Err	Median	Mode	Unique	Valid
D_VLC		Continuous	-0.000	1.238	16.116	1.238	0.101	0.015	0.192	0.037	3.428	0.192	13.474	0.383	0.025	-0.000*	--	[illegible]
D_RLC		Continuous	-10.237	1.367	-158.776	11.605	-0.999	0.148	1.872	3.506	-2.523	0.192	6.788	0.383	-0.266	-10.237*	--	[illegible]

FIGURE – 21: Reciprocal correlation between LC vulnerability and LC resilience

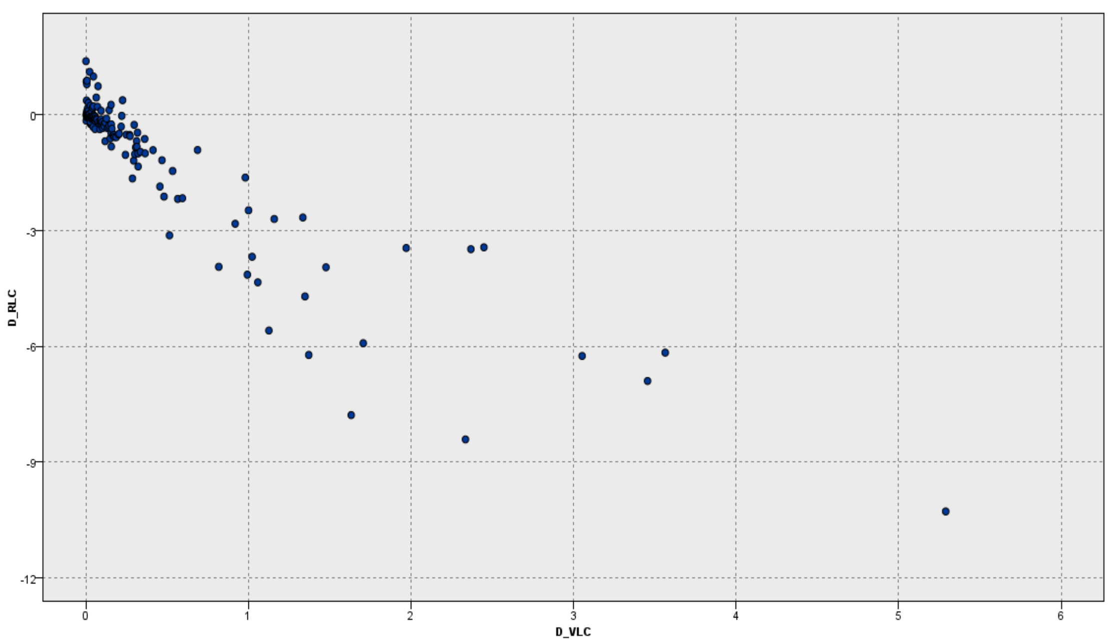


FIGURE – 22: Cumulative EHEE, land cover and morbidity per age-adjusted category

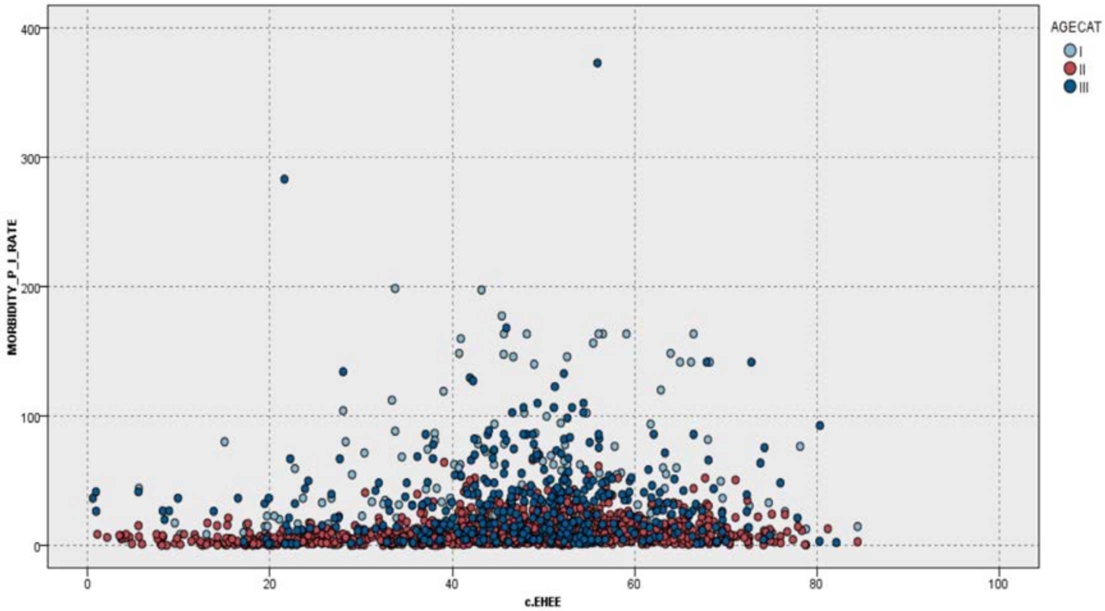


FIGURE – 23: Population adjusted morbidity, by year, period 1 (2000 – 2004)

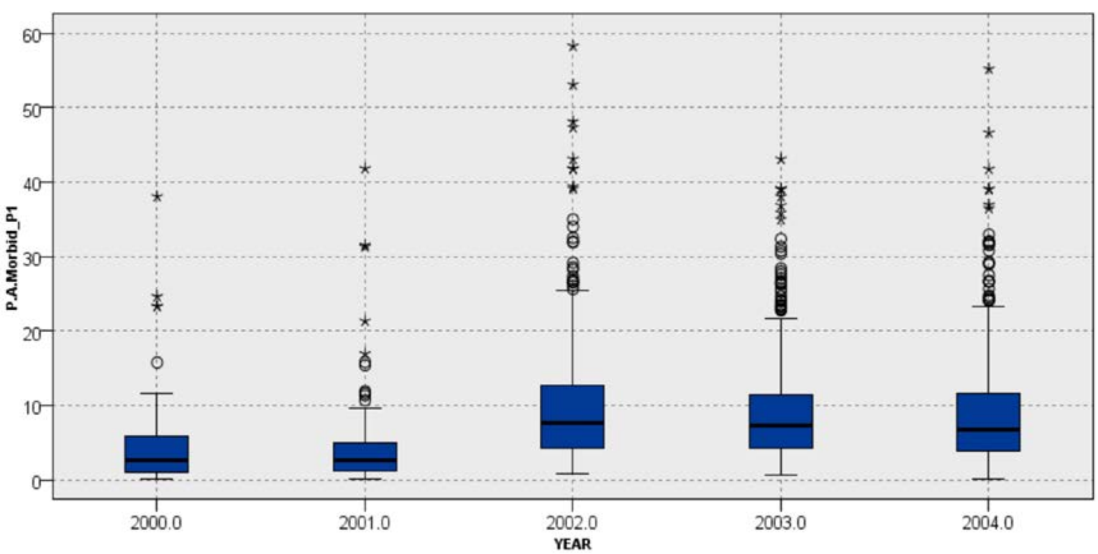


FIGURE – 24: P.A. Morbidity per month for period 1 (2000 – 2004)

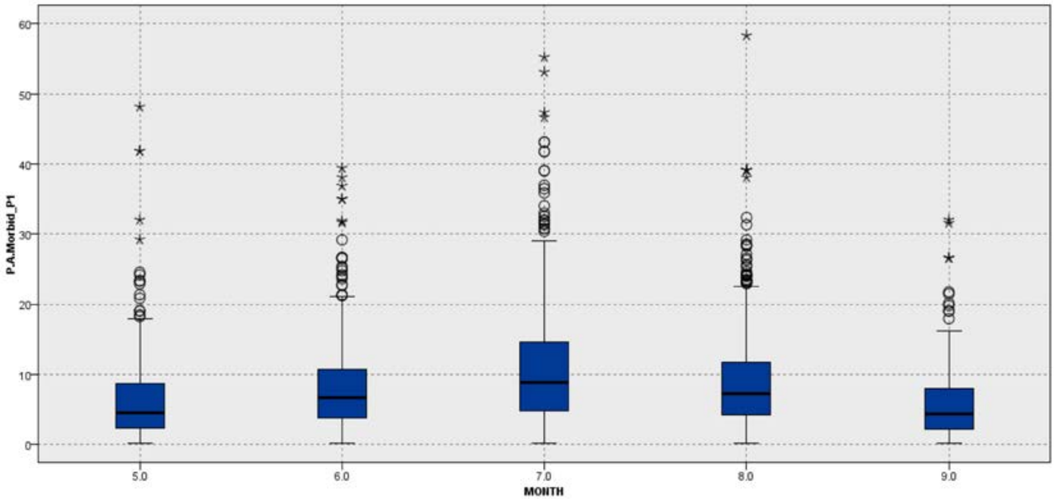


FIGURE – 25: Comparison of the two PAMs that resulted in an optimum solution for A.A. Morbidity CAT III

PREDICTED classes (EHEVI)		Classes of instances (Morb. Rate) 1	2	3	4
	1	15	5	11	3
	2	10	15	6	3
	3	5	6	12	11
	4	4	8	5	15
		CF =	12.7%	OUR=	29.1%
		FC =	12.7%	OOR=	28.4%
			OCP=	42.5%	

PREDICTED classes (EHEVI)		Classes of instances (Morb. Rate) 1	2	3	4
	1	19	6	7	2
	2	9	13	10	2
	3	5	9	9	11
	4	1	6	8	17
		CF =	8.2%	OUR=	28.4%
		FC =	9.0%	OOR=	28.4%
			OCP=	43.3%	

FIGURE – 26: Difference between actual and predicted morbidity for GA

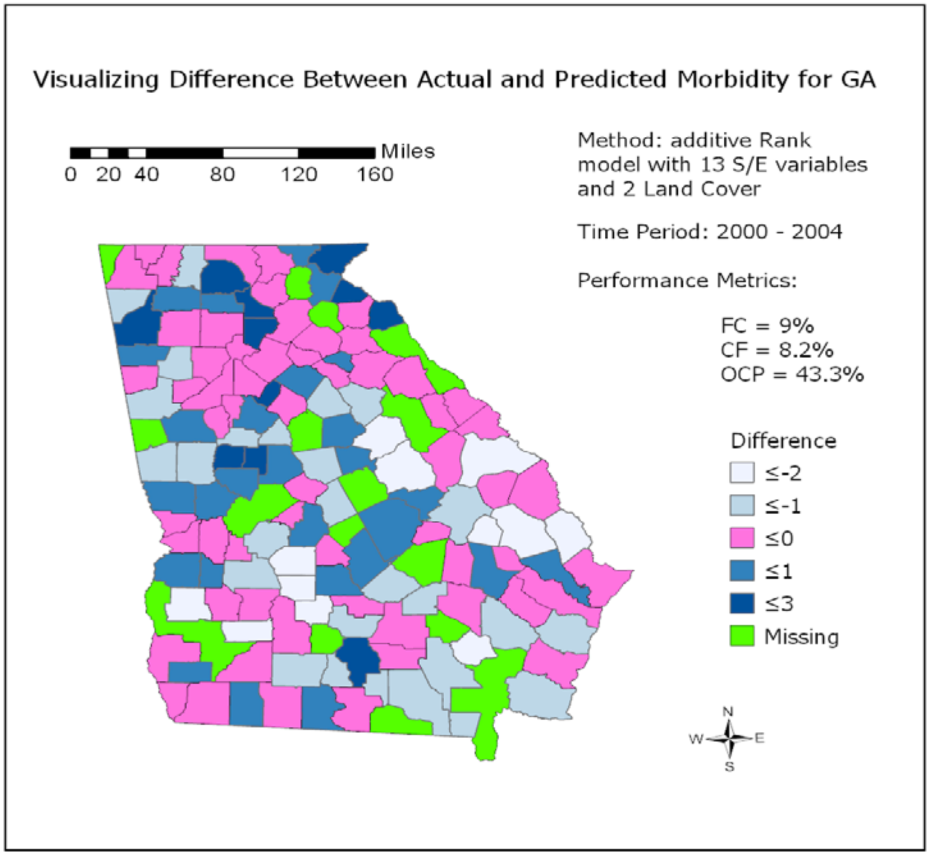


FIGURE – 28: PA matrices for Cat III morbidity classification, period-1, GA

PCA: 3 PCs 13 Var. 3 LC Varimax

PREDICTED classes (EHEVI)	Classes of instances (Morb. Rate) 1	2	3	4
1	19	6	7	2
2	9	13	10	2
3	5	9	9	11
4	1	6	8	17

CF =	8.2%	OUR=	28.4%
FC =	9.0%	OOR=	28.4%
OCP=	43.3%		

DT: 10 input variables

PREDICTED classes (EHEVI)	Classes of instances (Morb. Rate) 1	2	3	4
1	31	2	2	0
2	2	32	2	1
3	1	0	25	1
4	0	0	5	30

CF =	2.2%	OUR=	6.0%
FC =	0.7%	OOR=	6.0%
OCP=	88.1%		

FIGURE – 30: Input S/E variables in terms of importance for the 88.1% OCP DT solution

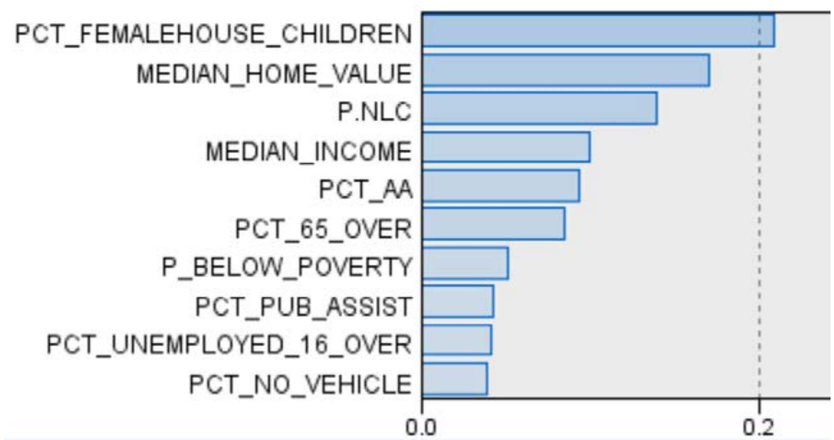


FIGURE – 31: Visualizing Difference Between Actual and Predicted Morbidity for GA (PCA)

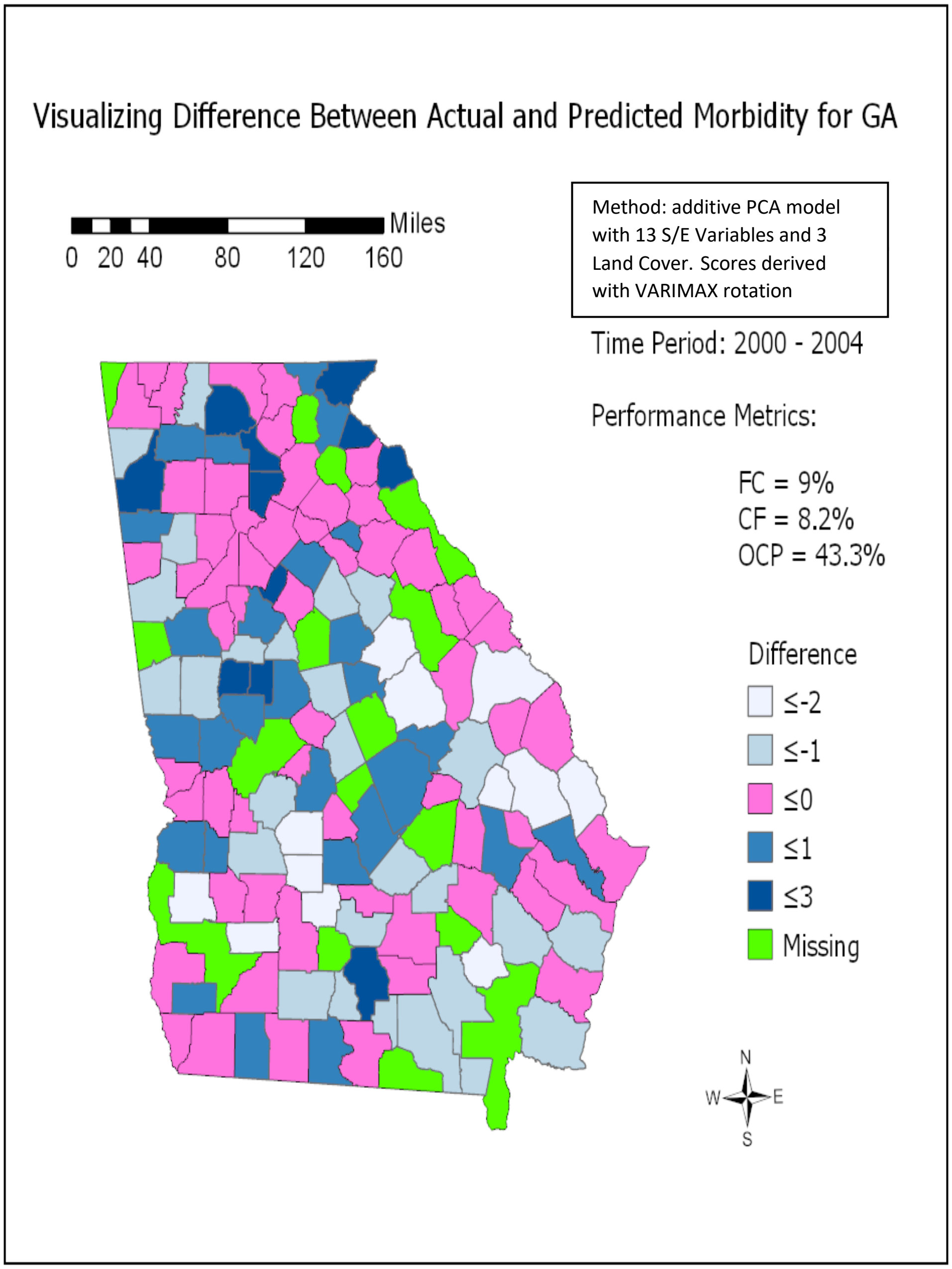


FIGURE – 32: Visualizing Difference Between Actual and Predicted Morbidity for GA (DT)

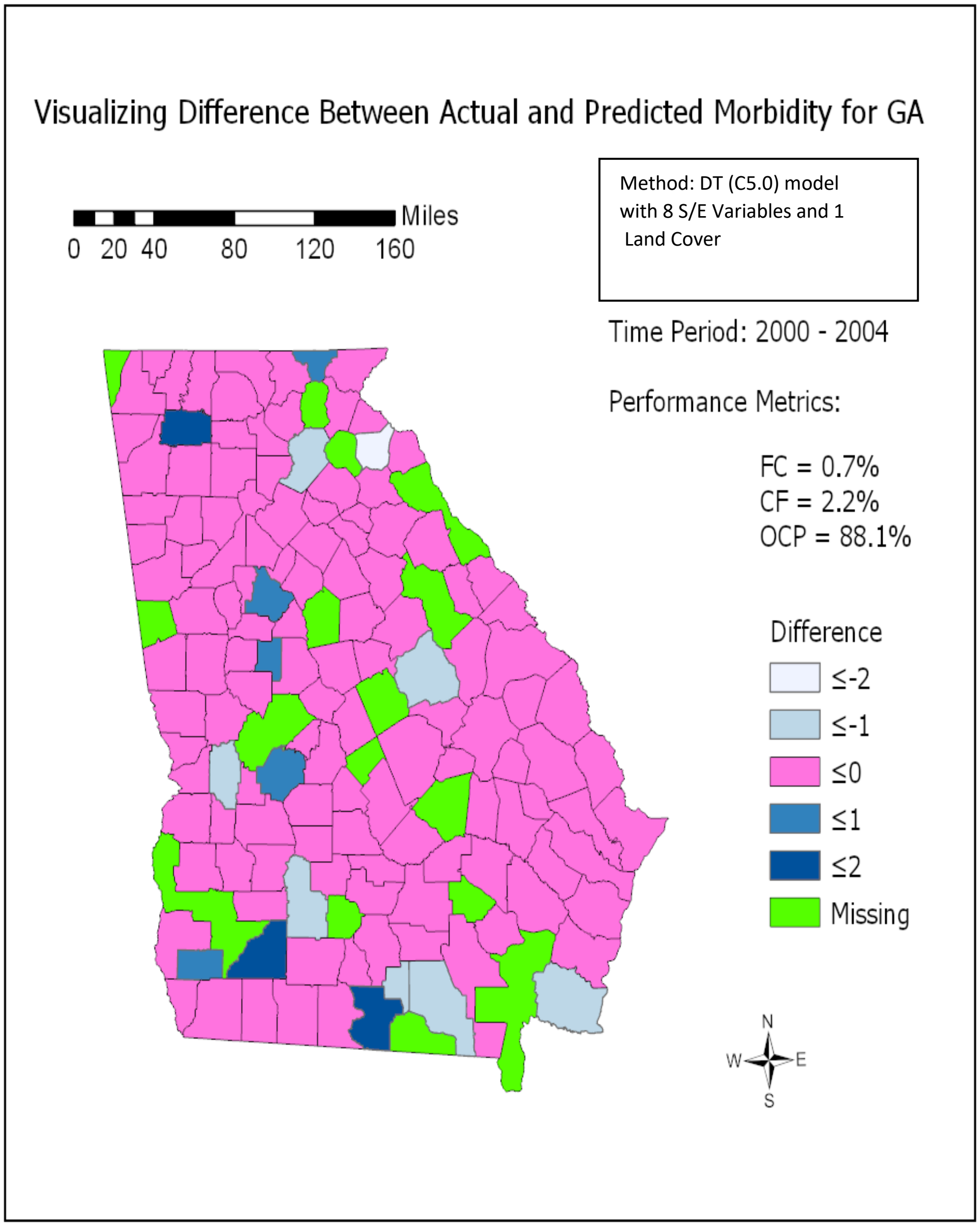


FIGURE – 33: Schematic of contextual sheres defining SoVP indices and the proposed validation approach

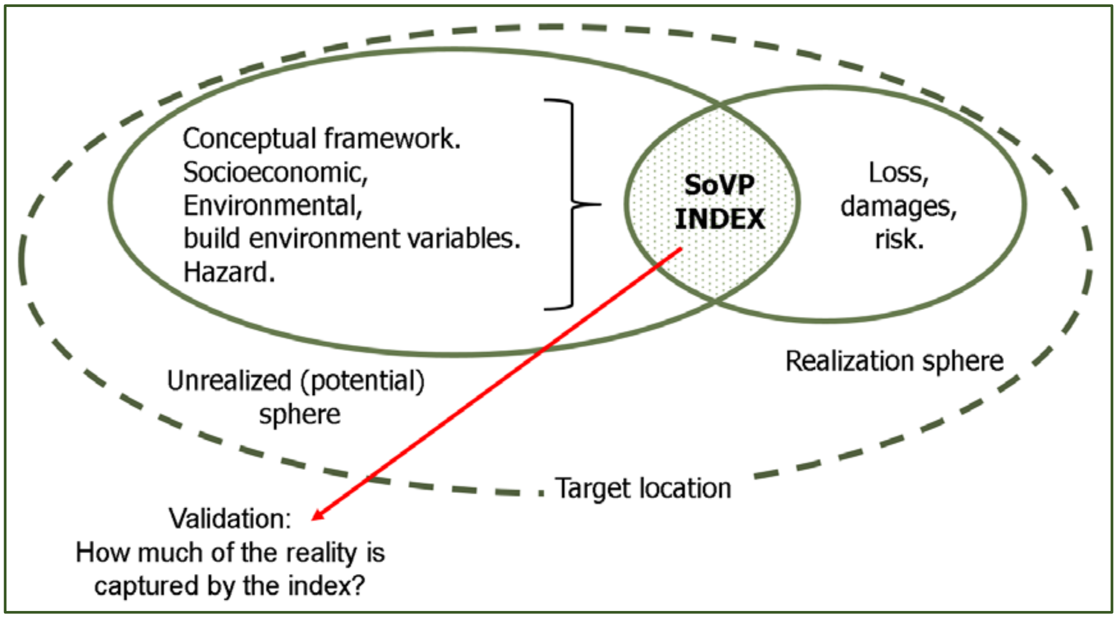


FIGURE – 34: Error or confusion matrix (or, in this study, performance assessment matrix)

			Unrealized/predicted, i				
			SoVP Categories $C(SoVP)_{ij}$				
			1	2	...	m	Σ
Realized /reference, j	Hazard Realization Categories, $C(HR)_{ij}$	1	c_{11}	c_{12}	...	c_{1m}	c_{1+}
		2	c_{21}	c_{22}			c_{2+}
		⋮					
		m	c_{m1}	...	...	c_{mm}	
		Σ	c_{+1}	c_{+2}	...		n

FIGURE – 35: The 4 × 4 PA matrix for the two disaster realizations (morbidity) used in this study (Population adjusted (left) and Cat. III morbidity

DLI.PII classes	DLI.PI Classes: 1	2	3	4
1	24	10	4	2
2	6	13	14	7
3	7	8	14	11
4	3	9	8	19

CF = 1.9%	UE = 25.8%
OCP = 44.0%	OE = 30.2%

DLI.PII classes, Cat III	DLI.PI Classes, Cat III: 1	2	3	4
1	19	4	4	4
2	6	14	7	8
3	6	11	12	6
4	3	5	10	11

CF = 2.3%	UE = 31.5%
OCP = 43.1%	OE = 25.4%

FIGURE – 36: The PA matrix for the Table-25: The 4 × 4 PA matrix for the EHEV derived with 10 input variables and 3 PCs. Adjacent is the performance threshold PA matrix

DLI.PI classes	EHEV Classes: 1	2	3	4
1	21	8	5	0
2	6	13	11	4
3	6	11	10	7
4	1	2	8	21

CF = 0.7%	UE = 25.4%
OCP = 48.5%	OE = 26.1%

DLI.PII classes, Cat III	DLI.PI Classes, Cat III: 1	2	3	4
1	19	4	4	4
2	6	14	7	8
3	6	11	12	6
4	3	5	10	11

CF = 2.3%	UE = 31.5%
OCP = 43.1%	OE = 25.4%

FIGURE – 37: Schematic of a simple decision tree

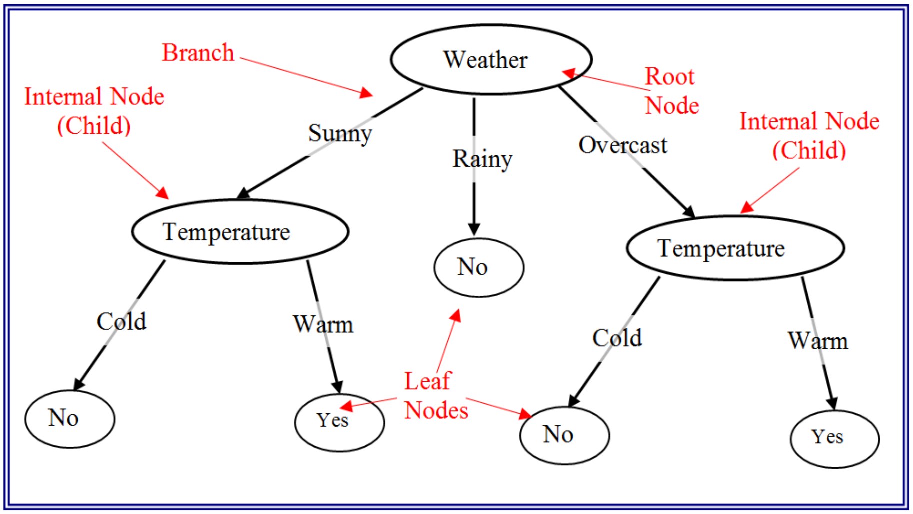


FIGURE – 38: Age adjusted morbidity (Cat III) for Periods 1 and 2

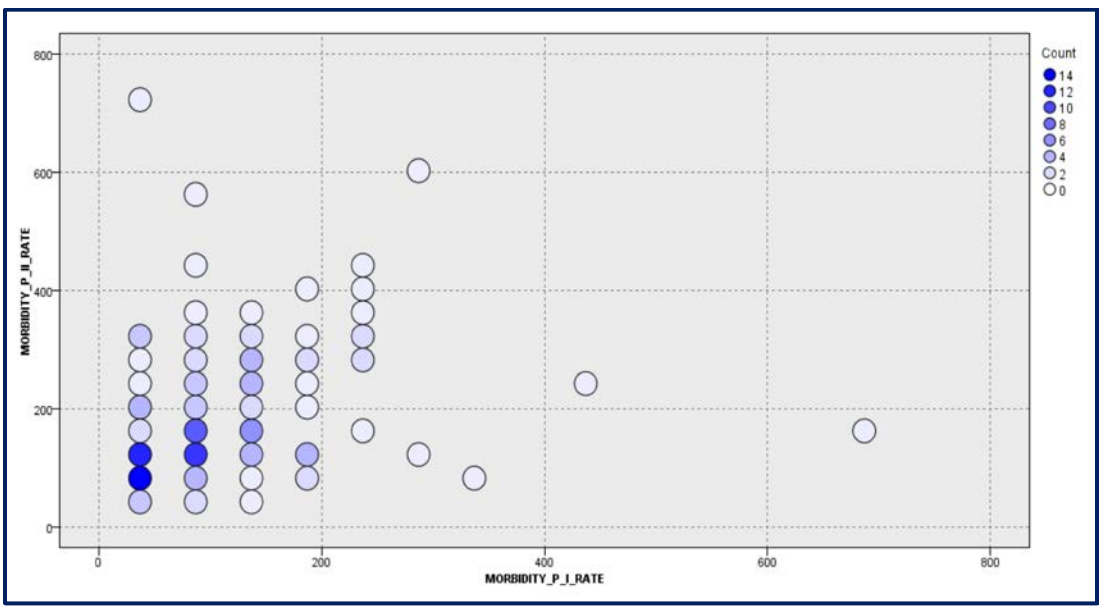


FIGURE – 39: Distribution of paired differences (red curve is the normal distribution)

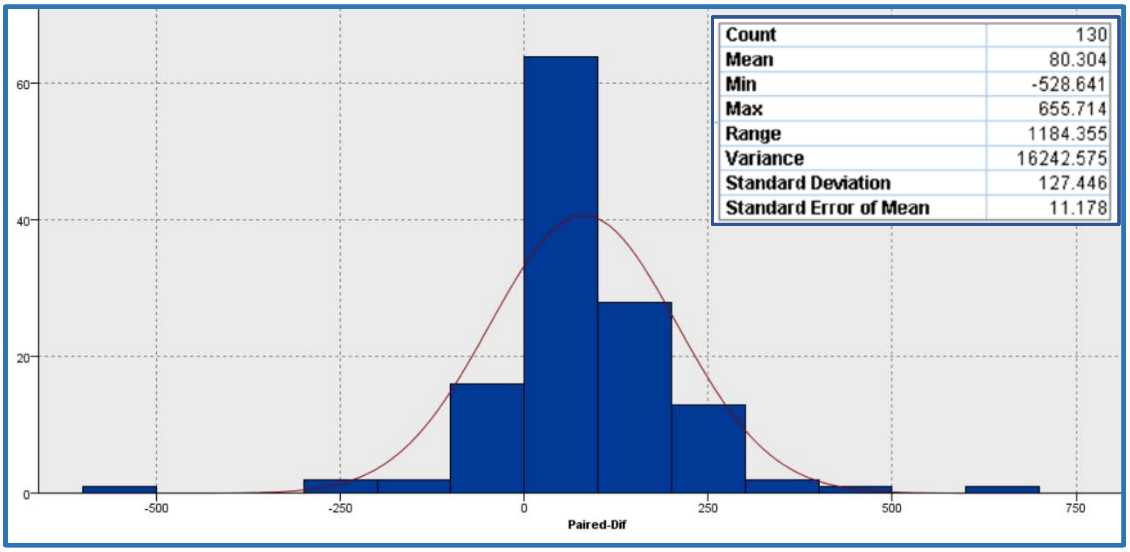


FIGURE – 40: PA matrix for validation model (P2 data) and original (P1)

Validation of DT(str. P1) model for predicting P2 Cat III Morbidity

		Classes of instances (Morb. Rate)			
		1	2	3	4
PREDICTED classes (EHEVI)	1	35	3	1	2
	2	1	28	1	3
	3	1	5	32	2
	4	1	2	4	30

CF =	4.0%	OUR=	7.9%
FC =	2.6%	OOR=	9.3%
	OCP=	82.8%	

DT Model for predicting Cat III Morbidity

		Classes of instances (Morb. Rate)			
		1	2	3	4
PREDICTED classes (EHEVI)	1	34	4	1	2
	2	0	33	2	1
	3	2	0	33	1
	4	2	1	2	33

CF =	2.6%	OUR=	7.3%
FC =	3.3%	OOR=	4.6%
	OCP=	88.1%	

FIGURE – 41: Predictor importance for DT model

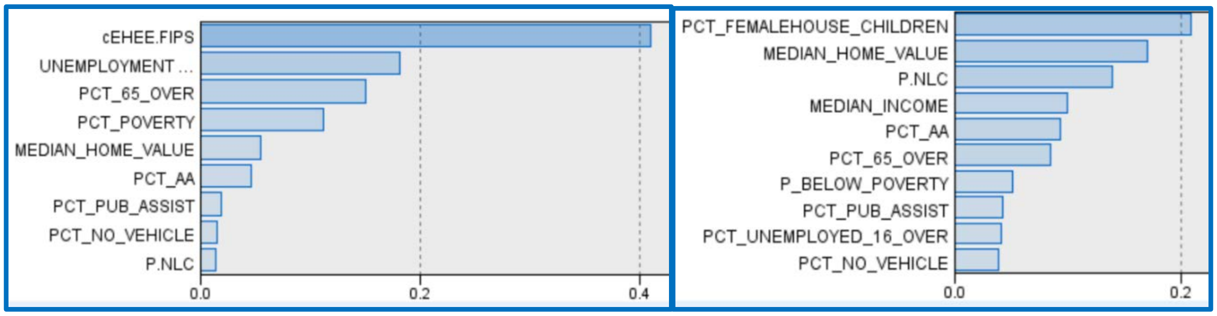


FIGURE – 42: Ten first predictor importance variables graph for deriving EHEV classification model C5.0

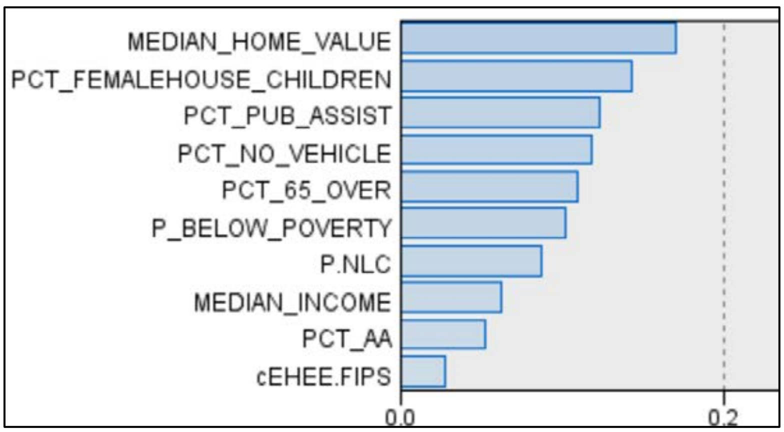


FIGURE – 43: Decision tree structure for application case (AC) 8 yielding an optimum EHEV classification model

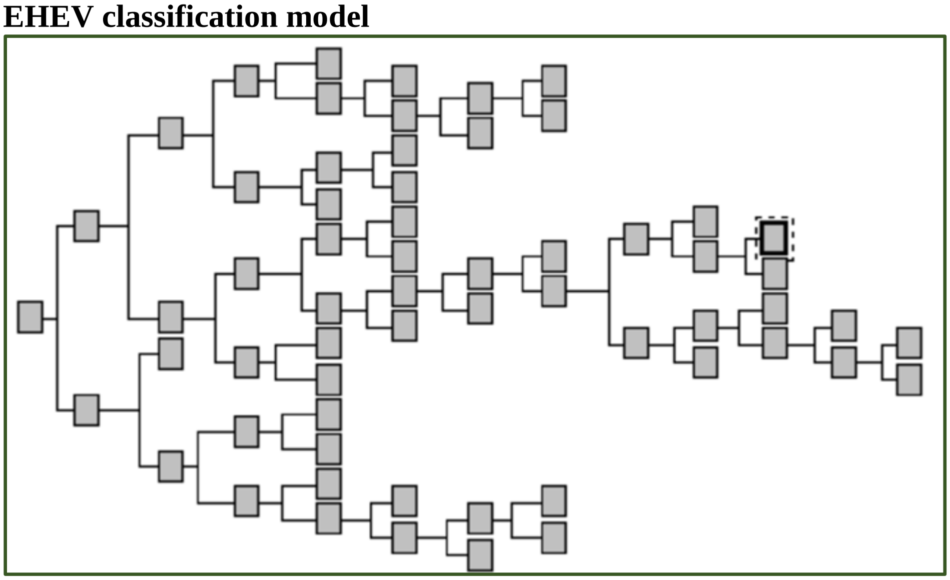


FIGURE – 44: PA matrix for application case (AC) 8 yielding an optimum EHEV classification model (CUE = 0.0%)

		EHEV DT classes			
		1	2	3	4
DLI.PI classes	1	32	2	1	0
	2	2	31	2	1
	3	0	1	26	1
	4	0	0	5	30

CF =	0.0%	UE =	6.0%
OCP =	88.8%	OE =	5.2%

FIGURE – 45: Comparative performance of DT, PCA, and PFR approaches for deriving EHEV indices; Georgia, Period I (CAT III)

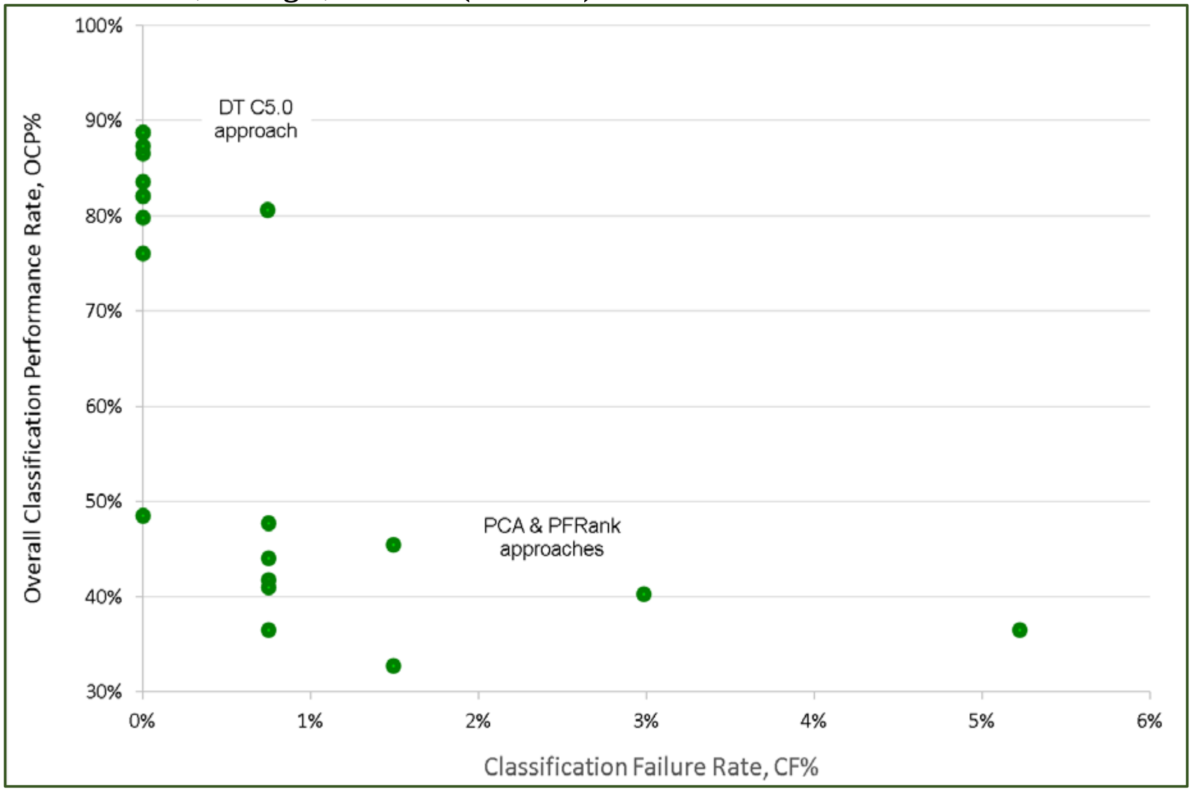


FIGURE – 46: Visualizing Difference Between Actual and Predicted Morbidity for GA (DT)

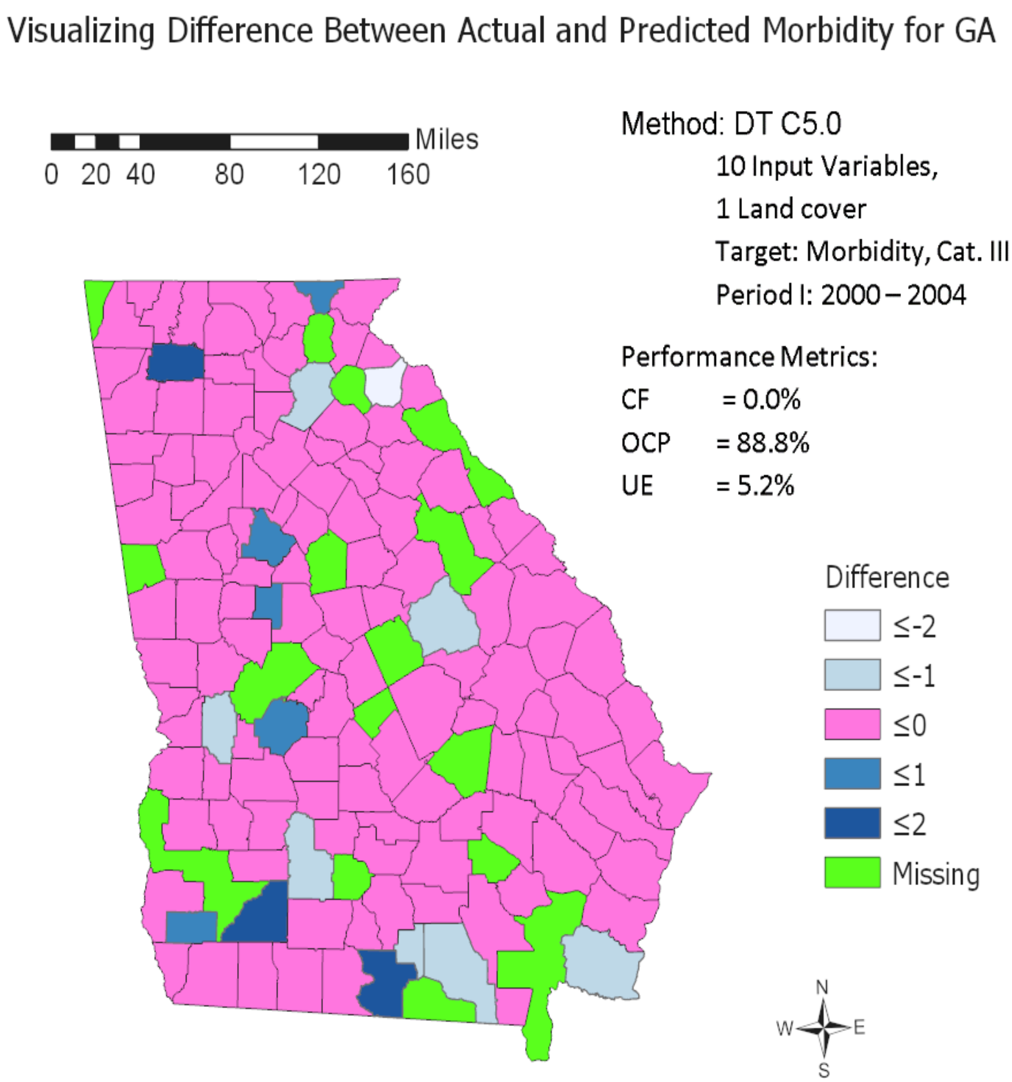
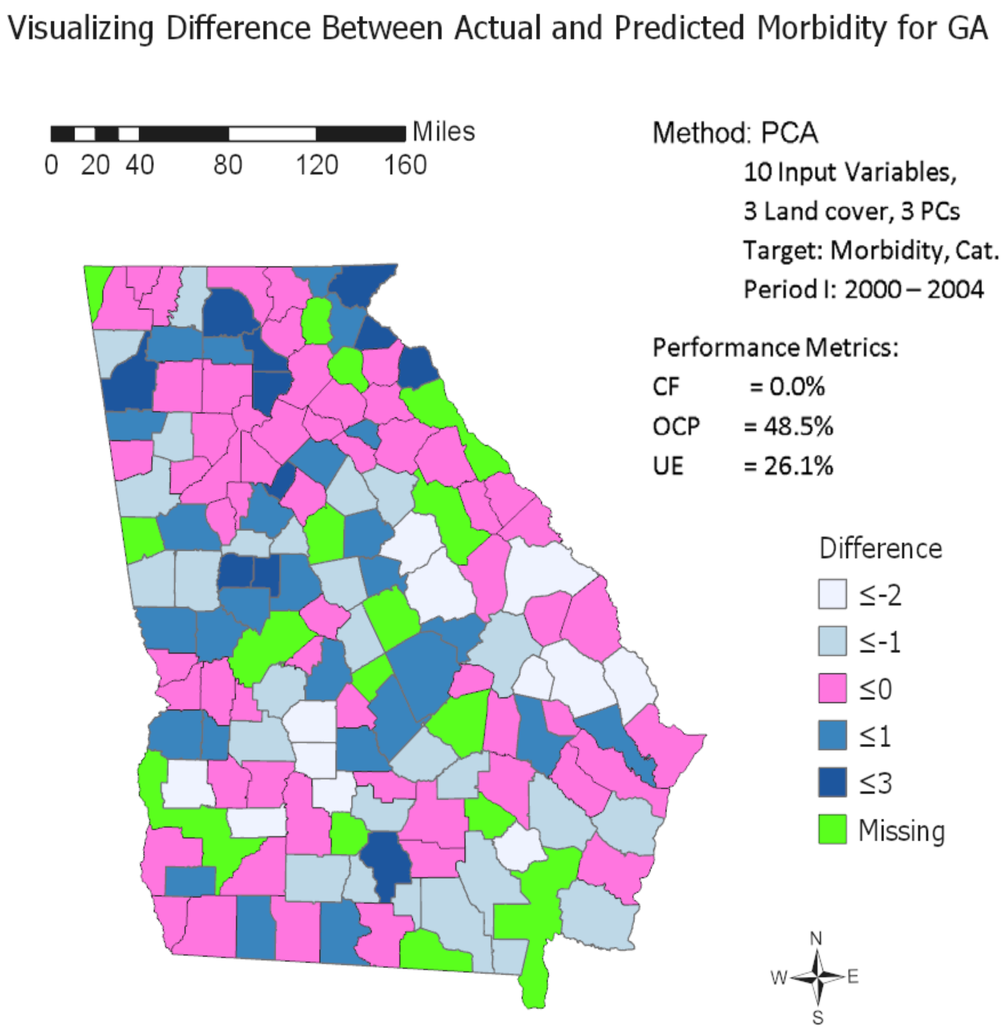


FIGURE – 47: Visualizing Difference Between Actual and Predicted Morbidity for GA (PCA)



APPENDIX B – TABLES

TABLE-I: 2010 GA DEMOGRAPHICS

2010 Demographic Population of Georgia, USA	
Demographic	**Percentage of the Population**
White American	59.7%
Black or African American (including Hispanics)	30.5%
American Indian or Alaska Native	0.5%
Asian	3.2%
Native Hawaiian or Pacific Islander	0.1%
Two or more Races	2.1%
Persons under 18 years	25.7%
Persons 65 years and over	10.7%
Total population	9,687,653

TABLE II: 2017 GA DEMOGRAPHICS

2017 Demographic Population of Georgia, USA	
Demographic	**Percentage of the Population**
White American	61.2%
Black or African American	32%
American Indian or Alaska Native	0.5%
Asian	4.1%
Native Hawaiian or Pacific Islander	0.1%
Hispanic or Latino	9.4%
Two or more Races	2.1%
Persons under 18 years	24.4%
Persons 65 years and over	13.1%
Total population	10,429,379

TABLE III: HEAT RELATED ICD-9 CODES & DESCRIPTIONS

Heat-related ER visits for 2000 – 2014 ICD-9 Codes	
ICD-9 Code	**Diagnosis**
705.1	Prickly Heat
992.0	Effects of Heat and Light
992.1	Heat syncope
992.2	Heat cramps
992.3	Heat exhaustion, anhidrotic
992.4	Heat exhaustion due to salt depletion
992.5	Heat exhaustion, unspecified
992.6	Heat fatigue, transient
992.7	Heat edema
E900.0	Injury to carotid artery, unspecified

TABLE IV: HEAT RELATED DIAGNOSIS CODES & DESCRIPTIONS

Heat-related ER visits for 2000 – 2014 ICD-9 Codes	
Diagnosis Code	**Diagnosis**
T67.0	Effects of heat and light
T67.1	Heat syncope
T67.2	Heat cramp
T67.3	Heat exhaustion, anhidrotic
X30	Exposure to excessive natural heat

TABLE V: ATLANTA METROPOLITAN WEATHER STATIONS USED TO OBTAIN APPARENT TEMPERATURE DATA

City	County	Zip Code	Local Site Name
Covington	Newton	30014	The Georgia FFA-FCCLA Center
Dallas	Paulding	30132	Paulding County High School
Dunwoody	Fulton	30075	Cherokee Town and Country Club
Griffin	Spalding	30223	Griffin Campus
Jonesboro	Clayton	30236	The Beach at Clayton County International Park
Roopville	Carroll	30217	Plant Wansley
Williamson	Pike	30292	Bledsoe Research Farm

TABLE VI: MONTHLY MEAN AVERAGE DAILY MAX HEAT INDEX

	MONTH (ADMHI mean)				
YEAR	5	6	7	8	9
2000	87.8	91.8	96.1	95.4	89.9
2001	84.8	89.3	93.9	93.3	88.2
2002	88.7	91.3	95.4	94.0	91.2
2003	88.6	90.5	92.3	93.7	87.8
2004	88.0	91.7	95.5	91.9	87.7
2005	84.6	90.1	98.2	96.0	91.0
2006	89.0	91.5	97.0	98.1	88.0
2007	85.3	92.5	94.1	102.2	90.6
2008	86.3	94.0	95.2	94.0	90.0
2009	86.3	97.1	95.3	97.7	92.0
2010	89.3	98.9	100.7	100.2	93.0
2011	91.5	98.6	99.7	101.2	89.1

TABLE VII: MONTHLY MEAN EXTREME HEAT EXPOSURE EXCEEDANCE

	MONTH (c.EHEE mean)				
YEAR	5	6	7	8	9
2000	57.5	43.3	60.8	68.0	29.0
2001	23.0	38.6	51.7	42.6	29.4
2002	38.9	55.9	49.5	63.2	40.8
2003	41.7	40.1	46.5	41.6	33.4
2004	43.5	49.9	48.6	49.6	33.3
2005	17.7	46.5	54.3	62.2	49.4
2006	40.5	54.9	51.8	47.4	48.1
2007	27.1	56.1	48.4	68.6	53.9
2008	34.4	52.3	47.9	72.7	74.3
2009	20.9	76.9	58.7	47.9	42.9
2010	40.3	66.1	68.3	56.4	52.6
2011	48.5	39.7	57.6	61.0	52.4

TABLE VIII: MEAN MORBIDIDTY RATE FROM 2000 – 2004

Mean Morbidity rate by age cat and month					
	Month				
AGECAT	5	6	7	8	9
I	39.6	45.2	45.5	38.5	43.7
II	7.3	9.5	13.0	10.7	7.0
III	36.0	29.2	39.6	30.7	26.6

TABLE IX: TOP COUNTIES IN GA WITH HIGHEST LAND COVER URBAN GROWTH FROM 2001 to 2011

County	FIPS Code	Total Difference in the Urban Sum	2010 Median Family Income	2010 Percent Living in Poverty	2010 Percent African American	2010 Percent Population White	2010 Percent Population 65 and over
Gwinnett	13135	33.6	70767	13.6	23.6	53.3	6.9
Forsyth	13117	26.5	96501	7.2	2.6	85.4	8.9
Henry	13151	24.4	70972	10.2	36.9	55	8.4
Clayton	13063	22.5	48064	22.6	66	18.9	6.6
Fulton	13121	19.8	75579	17.7	44	44.5	9.1
Cobb	13067	19.6	78920	14	25	62.2	8.7
Douglas	13097	18.6	62977	12.7	39.5	52.5	8.5
Barrow	13013	18.4	55415	12	11.4	78.8	9.3
Paulding	13223	18.1	67117	8.8	17.1	77.7	7.2
Cherokee	13057	15.2	77190	8.6	5.7	86.6	9.2
Jackson	13157	13.5	58239	15.7	6.8	86.8	11.9
Hall	13139	12.6	57774	17.8	7.4	74.1	11.1
Fayette	13113	11.7	92976	6.7	20.1	71.1	12.7
Houston	13153	11.6	67227	14	28.6	63.4	10.4
DeKalb	13089	11.31	60718	19.4	54	33.3	9
Clarke	13059	11.3	51687	33.3	26	61.9	8.5
Muscogee	13125	10.6	46283	18.5	8.2	89.8	15
Newton	13127	10.1	62445	18.8	26	67.6	15
Columbia	13073	9.1	74426	8.8	14.9	76.5	10.2

TABLE X: COUNTIES IN GA WITH ZERO OR NEGATIVE URBAN GROWTH

County	FIPS Code	Total Difference in the Urban Sum	2010 Median Family Income	2010 Percent Living in Poverty	2010 Percent African American	2010 Percent Population White	2010 Percent Population 65 and over
Clay	13061	0	31354	35.7	60.4	37.6	19.6
Talbot	13263	0	43694	22	59.2	39	16.3
Randolph	13243	0	29800	27.6	61.8	36.6	17.8
Glascock	13125	0	46283	18.5	8.2	89.8	15.1
Calhoun	13037	0	37309	36.8	61.3	34.8	11.9
Jenkins	13165	Negative	56382	28.9	40.5	54.9	15
Johnson	13167	Negative	35750	30.5	35	63.1	14
Taliaferro	13265	Negative	29375	42	59.6	37.3	20.5

TABLE XI: GA COUNTIES WITH THE HIGHEST MORBIDIDTY RATES FOR THE 65 YEARS OF AGE AND OLDER POPULATION (CAT III) WITH LOW URBAN LAND COVER

Hight morbidity (CAT III) and low ULC							
KEY_FIPS	MONTH	YEAR	c.EHEE	AGECAT	Geography	MORBIDITY _P_I_RATE	P.T.ULC
13307	5	2001	21.6	III	Webster County, Georgia	283.29	2.64
13249	7	2003	51.6	III	Schley County, Georgia	238.66	3.51
13165	7	2000	65.6	III	Jenkins County, Georgia	171.97	4.98
13183	7	2004	45.9	III	Long County, Georgia	168.35	4.77
13003	6	2002	67.9	III	Atkinson County, Georgia	141.84	5.43
13003	8	2003	34.1	III	Atkinson County, Georgia	141.84	5.43
13003	5	2003	72.8	III	Atkinson County, Georgia	141.84	5.43
13197	7	2003	52.2	III	Marion County, Georgia	132.98	2.88
13173	7	2004	41.9	III	Lanier County, Georgia	129.70	5.45
13065	8	2004	51.3	III	Clinch County, Georgia	122.85	4.85
13263	7	2001	53.1	III	Talbot County, Georgia	106.72	3.79
13263	7	2004	51.2	III	Talbot County, Georgia	106.72	3.79
13263	7	2003	47.8	III	Talbot County, Georgia	106.72	3.79
13259	6	2003	54.4	III	Stewart County, Georgia	102.78	2.13
13259	5	2004	46.6	III	Stewart County, Georgia	102.78	2.13
13019	7	2003	52.6	III	Berrien County, Georgia	98.67	4.9
13251	6	2002	80.3	III	Screven County, Georgia	92.81	4.08
13201	5	2002	51.6	III	Miller County, Georgia	91.58	4.78
13207	9	2002	44.0	III	Monroe County, Georgia	88.85	5.45
13315	7	2002	47.8	III	Wilcox County, Georgia	86.06	5.02
13315	7	2000	62.1	III	Wilcox County, Georgia	86.06	5.02
13315	6	2000	48.8	III	Wilcox County, Georgia	86.06	5.02
13165	6	2003	45.7	III	Jenkins County, Georgia	85.99	4.98
13165	8	2001	43.9	III	Jenkins County, Georgia	85.99	4.98
13165	5	2003	48.1	III	Jenkins County, Georgia	85.99	4.98
13165	8	2002	66.5	III	Jenkins County, Georgia	85.99	4.98
13165	8	2004	56.1	III	Jenkins County, Georgia	85.99	4.98
13165	8	2003	37.1	III	Jenkins County, Georgia	85.99	4.98

TABLE XII: GA COUNTIES WITH THE HIGHEST MORBIDIDTY RATES FOR THE 65 YEARS OF AGE AND OLDER POPULATION (CAT III) WITH HIGH URBAN LAND COVER

Counties with the lowest Morb. Rate (CAT III) and the highest ULC							
KEY_FIPS	MONTH	YEAR	c.EHEE	AGECAT	Geography	MORBIDITY_ P_I_RATE	P.T.ULC
13245	7	2003	52.6	III	Richmond County, Georgia	9.24	29.5
13215	6	2004	61.5	III	Muscogee County, Georgia	9.17	30.8
13215	8	2003	47.8	III	Muscogee County, Georgia	9.17	30.8
13215	8	2002	60.0	III	Muscogee County, Georgia	9.17	30.8
13215	5	2002	42.1	III	Muscogee County, Georgia	9.17	30.8
13121	7	2004	54.1	III	Fulton County, Georgia	7.25	48.3
13063	6	2004	52.8	III	Clayton County, Georgia	7.18	56.3
13063	8	2001	36.7	III	Clayton County, Georgia	7.18	56.3
13063	8	2002	66.4	III	Clayton County, Georgia	7.18	56.3
13067	6	2004	53.8	III	Cobb County, Georgia	7.14	59.6
13135	6	2002	47.9	III	Gwinnett County, Georgia	6.33	51.3
13135	8	2003	49.6	III	Gwinnett County, Georgia	6.33	51.3
13121	7	2000	53.6	III	Fulton County, Georgia	5.80	48.3
13089	8	2003	48.4	III	DeKalb County, Georgia	5.64	63.0
13089	7	2002	43.1	III	DeKalb County, Georgia	5.64	63.0
13021	5	2000	70.1	III	Bibb County, Georgia	5.10	27.8
13021	5	2002	44.4	III	Bibb County, Georgia	5.10	27.8
13021	7	2001	54.1	III	Bibb County, Georgia	5.10	27.8
13021	7	2003	53.0	III	Bibb County, Georgia	5.10	27.8
13021	8	2001	38.9	III	Bibb County, Georgia	5.10	27.8

TABLE XIII: COUNTIES WITH THE HIGHEST LAND COVER VULNERABILITY AND LOWEST LAND COVER RESILIENCE

	COUNTY	KEY_FIPS	D_VLC	D_RLC
1	Gwinnett County	13135	1.238	-10.237
2	Clayton County	13063	0.981	-6.897
3	Cobb County	13067	0.951	-6.120
4	Fulton County	13121	0.879	-6.228

TABLE XIV: LAND COVER CODE CLASS VALUE

Value	Definition
11	Open Water - All areas of open water, generally with less than 25% cover or vegetation or soil
12	Perennial Ice/Snow - All areas characterized by a perennial cover of ice and/or snow, generally greater than 25% of total cover.
21	Developed, Open Space - Includes areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses. Impervious surfaces account for less than 20 percent of total cover. These areas most commonly include large-lot single-family housing units, parks, golf courses, and vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes.
22	Developed, Low Intensity -Includes areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 20-49 percent of total cover. These areas most commonly include single-family housing units.
23	Developed, Medium Intensity - Includes areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 50-79 percent of the total cover. These areas most commonly include single-family housing units.
24	Developed High Intensity – Includes highly developed area where people reside or work in high numbers. Examples include apartment complexes, row houses and commercial/industrial. Impervious surfaces account fir 80 to 100 percent of the total cover.
31	Barren Land (Rock/Sand/Clay) - Barren areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, sand dunes, strip mines, gravel pits and other accumulations of earthen material. Generally, vegetation accounts for less than 15% of total cover.
41	Deciduous Forest – Areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. More than 75 percent of the tree species shed foliage simultaneously in response to seasonal change.
42	Evergreen Forest – Areas dominated by tree generally greater than 5 meters tall, and greater than 20% of total vegetation cover. More than 75 percent of the tree species maintain their leaves all year. Canopy is never without green foliage.

TABLE XIV: LAND COVER CODE CLASS VALUE (CONTINUED)	
Value	**Definition**
43	Mixed Forest – Areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. Neither deciduous nor evergreen species are greater than 75 percent of total tree cover.
51	Dwarf Scrub - Alaska only areas dominated by shrubs less than 20 centimeters tall with shrub canopy typically greater than 20% of total vegetation. This type is often co-associated with grasses, sedges, herbs, and non-vascular vegetation.
52	Shrub/Scrub – Areas dominated by shrubs; less than 5 meters tall with shrub canopy typically greater than 20% of total vegetation. This class includes true shrubs, young trees in an early successional stage or trees stunted from environmental conditions.
71	Grassland/Herbaceous – Areas dominated by grammanoid or herbaceous vegetation, generally greater than 80% of total vegetation. These areas are not subject to intensive management such as tilling, but can be utilized for grazing.
72	Sedge/Herbaceous - Alaska only areas dominated by sedges and forbs, generally greater than 80% of total vegetation. This type can occur with significant other grasses or other grass like plants, and includes sedge tundra, and sedge tussock tundra.
73	Lichens - Alaska only areas dominated by fruticose or foliose lichens generally greater than 80% of total vegetation.
74	Moss - Alaska only areas dominated by mosses, generally greater than 80% of total vegetation.
81	Pasture/Hay – Areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops, typically on a perennial cycle. Pasture/hay vegetation accounts for greater than 20 percent of total vegetation.

TABLE XIV: LAND COVER CODE CLASS VALUE (CONTINUED)	
Value	**Definition**
82	Cultivated Crops – Areas used for the production of annual crops, such as corn, soybeans, vegetables, tobacco, and cotton, and also perennial woody crops such as orchards and vineyards. Crop vegetation accounts for greater than 20 percent of total vegetation. This class also includes all land being actively tilled.
90	Woody Wetlands – Areas where forest or shrub land vegetation accounts for greater than 20 percent of vegetation cover and the soil or substitute is periodically saturated with or covered with water.
95	Emergent Herbaceous Wetlands – Areas where perennial herbaceous vegetation accounts for greater than 80 percent of vegetative cover and the soil or substrate is periodically saturated with or covered with water.

TABLE XV: METHODOLOGIES USED TO DETERMINE EXTREME HEAT SOCIAL VULNERABILITY

Title	Author	Purpose	Methodology	Verification Results	Was Minority Heat Social Vulnerability (HSV) discussed	Predictive Performance Verification
Social Vulnerability to Environmental Hazards	Susan L. Cutter; Bryan J. Boruff & W. Lynn Shirley (2003)	Analyze county-level SES and demographic data to construct a Social Vulnerability Index (SoVI)	Factor analytical approach & additive model approach	Principle Component Analysis (PCA) Correlation between presidential disaster declarations by county and the SoVI score	Yes – race was considered as a contributor to social vulnerability	No. The article states: the next step is to examine the overall vulnerability as measured by the SoVi has changed over time and space. (Cutter et. al. 2003)
Assessing the Performance of a Vulnerability Index during Oppressive Heat across Georgia, United States	Maier, Grundstein; Jang, Li; Naeher & Shepherd (2003)	To broaden the geographic context of earlier work and compute heat vulnerability across the state of Georgia	Modified Heat Vulnerability Index developed by Reid is used to characterize vulnerability by county	Prevalence Principle Component Analysis (PCA) Apparent Temperature (AT) - The warm season 95th percentile max AT was used as a threshold Mortality data was compared to meteorological data (Poisson mixed effect model with natural splines)	Yes- urban and rural areas were examined as well as nonwhite residents (looked at land use/land cover – Natural resource spatial analysis laboratory)	No. The article states: the modified HVI can be tested in other regions of the country such as the Midwest, which has not yet been empirically examined. (Grunstein et. al. 2003)

TABLE XV: METHODOLOGIES USED TO DETERMINE EXTREME HEAT SOCIAL VULNERABILITY (CONTINUED)

Title	Author	Purpose	Methodology	Verification Results	Was Minority Heat Social Vulnerability (HSV) discussed	Predictive Performance Verification
Climate change vulnerability assessment in Georgia	Binta, KC; Shepherd, Marshall & Gaither, Cassandra (2015)	This study focuses on climate change in Georgia, considering both biophysical and socio-demographic indicators of vulnerability.	Compared geographic vulnerability, historical climate data, climate change vulnerability, socioeconomic vulnerability and social vulnerability	Anomalies in decadal temperature were compared to the 30-yr climate normal. Social vulnerability was verified via reviewing other literature.	Yes – The Black Belt of the South, Hispanics and Asian Immigrants	No. The article states: Attribution studies are emerging as a challenging new field of study and beyond our scope. (Binta et. al. 2015)
Mapping Community Determinants of Heat Vulnerability	Reid, Colleen; O'Neill, Marie; Gronlund, Carina; Brines, Shannon; Brown, Daniel; Diez-Roux & Schwartz, Joel (2009)	To create a cumulative heat vulnerability index for nationwide comparison.	Mapped 10 vulnerability factors for heat -related morbidity/mortality in the US in geographic space and identify potential areas for intervention and further research.	Performed a factor analysis Spearman's correlation coefficients were calculated between the 10 vulnerability variables A varimax rotation was used to minimize the number of the original variables Factor analysis was performed in SAS	Yes – they combined education, poverty, race and green space in urban areas	No. The article states: This study can be considered a first step toward tools that can help public health professionals prepare climate change adaptation plans for their communities. (Reid et. al. 2009)

TABLE XV: METHODOLOGIES USED TO DETERMINE EXTREME HEAT SOCIAL VULNERABILITY (CONTINUED)

Title	Author	Purpose	Methodology	Verification Results	Was Minority Heat Social Vulnerability (HSV) discussed	Predictive Performance Verification
Urban Form and Extreme Heat Events: Are Sprawling Cities More Vulnerable to Climate Change Than Compact Cities?	Stone, Brian; Hess, Jeremy & Frumkin, Howard (2010)	Examine the association between urban form at the level of the metropolitan region and the frequency of EHEs over five-decade period	Reviewed previously published journals to measure the association between urban form in 2000 and the mean annual rate of change in EHEs between 1956 and 2005	Based on previous index analyses. Performed a t-test to gauge the statistical significance of a linear association.	No	No
Developing an applied extreme heat vulnerability index utilizing socioeconomic and environmental data	Johnson, Daniel; Stanforth, Austin; Lulla, Vijay and Luber, George (2012)	Create an extreme heat vulnerability index that could be utilized by city officials to assist in the migration of extreme heat events.	Combine 25 well-known indicators of extreme heat-health risk into an applied index utilizing a PCA	Principle component analysis A varimax rotation	Yes – Black population is included, but there is more focus on the socioeconomic states	No
Warm season temperatures and emergency department visits in Atlanta, GA	Winquist, Andrea; Grundstein, Andrew; Chang, Howard; Hess, Jeremy & Sarnat, Stefanie (2016)	Assess the association between warm-season ambient temperature and emergency department visits in Atlanta from 1993-2012	Examined daily counts of ED visits with primary diagnosis of heat illness, by age group in relation to daily TMX	Poisson time series models. Estimated the relative risk and 95% CI for TMX changes and conducted sensitivity analyses.	No	No

TABLE XV: METHODOLOGIES USED TO DETERMINE EXTREME HEAT SOCIAL VULNERABILITY (CONTINUED)						
Title	Author	Purpose	Methodology	Verification Results	Was Minority Heat Social Vulnerability (HSV) discussed	Predictive Performance Verification
Vulnerability to extreme heat and climate change: is ethnicity a factor?	Hansen, Alana; Bi, Linda; Saniotis, Arthur & Nitschke, Monika (2013)	To investigate the reasons why ethnicity may be associated with susceptibility to extreme heat in Australia	Not really mentioned The article states “literary evidence was sourced from relevant peer-reviewed and grey literature.”	Verification not mentioned	Yes	No
A social vulnerability index for disaster management	Flanagan, Barry E.; Edward, Gregory W.; Hallisey, Elaine J.; Haitgert, Janet L.; Lewis, Brian (2011)	This paper, therefore, addresses an important subcomponent of the disaster management risk equation—social vulnerability—with the goal of improving all phases of the disaster cycle.	An overall PR for each tract was calculated as the sum of the domain PRs SVI values and flag counts were calculated for each of the 15 variables, for the four domains, and for the overall results.	Yes. Verified by applying the methodology to previous extreme weather events (cases studies)	Yes. Percent minority was considered	Yes – A unique toolkit consisting of SVI data along with a simple mapping application was initially distributed to 24 state and local public health departments for review and feedback. (Flanagan et. al. 2011)

TABLE XVI: COMONLY USED EXTREME HEAT SOCIAL VULNERABILITY INDEX FACTORS

Flanagan et. al. 2011	Reid et al. 2009	Maier et.al. 2014 (Based on Reid)	Binta et. al. 2015	Bahksh 2015	Cutter et. al. (2003)
Socioeconomic Status: Income (per capita income) Poverty (% below poverty) Employment (% civilian unemployed) Education (%w/o high school diploma)	Percent Below Poverty Line Percent w/less than a high school diploma	Percent Below Poverty Line Percent w/less than a high school diploma	Poverty Occupation Unemployment	Low education Unemployed	Personal Wealth: Per capita income Median house values / median rents % of households earning < $75K per year
Household Composition/Disability • Age (dependents <18 yrs old, 65 and older) • Single Parenting (% male/female w/o spouse & children under 18) • Disability (more than 5 yrs old w/disability)	Percent Population Living Alone Percent Population ≥ 65 years of age Percent Population ≥ 65 living alone	Percent Population Living Alone Percent Population ≥ 65 years of age Percent Population ≥ 65 living alone	Age group > 65 Age Group < 5 Female Head of Household	Age group > 65 Age Group < 5 Disabled over 65 Living alone	Age Children in the community % of the population > 65
Minority Status/Language • Race (%minority) • Ethnicity • English-language proficiency	Percent population other than white	Percent population other than white	Racial and ethnic minorities Non-English speaking	Race other than White Low English skills	Race, specifically African American
Housing/Transportation • Housing Structure (quality/ high-rise) • % multi-unit • % mobile homes • Crowding • Vehicle Access • (% of persons in group quarters: inmates, nursing homes, dormitory)			Urban/Rural population Inmate population Renter population Dwelling in mobile homes	High Population Density Lack of access to transit/car Housing Quality Top floor of high-rise High Crime Area Outdoor Workers/athletes	Density of the Built Environment Housing Stock and Tenancy Occupation
				Heat Exposure	
				Urban Heat Islands	

TABLE XVI: COMONLY USED EXTREME HEAT SOCIAL VULNERABILITY INDEX FACTORS (CONTINUED)

Flanagan et. al. 2011	Reid et al. 2009	Maier et.al. 2014 (Based on Reid)	Binta et. al. 2015	Bahksh 2015	Cutter et. al. (2003)
Environmental Variables					
None Looked at previous extreme weather events (i.e. Hurricane Katrina)	Calculated the mean apparent temperature for MSAs from 1985 to 2003 Land cover - % census tract area not covered in vegetation (NLCD 2001) Diabetes prevalence - % pop Air conditioning % household w/o central AC % household w/o any AC	Apparent Temperature (AT) - The warm season 95th percentile max AT was used as a threshold Land cover - % county with land use/ cover described as urban (NRSAL 2011) Diabetes prevalence - % pop	Anomalies in decadal temperature were compared to the 30-yr climate normal. The mean temperature was used	Daily maximum temperature (Tmax) was used from the National Climate Data Center Summer season percentage of days when Tmax was above 35°C Summer season percentage of the days where Tmax was above the 95th percentile temperature for each station. A monthly percentile approach where the percentage of days above the 95th percentile was calculated for each month per station	Infrastructure development Single-sector Economic Dependence
				Land Cover No AC	

TABLE XVII: SAMPLE OF PREVIOUS STUDIES AND CLASSIFICATION OF SOV INDICES

Author	**Derivation Methodology**	**Map Number of Vulnerability Classes (and Type)**
Cutter et al. (2003)	Principal Component Analysis	Five (Standard Deviation)
Chakraborty et al. (2005)	Maximum value transformation (ratio of value)	Five (Defined interval)
Flanagan et al. (2011)	Percentile Rank	Three (Equal Interval)
Ryder et al. (2006)	Principal Component Analysis	Five (Standard Deviation)
Schmidtlein et al. (2008)	Principal Component Analysis	Five (Standard Deviation)
Yoon (2012)	Percentile Rank, rescaling, and Principal Component Analysis	Five (Standard Deviation)
Zhou et al. (2014)	Principal Component Analysis	Four (Standard Deviation)

TABLE XVIII: 2000 AND 2010 GEORGIA, USA CENSUS VARIABLES USED

Category	**Census Variables**
Housing Characteristics	% Occupied Housing Units: Renter Occupied, % Housing Units: 5 or more, % Housing Units: Mobile home or trailer, etc.
Children	% Under 5 years
Elderly	% 65 years and over
Race – African American	% African American
Race – White	% White
Female Head Household with children (under18 years old)	% Households with one or more people under 18 years: Female householder, no husband present
Institutionalized Persons	% Population in group quarters: Institutionalized Population
Education - Less than High School Degree	% Population 25 years and over: Less Than High School
Unemployed	% Population in Labor Force 16 Years and Over: Unemployed
Household Income	Median household income in 1979/1999 Dollars
Below Poverty	% Below Poverty
Mobility	% No Vehicle Available
Social Welfare Recipient	% social welfare
Housing Value	Median value
Occupation Type	Bottom quantile occupation type (e.g. Manual material moving, etc.)

TABLE XIX: P.A. MORBIDIDTY PER MONTH FOR PERIOD 1 (2000 – 2004)

	c.EHEE	MEDIAN_INCOME	PCT_NO_VEHICLE	PCT_PUB_ASSIST	PCT_FEMALEHOUSE_CHILDREN	PCT_5_UNDER	PCT_65_OVER	PCT_AA	PCT_NO_HSEDU	PCT_BELOW_POVERTY	PCT_UNEMPLOYED_16_OVER	PCT_MOB_HOME	PCT_RACE_OTHER_THAN_WHITE	MEDIAN_HOME_VALUE	P.MLC	P.NLC	P.T.ULC	P.A.Morbid_P1
c.EHEE	1	-0.219	0.223	0.266	0.259	0.034	0.014	0.266	0.064	0.264	0.222	0.192	0.245	-0.29	-0.036	0.186	-0.187	0.201
MEDIAN_INCOME	-0.219	1	-0.692	-0.715	-0.497	0.256	-0.647	-0.444	-0.571	-0.867	-0.56	-0.601	-0.392	0.88	0.404	-0.56	0.553	-0.403
PCT_NO_VEHICLE	0.223	-0.692	1	0.819	0.809	-0.138	0.466	0.785	0.345	0.801	0.724	0.223	0.748	-0.569	-0.166	0.203	-0.199	0.251
PCT_PUB_ASSIST	0.266	-0.715	0.819	1	0.733	-0.002	0.41	0.719	0.298	0.825	0.667	0.364	0.677	-0.674	-0.291	0.347	-0.341	0.326
PCT_FEMALEHOUSE_CHILDREN	0.259	-0.497	0.809	0.733	1	0.025	0.12	0.921	0.011	0.691	0.747	-0.024	0.915	-0.416	-0.068	-0.087	0.091	0.072
PCT_5_UNDER	0.034	0.256	-0.138	-0.002	0.025	1	-0.568	-0.067	-0.088	-0.161	-0.103	-0.088	0.002	0.097	0.007	-0.146	0.148	-0.094
PCT_65_OVER	0.014	-0.647	0.466	0.41	0.12	-0.568	1	0.162	0.473	0.463	0.208	0.337	0.075	-0.519	-0.223	0.486	-0.484	0.315
PCT_AA	0.266	-0.444	0.785	0.719	0.921	-0.067	0.162	1	0.014	0.632	0.726	0.03	0.983	-0.395	0.066	-0.053	0.052	0.16
PCT_NO_HSEDU	0.064	-0.571	0.345	0.298	0.011	-0.088	0.473	0.014	1	0.366	0.127	0.477	-0.018	-0.616	-0.28	0.487	-0.483	0.32
PCT_BELOW_POVERTY	0.264	-0.867	0.801	0.825	0.691	-0.161	0.463	0.632	0.366	1	0.707	0.48	0.602	-0.713	-0.373	0.382	-0.373	0.37
PCT_UNEMPLOYED_16_OVER	0.222	-0.56	0.724	0.667	0.747	-0.103	0.208	0.726	0.127	0.707	1	0.138	0.715	-0.424	-0.145	0.082	-0.078	0.19
PCT_MOB_HOME	0.192	-0.601	0.223	0.364	-0.024	-0.088	0.337	0.03	0.477	0.48	0.138	1	-0.026	-0.625	-0.319	0.672	-0.669	0.47
PCT_RACE_OTHER_THAN_WHITE	0.245	-0.392	0.748	0.677	0.915	0.002	0.075	0.983	-0.018	0.602	0.715	-0.026	1	-0.334	0.087	-0.164	0.163	0.125
MEDIAN_HOME_VALUE	-0.29	0.88	-0.569	-0.674	-0.416	0.097	-0.519	-0.395	-0.616	-0.713	-0.424	-0.625	-0.334	1	0.377	-0.636	0.63	-0.436
P.MLC	-0.036	0.404	-0.166	-0.291	-0.068	0.007	-0.223	0.066	-0.28	-0.373	-0.145	-0.319	0.087	0.377	1	-0.367	0.335	-0.166
P.NLC	0.186	-0.56	0.203	0.347	-0.087	-0.146	0.486	-0.053	0.487	0.382	0.082	0.672	-0.164	-0.636	-0.367	1	-0.999	0.385
P.T.ULC	-0.187	0.553	-0.199	-0.341	0.091	0.148	-0.484	0.052	-0.483	-0.373	-0.078	-0.669	0.163	0.63	0.335	-0.999	1	-0.384
P.A.Morbid_P1	0.201	-0.403	0.251	0.326	0.072	-0.094	0.315	0.16	0.32	0.37	0.19	0.47	0.125	-0.436	-0.166	0.385	-0.384	1

TABLE XX: SUMMARY PA METRICS; PERIOD 1 P.A. MORB; OPTIMIZATION APPROACH FOR RANKS AND PCA

Period 1: P.A. Morbidity			FC	CF	OCP	OOR	OUR	Cum
Add. Ranks	13 Var	NoLC	11.3%	10.1%	36.5%	29.6%	34.0%	n/a
Add. Ranks	13 Var	3 LC	10.7%	9.4%	35.2%	30.2%	34.6%	n/a
Add. Ranks	13 Var	2 LC	10.7%	9.4%	36.5%	29.6%	34.0%	n/a
PCA 3 PC +++	13 Var	No LC	11.3%	11.3%	39.6%	29.6%	30.8%	79.7%
PCA 3 VARIM	13 Var	No LC	12.6%	13.2%	35.8%	30.2%	34.0%	79.7%
PCA 3 PC +++	13 Var	3 LC	14.5%	12.6%	34.6%	31.4%	34.0%	74.5%
PCA 3 PC +++	13 Var	2 LC	14.5%	11.9%	35.8%	30.2%	34.0%	78.5%
PCA 2 PC +-	13 Var	No LC	13.2%	13.8%	30.8%	32.7%	36.5%	68.5%
PCA 2 PC +-	13 Var	3 LC	13.8%	12.6%	38.4%	28.3%	33.3%	63.9%
PCA 2 PC VARIM	13 Var	3 LC	10.7%	10.1%	34.6%	30.2%	35.2%	63.9%
PCA 3pc +-+	13 Var	2 LC	13.8%	10.1%	34.6%	28.9%	36.5%	78.5%
PCA 3pc VARIM	13 Var	2 LC	11.9%	8.8%	40.9%	27.0%	32.1%	78.5%

TABLE XXI: TABLE REMOVED

TABLE XXII: DT (5.0 ALGORITHM) MODELING OF PERFORMANCE ASSESSMENT MORBIDITY

Period 1 P.A. Morb.			FC	CF	OCP	OOR	OUR
DT C5.0 4 Bin	13 Var	No LC	10.1%	1.9%	77.4%	16.4%	6.3%
DT C5.0 4 Bin	13 Var	3 LC	6.3%	3.8%	74.8%	10.1%	15.1%
DT C5.0 4 Bin	10 Var	2 LC	2.5%	1.9%	83.6%	5.0%	11.3%
DT C5.0 4 Bin	9 Var	2LC	7.5%	2.5%	80.5%	11.3%	8.2%
DT C5.0 4 Bin	10 Var	1LC	4.4%	1.9%	83.0%	10.7%	6.3%
DT C5.0 4 Bin	7 Var	1LC	4.4%	1.9%	84.3%	12.6%	3.1%

TABLE XXIII: VALIDATION METRICS FOR PERCENTAGE FRACTONAL RANKS DERIVAITON APPROACH, GEORGIA

	Method (for population	**Var**	**LC**	**OCP**	**UE**	**CF**
PT	DLI.PI → DLI.PII	1	n/a	44.0%	25.8%	1.9%
R1	EHEV(PFR) → DLI.PI	13	no	36.5%	34.0%	0.6%
R2	EHEV(PFR) → DLI.PI	13	3	33.3%	35.2%	0.6%
R3	EHEV(PFR) → DLI.PI	13	2	36.5%	34.0%	0.6%
R4	EHEV(PFR) → DLI.PII	13	2	38.4%	34.0%	0.6%
R5	EHEV(PFR) + DLI.PI → DLI.PII	13	2	40.3%	33.3%	0.6%
Notes:		DLLPI: Disaster Loss Period 1 VARI: Input Variable EHEV: Extreme Heat Event Vulnerability PFR: Percentage Fractional Ranks				

TABLE XXIV: VALIDATION METRICS FOR PERCENTAGE FRACTONAL RANKS DERIVAITON APPROACH, GEORGIA 2.0

	Method (Age adjusted	**Var**	**LC**	**OCP**	**UE**	**CF**
PT	PT DLI.PI → DLI.PII	1	n/a	44.0%	25.8%	1.9%
R6	EHEV(PFR(00)) → DLI.PI	13	no	35.8%	31.3%	2.2%
R7	EHEV(PFR(00)) → DLI.PI	13	3	38.8%	30.6%	2.2%
R8	EHEV(PFR(00)) → DLI.PI	13	2	35.8%	32.1%	2.2%
R9	EHEV(PFR(00)) → DLI.PII	13	2	36.9%	33.8%	2.3%
R10	EHEV(PFR(00)) + DLI.PI → DLI.PII	13	2	40.0%	30.8%	1.5%
Notes:		DLLPI: Disaster Loss Period 1 VARI: Input Variable: EHEV: Extreme Heat Event Vulnerability PFR: Percentage Fractional Ranks				

TABLE XXV: METHOD AND VALIDATION METRICS FOR EHE INDICES DERIVED WITH PCA; GEORGIA

Method		Input Var	PCs (rot)	Var. Exp	OCP	UE	CF
AC	Perf. Threshold	1	n/a	n/a	44.0%	25.8%	1.9%
1	PCA → DLI.PI	13	VARI	81.2%	36.6%	30.6%	1.5%
2	PCA → DLI.PI	12	no	76.2%	40.3%	29.9%	2.2%
3	PCA → DLI.PI	12	VARI	76.1%	45.5%	26.9%	1.5%
4	PCA → DLI.PI	12	VARI	81.5%	41.8%	27.6%	1.5%
5	PCA → DLI.PI	12	VARI	79.8%	44.0%	27.6%	1.5%
6	PCA → DLI.PI	13	VARI	82.1%	41.0%	29.1%	0.7%
7	PCA → DLI.PI	12	VARI	82.3%	36.6%	35.1%	2.2%
8	PCA → DLI.PI	12	VARI	87.1%	32.8%	32.1%	2.2%
9	PCA → DLI.PI	11	VARI	72.6%	47.8%	26.1%	0.7%
10	PCA → DLI.PI	10	VARI	71.5%	48.5%	25.4%	0.7%
11	PCA(10) → DLI.PII	10	VARI	71.5%	43.1%	24.6%	1.5%
12	PCA(10) + DLI.PI → DLI.PII	n/a	n/a	n/a	46.2%	24.6%	2.3%
Notes:		PCA: Principle Component Analysis DLLPI: Disaster Loss Period 1 VARI: Input Variable					

TABLE XXVI: CATEGORY III MEANS OF PERIOD 1 AND 2

Mean One*	Mean Two*	Correlation	Mean Difference*	95% Conf. ...	T-Test	df
190.302	109.998	0.236	80.304	58.189	7.184	129
115.472	87.604	Strong	127.446	102.420		
10.128	7.683		11.178			
130	130		130			

*Cells contain: Mean, Standard Deviation, Standard Error, Count

TABLE XXX: VARIABLES ARE INCLUDED FOR THE VALIDATION MODEL STRUCTURE

MORBIDITY_P_I_RATE_TILE4

MEDIAN_INCOME
PCT_NO_VEHICLE
PCT_PUB_ASSIST
PCT_FEMALEHOUSE_CHILDREN
PCT_AA
PCT_UNEMPLOYED_16_OVER
MEDIAN_HOME_VALUE
P_BELOW_POVERTY
P.NLC
PCT_65_OVER
cEHEE.FIPS

TABLE XXXI: DT APPROACH FOR DERIVING EHEV INDEX, GEORGIA

Method	In. Var.	LC	OCP	UE	CF	CUE	38.7 %
AC	PT	1	n/a	44.0%	25.8%	1.9%	14.7 %
1	DT → DLI.PI	13	No LC	80.6%	9.0%	0.7%	8.8%
2	DT → DLI.PI	13	3	82.1%	12.7%	0.0%	8.8%
3	DT → DLI.PI	13	2	79.9%	11.9%	0.0%	5.9%
4	DT → DLI.PI	13	2	76.1%	11.9%	2.2%	11.8 %
5	DT → DLI.PI	13	2	83.6%	12.7%	0.0%	20.5 %
6	DT → DLI.PI	12	2	82.1%	9.0%	0.0%	8.6%
7	DT → DLI.PI	11	1	88.8%	6.0%	0.0%	8.6%
8	DT → DLI.PI	10	1	88.8%	6.0%	0.0%	9.4%
9	DT → DLI.PI	9	1	86.6%	8.2%	0.0%	9.1%
10	DT → DLI.PI	8	1	87.3%	7.5%	0.0%	33.3 %

TABLE XXXII: VARIABLE SELECTION AND PA METRICS

Variable selection and PA metrics						
AC		2	6	7	8	10
CUE		10.4%	2.2%	0.0%	0.0%	1.5%
CF		0.0%	0.0%	0.0%	0.0%	0.0%
UE		12.7%	9.0%	6.0%	6.0%	7.5%
OCP		82.1%	82.1%	88.8%	88.8%	87.3%
Variabl.		13 + 3	12 + 2	11 + 1	10 + 1	8 + 1
Target		DL.PI	DL.PI	DL.PI	DL.PI	DL.PI
INPUT VARIABLES	Socioeconomic and exposure	MedInc	MedInc	MedInc	MedInc	MedInc
		%NoVeh	%NoVeh	%NoVeh	%NoVeh	
		%PubAssi	%PubAssi	%PubAssi	%PubAssi	%PubAssi
		%FHous	%FHous	%FHous	%FHous	%FHous
		%5Und				
		%65Abov	%65Abov	%65Abov	%65Abov	%65Abov
		%AA	%AA	%AA	%AA	%AA
		%NoHSE	%NoHSE	%NoHSE		
		%BelPov	%BelPov	%BelPov	%BelPov	%BelPov
		%Unep>16	%Unep>16	%Unep>16	%Unep>16	
		%MobHome	%MobHome			
		MedHomVal	MedHomVal	MedHomVal	MedHomVal	MedHomVal
		cEHEE	cEHEE	cEHEE	cEHEE	cEHEE
	Land Cov	NLC	NLC	NLC	NLC	NLC
		MLC				
		T.ULC	T.ULC			

VITA

NAME: Bakari Baker

EDUCATION: B.S., Biology, Clark Atlanta University, Atlanta, GA 1999

MPH, Public Health Sciences, University of Illinois at Chicago, Chicago, IL 2010

Ph.D., Public Health Sciences, University of Illinois at Chicago, Chicago, IL 2019

TEACHING: Environmental and Occupational Health Science Division, University of Illinois at Chicago, 2013 – 2019

HONORS: Diversifying Higher Education in Illinois Fellow, University of Illinois at Chicago, Chicago, IL 2014 – 2019

EXPERIENCE: Environmental Health Scientist, U.S. EPA, February 2012 to Present

Nuclear Oversight Auditor, Exelon Corporation, November 2010 to February 2012

Operational Chemistry Supervisor, Exelon Corporation, October 2008 to November 2010

Environmental Chemist, Exelon Corporation, August 2008 to October 2008

Analytical Chemist II, Exelon Corporation, May 2006 to August 2008

PUBLICATIONS: Bakari Baker, Patrick Milligan, Anne Bavuso, Kate Robinson: "EPA Needs to Improve Its Process for Accurately Designating Land as Clean and Protective for Reuse"; U.S. Environmental Protection Agency, Office of Inspector General; Report No. 14-P-0364; 2014; September 2014.

Bakari Baker, Patrick Milligan, Anne Bavuso, Kate Robinson, Natasha Henry: "Hyperspectral Imaging Can Be a Useful Evaluation Tool for Office of Inspector General Reviews Focused on Contaminated Land"; U.S. Environmental Protection Agency, Office of Inspector General; Report No. 14-N-0360, September 2014.

Bakari Baker, Lauretta Joseph, Heather Drayton, Kalpana Ramakrishnan; Jeffery Harris: EPA's Risk Assessment Division Has Not Fully Adhered to Its Quality Management Plan; Report No. 14-P-0350; September 2014.