APPLICATIONS OF GEOSPATIAL ANALYSIS TECHNIQUES FOR PUBLIC HEALTH

Austin Curran Stanforth

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the Department of Earth Science,
Indiana University

September 2016

Accepted by the Graduate Faculty, Indiana University, in partial
fulfillment of the requirements for the degree of Doctor of Philosophy.

_____
Gabriel Filippelli, Ph.D., Co-Chair

_____
Daniel P. Johnson, Ph.D., Co-Chair

Doctoral Committee

_____
Lixin Wang, Ph.D.

May 2, 2016

_____
Jeffrey Wilson, Ph.D.

_____
Max J. Moreno-Madriñán, Ph.D.

_____
Pierre-André Jacinthe, Ph.D.

ACKNOWLEDGEMENTS

I would like to dedicate this production to my family, who provided me a strong foundation through their endless source of love, encouragement, and support. Without them this project and degree would not have been possible.

I would like to thank Jack and all my un-named Buds who helped keep me in a good mindset during this journey.

Austin Curran Stanforth

APPLICATIONS OF GEOSPATIAL ANALYSIS TECHNIQUES FOR PUBLIC HEALTH

Geospatial analysis is a generic term describing several technologies or methods of computational analysis using the Earth as a living laboratory. These methods can be implemented to assess risk and study preventative mitigation practices for Public Health. Through the incorporation Geographic Information Science and Remote Sensing tools, data collection can be conducted at a larger scale, more frequent, and less expensive that traditional *in* situ methods. These techniques can be extrapolated to be used to study a variety of topics. Application of these tools and techniques were demonstrated through Public Health research. Although it is understand resolution, or scale, of a research project can impact a study's results; further research is needed to understand the extent of the result's bias. Extreme heat vulnerability analysis was studied to validate previously identified socioeconomic and environmental variables influential for mitigation studies, and how the variability of resolution impacts the results of the methodology. Heat was also investigated for the implication of spatial and temporal resolution, or aggregation, influence on results. Methods studying the physical and socioeconomic environments of Dengue Fever outbreaks were also studied to identify patters of vector emergence.

Gabriel Filippelli, Ph.D., Co-Chair

Daniel P. Johnson, Ph.D., Co-Chair

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Dissertation Introduction

Geospatial research is conducted using the Earth as a living laboratory, which requires an understanding of neighborhood interactions and a method to quantify spatial interactions. Geographic Information Science (GIS) is a field which was developed with the intention of studying the natural environment and consider relational space. Although the lay community has generally referred to GIS as a tool for the collection, storage, processing, and retrieval of spatial data (Kurland & Gorr, 2012). GIS allows for the more holistic collection, study, and use of spatial data than many other research methods (Bolstad, 2015). Tools developed for this field allow for the quantification of variables, and consider their physical locational relationships. The following chapters describe research projects which were developed and processed with tools developed for the GIS community. Data used in these projects include Earth satellite remote sensing products, population demographics, medical records, and similar survey-derived products used to quantify social conditions of the study populations.

One of the earlier epidemiological and health GIS projects was the work of Dr. John Snow during the London Cholera epidemic in 1831-1832. Through his application of a germ theory of disease and basic cartographic techniques, John Snow quantified the distribution of Cholera incidence within the European city (Cameron & Jones, 1983). Even without the technology to identify the specific pathogen, John Snow was able to correctly identify contaminated water as the cause of the illness, support that theory with cartographic data, and implement measures to reduce its spread (Brody, Rip, Vinten-Johansen, Paneth, & Rachman, 2000). John Snow initiated mitigation methods, including the removal of the Broad Street pump handle, to reduce use of the contaminated water source. Similarly, the following chapters describe the use of GIS techniques to study medical conditions.

Assessment of fine scale modeling of Extreme Heat Vulnerability Indexes

The Extreme Heat Vulnerability Index (EHVI) was developed to be used by public health and emergency management personnel to improve the collective understanding of health risks during extreme heat events (EHEs) (Dan Johnson, Lulla, Stanforth, & Webber, 2011; Daniel P Johnson, Stanforth, Lulla, & Luber, 2012). The model uses residential space to identify vulnerable populations who are more susceptible to the negative influences of EHEs. The EHVI was developed to integrate previously separate physical and social vulnerability modeling approaches (Daniel P Johnson, Webber, Urs Beerval Ravichandra, Lulla, & Stanforth, 2014; Stanforth, 2011).

Research into the vulnerability of populations has commonly focused on either socioeconomic or environmental conditions of the population. Social studies focus predominantly on socioeconomic or population variables which identify groups or individuals who are less capable of surviving or recovering from an oppressive event. These studies often include populations with lower educational attainment or financial status (Susan L. Cutter, Boruff, & Shirley, 2003; Davis, 1997; S. L. Harlan, Brazel, Prashad, Stefanov, & Larsen, 2006). Age is also frequently cited in vulnerability studies, as older individuals have a higher rate of chronic health issues and are less physically fit (Ebi, Teisberg, Kalkstein, Robinson, & Weiher, 2003; Luber & McGeehin, 2008; Semenza et al., 1996). Physical vulnerability studies the degree to which a population is influenced by the oppressive event, generally without considering their ability to recover. Extreme heat can affect populations differently due to the thermal influence of the local environment (McMichael et al., 2008; Robinson, 2001; Stone, Hess, & Frumkin, 2010; Tan et al., 2010). The urban heat island, a phenomenon caused by impervious surface's increased capacity to store energy, can disproportionately impact local populations (Voogt &

Oke, 2003; Q. Weng & Lu, 2008; Q. H. Weng & Quattrochi, 2006; Xiao et al., 2007; X. Zhang, Zhong, Wang, & Cheng, 2009; Zhou & Shepherd, 2010).

Previous work has suggested averaging data across large areas, such as a state or county, can cause aggregation bias, commonly identified as the Modifiable Aerial Unit Problem (MAUP) (Kwan, 2012a, 2012b). Therefore, further research is needed to improve the understanding of how social and physical variables respond to changes in resolution, such as between census tract or block boundaries (Stanforth, 2011). Larger boundaries also do not provide a means to identify where people are located, which can lead to assumptions of continuous distributions of variables throughout the boundary without concern for habitable space (Johnson et al. 2011, Johnson et al. 2009, Johnson et al. 2012, Reid et al. 2009). This can complicate mitigation practices, since the spaces occupied by vulnerable populations have not been properly identified. This can lead to resources being misallocated to populations, or locations, who might not require the assistance.

Current weather warnings are plagued by these aggregation limitations. Most forecasting tools and weather alerts are conducted at county or municipality boundaries, which are unable to distinguish vulnerable populations within the community (Susan L. Cutter et al., 2003; Laurence S. Kalkstein, 1991; Laurence S. Kalkstein, Jamason, Greene, Libby, & Robinson, 1996; Robinson, 2001). This can contribute to a response based mitigation plan, such as is the case with 911 assistance, where medical assistance is provided after individuals are already in danger rather than providing preventative relief. Vulnerability modeling, such as the EHVI, can identify vulnerable populations before oppressive events to facilitate preventative actions and reduce the medical or financial burden of oppressive events (Cutter et al. 2010, Ebi et al. 2003, Johnson et al. 2012, Johnson et al. 2013, Stanforth 2011, Harlan et al. 2012). If smaller

boundaries are used for modeling practices, it could allow for improved allocation of limited mitigation resources.

Previous research compared the results of census tract and block group modeling during the Chicago, IL 1995 heat wave (Stanforth 2011). Those results demonstrated the block group analysis was better suited to model vulnerability. This project compared census block group and block level resolution data. Principal component analysis was used to examine the explanation of variance and variable correlation between the two boundary resolutions (Johnson et al. 2011, Johnson et al. 2012, Stanforth 2011). The results compared the use of higher resolution modeling, beyond what had been previously documented in the literature (Susan L. Cutter et al., 2003; C. E. Reid et al., 2009; Stanforth, 2011), and discussed potential future investigations on the aggregation bias's impact on both social and environmental variables commonly employed in vulnerability studies (Kwan, 2012a, 2012b). Environmental and social variables were also discussed whether they could continue to be used in future applications of the EHVI or if alternative variables are better suited for this type of modeling.

2. Remote Dengue Fever Modeling in the Río Magdalena Watershed

Remote sensing data is now frequently used to study the distribution of vector diseases. Satellite and geospatial products have become more available, allowing for more efficient data collection of large study areas compared to more traditional *in* situ field methods. Geographic Information Science (GIS) tools can use this data to estimate vector spread, virus activity, and population vulnerability, allowing for monitoring and research programs to become more affordable and prevalent. Research and warning systems are currently being developed for a variety of diseases across the globe; the NASA Goddard Space Flight Center's Healthy Planet program has already initiated a global Malaria transmission model (Cleckner & Allen, 2014). Recent literature suggests similar research projects are similarly focused on mosquito

monitoring, or modeling, for a variety of other mosquito vector diseases (Beebe, Cooper, Mottram, & Sweeney, 2009; Cleckner & Allen, 2014; Delmelle, Dony, Casas, Jia, & Tang, 2014; Fuller, Troyo, & Beier, 2009; Z. Li et al., 2016; Liu, Weng, & Gaines, 2008). One such vector is the *Aedes aegypti* mosquito, which is capable of spreading a variety of diseases, including Dengue Fever, Yellow Fever, Chikungunya, and Zika. These diseases are commonly referred to as tropical and subtropical climate disease, due to the more moderate range of temperatures required for the reproduction of the mosquito vector and virus incubation. Recent climate change models now suggest the vector could expand into higher latitude areas due to warming climates (Brady et al., 2013; Ruiz-López et al., 2016). Therefore, improved modeling is needed to identify the current populations who are at risk of coming in contact with infected vectors, and to identify expanding habitable zones.

Dengue Fever is a viral disease with symptoms including fever, rash, and orthopedic discomfort (CDC, 2016; WHO, 2016). As no vaccine is currently available for wide distribution, the best prevention practices include the reduction of mosquito rearing sites and contact, through proper hygiene, repellant, and screen/netting (Gubler, 1998; WHO, 2016). This makes it critical to identify potential vector breeding sites, so local population can take the proper precautions. The literature indicates that the mosquito tends to breed near populated areas by using anthropogenic or peridomestic water sources, such as those found in discarded tires, flower pots, water storage containers, or impervious surface catchments (Beebe et al., 2009; Gubler, 1998). This can increase the prevalence for vector - human interaction and virus dispersion in higher populated areas (Beebe et al., 2009; Cleckner & Allen, 2014; Delmelle et al., 2014). Meteorological and elevation data have been identified to influence the propensity of local mosquito habitat sites (Brunkard, Cifuentes, & Rothenberg, 2008; Moreno-Madriñán et al., 2014). Therefore, remote sensing data can provide meteorological, elevation, and

environmental variables to model with medical records to identify patterns of mosquito vector and virus habitation within an area of interest. This can improve the ability to accurately identify the current extent, and expansion potential, of Dengue Fever transmission, as has been done with other vector viruses (Ceccato, Connor, Jeanne, & Thomson, 2005; Cleckner & Allen, 2014; Liu et al., 2008).

Vector habitat and virus transmission investigation was conducted in the Colombian Río Magdalena watershed in South America. The watershed is bounded by the Andes mountain range on three sides and the Caribbean Sea to the north. Satellite data provided environmental variables to identify potential habitable areas within the watershed. Dengue Fever cases were identified by the Instituto National De Salud (National Institute of Health) of Colombia at the municipality level over the years of 2012-2014. Exploratory factor analysis was conducted as a preliminary analysis to identify environmental and socioeconomic variables which were useful to identify transmission areas, which could also be incorporated into for future applications of forewarning models. The results identified the variables with the strongest correlation to rates of Dengue Fever outbreak, such as Elevation and Temperature.

3. Space and Time: Assessing Heat Mortality Distribution

Previous studies have demonstrated the impact socioeconomic and environmental variables have on vulnerability modeling, or patterns of negative health outcomes (S. L. Harlan et al., 2006; Daniel P Johnson et al., 2012; Naughton et al., 2002; O'Neill, Zanobetti, & Schwartz, 2005). Populations classified as elderly, age 65 and older, or living in poverty have been repeatedly documented as having higher rates of health complications during extreme heat events (Andrew, Mitnitski, & Rockwood, 2008; Changnon, Kunkel, & Reinke, 1996; Susan L. Cutter et al., 2003; Davis, 1997). Although the socioeconomic and environmental variables have

been studied, less information has been presented on the distribution of heat mortalities across space and time.

Patterns of heat mortality were investigated for spatial and temporal clusters over a three year period, 2009 - 2011. The study was conducted in Maricopa County, AZ, which incorporates Phoenix, and is an arid region in the southwest United States. Spatial analysis tools from Esri ArcMap were used to identify statistical patterns or clusters. Moran's I was used to describe the measure of spatial autocorrelation (Lozano-Fuentes et al., 2012). Additionally, a High-Low Clustering Report using a Getis-Ord General G analysis, compared a feature's weight, with its neighbors, against the overall population's local sum. The Getis-Ord is not designed to detect individual anomalies or outliers, like the Moran's I, but clusters or neighboring features with similar deviation values from the population (Beebe et al., 2009). This allows the Getis-Ord to be viewed more as a population based metric. These neighboring clusters were visually represented through the use of a Mapping Clusters - Hot Spot Analysis tool (Lozano-Fuentes et al., 2012). Results from the individual three years and a composite three year assessment were compared for similarities. The results demonstrate the propensity for heat mortalities to occur in similar, or adjacent, neighborhoods within Maricopa County. Future mitigation practices could use these results to investigate or implement new mitigation practices. Additionally, the ArcMap Spatial Time Cube and Emerging Hot Spot analysis tools were tested as a validation method for the previously identified spatial clusters, but were unable to provide consistently supportive results due to inefficient length of data.

This study was designed to assess whether reported cases of heat induced mortality occur at random, or cluster in space and time. Previous research has operated under the assumption vulnerability, and general impacts on health, are static when considering geography and time. The collected mortality data and cluster results support this assumption, depending

on resolution. However, due to the data limitations in this study, the results do not facilitate an extrapolation to other locations or study periods.

Chapter 1 Assessment of Fine Scale Modeling of Extreme Heat Vulnerability Indexes

1.1 Abstract

Heat is considered the leading cause of weather-related mortality throughout the world. Research and application of warning / mitigation plans can improve community resilience. Previous research has focused on identifying socioeconomic or environmental variables common to victims of extreme heat events. Less research has been conducted on the interrelationship between Earth observation derived environmental variables and socioeconomic variables, or whether these variables are sensitive to aggregation error. This study investigated the influence of environmental and socioeconomic variables' impact on extreme heat vulnerability by comparing their results at varying spatial resolutions.

Studies of this nature have been commonly conducted at the State or County resolution, while higher resolution models, such as block groups, have been found to improve applicability. U.S. Census socioeconomic and environmental variables, obtained by satellite imagery, were statistically studied against heat-related mortality using different spatial resolutions. Following the methods of the previously developed Extreme Heat Vulnerability Index (EHVI), principal component analysis was conducted on twenty-nine documented indicators of heat-health hazards at the U.S. Census block group and block scale. Results suggest block scale analysis is beneficial for mitigation management, but requires additional variable preparation since not all variables are readily available at this higher resolution. Environmental variables collected using lower resolution platforms were less prevalent within the smaller block boundaries, suggesting an aggregation bias is present. Both the block and block group scales demonstrated an improved mitigation potential over currently implemented weather service warnings.

1.2 Background

Extreme heat is currently considered to be the deadliest weather pattern on the planet (Patz et al., 2000; Robinson, 2001). The 1995 Chicago, IL heat wave, where 700 mortalities were

reported during a five-day heat wave, is one representation of how dangerous these events can be (Donoghue et al., 1997; Semenza et al., 1996; Shen, Howe, Alo, & Moolenaar, 1998; Whitman et al., 1997). Extreme heat often does not exhibit discernible warning signs; tornados, hurricanes, and blizzards commonly exhibit clouded air mass fronts, a change in wind pattern, or even swift temperature changes which can alert local populations of the approach and severity of inclement weather. Without the use of meteorological measuring equipment, heat waves may not be perceived as dangerous until medical complications have ensued. The cumulative influence of several consecutive oppressive days further increases the hazardous nature of heat waves (Robinson, 2001). This study was designed to build upon previous research of the Extreme Heat Vulnerability Index (EHVI). The index has previously explored socioeconomic and environmental variables used in heat wave vulnerability modeling. This study further explored the influence resolution, or scope of view, has on the results (Daniel Johnson, Wilson, & Luber, 2009; Daniel P Johnson et al., 2012; Daniel P Johnson et al., 2014; D. P. Johnson & Wilson, 2009; Stanforth, 2011). The current study examined census block resolution in comparison to the previous block group and census tract results.

Due to the health risks associated with extreme heat, the National Weather Service (NWS) has mandatory population notifications during heat waves (Ebi et al., 2003). The standard NWS uses a single risk threshold of 105°F, for two or more days, within the continental United States. This single criteria does not account for the regional climate differences found within the nation. The cool, moist conditions of the Pacific Northwest and the hot, dry weather of the Southwest experience dramatically different climates. This single threshold may rarely indicate dangerous weather conditions in Seattle, WA, but could lead to over estimation of risk in the arid climate of Phoenix, AZ. Previous researchers have noted a single national threshold cannot appropriately identify local dangerous conditions (Robinson, 2001). Another popular local metric

is Robinson's Threshold, which tracks local temperature conditions over several years to document the distribution of the local climate. A Robinson's heat wave alert is issued when temperatures exceed the weather record's 95th percentile over consecutive days, but does not consider impacts on health (Robinson, 2001). Kalkstein's Heat/Health Watch Warning System (HWWS) also uses local measurements (including temperature, humidity, wind) to identify weather patterns which corresponded to a statistical increase in heat mortalities (Laurence S. Kalkstein, 1991; Laurence S. Kalkstein & Greene, 1997; Laurence S. Kalkstein et al., 1996). The HWWS is currently used in several cities across the globe, and has been suggested to improve resilience to adverse weather conditions by increasing notification adherence (L. S. Kalkstein, Greene, Mills, & Samenow, 2011; Whitman et al., 1997). All of these systems rely on low spatial resolution weather forecast data, making them unable to identify which neighborhoods are more vulnerable to the conditions, and impede the ability to identify local mitigation needs or risk prevention practices (S. L. Cutter et al., 2008; Susan L. Cutter et al., 2003; S. L. Harlan et al., 2006; Daniel P Johnson et al., 2012; Daniel P Johnson et al., 2014; Naughton et al., 2002; O'Neill et al., 2005; Rinner et al., 2010; Semenza et al., 1996; Whitman et al., 1997).

Vulnerability studies frequently use non-meteorological variables to ascertain the impact oppressive weather events have on local populations. These are commonly known as either Social Vulnerability or Hazard-first Vulnerability studies. Social vulnerability modeling focuses on the ability of individuals to adapt or recover from oppressive events. Some of the more cited variables in these models include age, financial status, and educational attainment (S. L. Cutter et al., 2008; S. L. Cutter, Burton, & Emrich, 2010; Davis, 1997; S. L. Harlan et al., 2006; Naughton et al., 2002; Semenza et al., 1996; Stanforth, 2011). These variables can help identify the resilience or location of vulnerable populations, but are often not specific to an explicit type of event (ex. Heat).

The Social Vulnerability Index (SoVI) was designed to use socioeconomic variables collected by the U.S. Census to identify higher risk populations (Susan L. Cutter et al., 2003). The SoVI was created as a national assessment using county boundaries, and therefore is unable to represent local population variations. Reid et al. (2009) similarly investigated the implementation of a national assessment at the census tract resolution for urban environments (C. E. Reid et al., 2009). Their study used socioeconomic data from the U.S. Census and land cover classification data to incorporate a measure of vegetated land within boundaries. However, since the national land cover maps used are only produced on average every ten years, this approach may not provide a reliable measure of current environmental conditions (Daniel P Johnson et al., 2012; C. E. Reid et al., 2009; Stanforth, 2011).

Several social vulnerability studies have consistently identified variable categories which increase a population's risk. Age, low income, and lower educational attainment are commonly cited indicators of social vulnerability (Davis, 1997; McMichael et al., 2008; Naughton et al., 2002). Populations with higher education tend to have an improved financial status and more employment security. Populations with higher income also tend to be able to afford better housing options with improved insulation, air condition availability, and are constructed in less dense configurations. Expendable income, in particular, is considered a protective variable during extreme heat events; it increases the opportunity to visit medical facilities when health is impaired and grants access to air conditioned establishments should their residence not be equipped with an operating cooling system (Davis, 1997; O'Neill et al., 2005; C. E. Reid et al., 2009; Semenza et al., 1996). Restaurants, movie theaters, and even community pools often require a financial transaction to use their facilities. Individuals with less income may not be able to access these areas which provide an opportunity to restore safe body temperatures.

Previous socioeconomic vulnerability studies include age demographics as an indicator of the overall health status of a population. Both very young and very old individuals may require enhanced assistance during extreme heat events; reduced mobility can impede an individual from relocating to safer environments or acquire fluids from an adjacent room to maintain proper hydration (Changnon et al., 1996; Naughton et al., 2002; Patz et al., 2000; C. E. Reid et al., 2009; Whitman et al., 1997). There are also higher documented rates of chronic illnesses, such as diabetes and cardiovascular disease, which can negatively influence survivability in older populations (CDC, 2002, 2006; Susan L. Cutter et al., 2003; S. L. Harlan et al., 2006; Hondula et al., 2012; Robinson, 2001; Semenza et al., 1996; Stanforth, 2011).

Hazard-first, or environmental, vulnerability studies focus on the proximity and intensity of oppressive events. These studies focus on an event's magnitude and the population's proximity to the impacted areas, but do not indicate how well the local population could adapt or recover (Wisner, 2004). Proximity to heat exasperating features can increase the oppressive impact on populations during heat waves, demonstrating a need to identify populations located in higher hazard areas.

Vegetation has been found to be a protective element during extreme heat events. Photosynthesis and evapotranspiration in healthy vegetation absorb and reduce solar energy in the near environment (Jensen, 2005). Increased vegetation can reduce the overall thermal load within a neighborhood as well as provide shade. Vegetation has also been found to reduce erosion, improve air quality, and reduce local flooding, making it widely used in a variety of vulnerability studies (Davis, 1997; Daniel Johnson et al., 2009; Daniel P Johnson et al., 2012; Lo & Quattrochi, 2003; Luber & McGeehin, 2008; Naughton et al., 2002; Uejio et al., 2011; Wilhelmi & Hayden, 2010; X. Zhang et al., 2009). The normalized difference vegetation Index (NDVI) is one

of the more commonly used vegetation metrics due to its ability to identify vegetation across a wide variety of climate and geographical features (Jensen, 2005; McCabe, Wang, & Lu, 2016).

Conversely, impervious surfaces, such as concrete and asphalt, absorb and re-emit solar energy into the local environment, increasing thermal impact (Chen, Zhao, Li, & Yin, 2006; Lo & Quattrochi, 2003; Voogt & Oke, 2003; Wang, Zhu, & Wang, 2004; Q. Weng & Lu, 2008; Q. H. Weng, Lu, & Schubring, 2004; Q. H. Weng & Quattrochi, 2006; Xiao et al., 2007; Yuan & Bauer, 2007; X. Zhang et al., 2009; Y. S. Zhang, Odeh, & Han, 2009; Zhou & Shepherd, 2010). This causes an increase in experienced daytime temperature, but the built environment also continues to emit thermal energy at night (Voogt & Oke, 2003). This can inhibit normal diurnal thermoregulation among urban residents (Voogt & Oke, 2003). These variables can be measured with *in situ* sensors for a direct measure of environmental variables, such as ambient temperature, but the cost and placement of such sensors is often prohibitory. Remote sensing applications are commonly used to provide indirect measures.

Satellite data can provide a measure of the UHI influence by quantifying the temperature of surface features through the collection of data within the thermal electromagnetic spectrum (Chander, Markham, & Helder, 2009; Jensen, 2005). Since the UHI is a manmade phenomenon, building practices have contributed to the severity of heat waves. Growing urban populations have been found to increase urban sprawl and development density, increasing the UHI (Stone et al., 2010; Voogt & Oke, 2003; Wang et al., 2004; Q. H. Weng & Quattrochi, 2006; X. Zhang et al., 2009). Therefore, landscape temperature variables have been found to help understand the severity of thermal loading on local vulnerable populations (S. L. Harlan et al., 2006; Daniel P Johnson et al., 2012; Daniel P Johnson et al., 2014; K. Li & Yu, 2008; Q. H. Weng et al., 2004; Xiao et al., 2007; Yuan & Bauer, 2007; Zhou & Shepherd, 2010). It should also be noted there is limited research on the relationship between

surface and ambient temperature, but they are anticipated to have a strong positive relationship (Voogt & Oke, 2003; Q. H. Weng & Quattrochi, 2006). The influence of impervious surfaces and built density can assist with identifying local thermal impact through the normalized difference built-up index (NDBI).

A relationship can also exist between the physical environment and socioeconomic variables (Dan Johnson et al., 2011). Lower income residences have been shown to experience more dense construction configurations, leaving a lower quantity of vegetated space, and increase the local UHI severity (S. L. Cutter et al., 2008; Hondula et al., 2012; Jensen, 2005; Papanastasiou, Melas, Bartzanas, & Kittas, 2010; Stone et al., 2010; Q. H. Weng et al., 2004). This project was designed to further study the interrelationship between variables through a composite vulnerability study, building upon the Johnson et al. (2012) development of the Extreme Heat Vulnerability Index (EHVI) (Daniel P Johnson et al., 2012; Stanforth, 2011). The EHVI methods combine social and physical modeling capabilities to improve vulnerability assessment for extreme heat monitoring. Prior research assessed the model's ability to better identify locations of vulnerable populations over previous independent socioeconomic or environmental systems. The contribution of the current study was to evaluate the effectiveness of the EHVI, and the vulnerability variables it detects, at a finer resolution. The previous research analyzed the census tract and block group boundaries using 1990 Census data, while this study used block group and block resolutions with 2000 Census data (Daniel P Johnson et al., 2012; Stanforth, 2011).

Principal component analysis (PCA) is a statistical method for dimension reduction, which is useful to identify and decrease redundancy between input variables by identifying correlations between them (DeCoster, 2004). This method can help researchers identify the prominent variables in a dataset. The method also compares the variables' relationship to a

single dependent variable to identify the more influential variables within the dataset (StatSoft, 2015). The PCA identified independent variables' similarities with the dependent variable through a varimax rotation of plotted variable values. This rotation allows for the creation of new composite, or principal, components which incorporate a mix of the attributes from the input variable, to varying degrees. The strength of the PCA is demonstrated by the quantity of these principal components, and the explanation of variance from the dataset which they describe. Principal components contain at least as much information as one original variable. Following the Kaiser Criterion, principal components must have an eigenvalue greater than one (1) (Kaiser, 1960; UCLA). Although the PCA will produce as many component variables as the number of input variables, only components which pass the Kaiser Criterion are used for modeling, reducing the number of input variables used to explain a majority of the variance, which simplifies the analysis. The strength of the model is demonstrated by the quantity of principal components. Fewer principal components indicate the variables are more correlated, or relatively similar, to the dependent variable. The cumulative explanation of variance of the principal components demonstrate how much of the entire dataset's variability is explained by them; a higher explanation of variance indicates less noise in the data set and the input variables were appropriately picked to identify the dependent variable.

Zonal statistics provide a measure of the environmental variables for the PCA, by calculating the raster variable values from within predefined boundaries. The EHVI methodology is focused on residential space, where it was anticipated people spend the majority of their time. Since urban environments, such as Chicago, can have mixed residential and commercial space, the National Land Cover Dataset (NLCD) raster was used to better filter out non-residential space before collecting local environmental variable values (Daniel P Johnson et al., 2012; Stanforth, 2011). The NLCD dataset is a land use and land classification product developed

by the U.S. Geological Survey (USGS), through the use of Landsat 5 TM satellite data, circa 1992. NLCD classes 21 and 22 (identified as low and high intensity residential space, respectively) were used to identify habitable or residential land cover within the study area (Vogelmann et al., 2001). Low density residential areas represent low building density with larger vegetated space, as you would find in suburban or singly family dwelling residences. High density residential areas contain more closely built homes and multi-family dwellings, such as you would find in apartment complexes or mixed commercial space. The literature identifies NLCD class 23 as larger Commercial, Industrial, or Transportation features, suggesting most mixed residential/commercial space should be identified within the 21 and 22 classes. This procedure can reduce the influence of potential confounding errors caused by large industrial or park spaces which could skew the results (Daniel P Johnson et al., 2012; Stanforth, 2011). Heat waves are commonly not associated with windy conditions, so local environmental weights will have limited impact by external features. This supports the removal of adjacent commercial features which would skew the residential environmental variables. This method was successfully implemented during the 2012 EHVI analysis, which used 1990 census boundaries (Daniel P Johnson et al., 2012; Stanforth, 2011). Building codes were initially considered for this designation of land use, but the data was not available at the Cook County boundary level during the time of the project. Though it should be noted the high density and mixed land use of the highly urbanized city of Chicago may still inflict a land use misclassification limitation in the study.

1.3 Methods

The study was conducted in Cook County, Chicago, IL in response to the July 12-16th, 1995 heat wave (Susan L. Cutter et al., 2003; S. L. Harlan et al., 2006; Daniel P Johnson et al., 2012; Semenza et al., 1996). Socioeconomic data was collected during the 2000 U.S. Census at

the block and block group levels (U.S.Census, 2000). These political boundaries are used to identify the distribution of populations, and decrease in size from: county > census track > block group > block. Census data included: race, age, educational attainment, and financial status (Susan L. Cutter et al., 2003; S. L. Harlan et al., 2006; Daniel P Johnson et al., 2012; Semenza et al., 1996; Stanforth, 2011). It should be noted the U.S. Census did not collect the entire spectrum of variables incorporated at the block resolution, therefore some variables were interpreted from the block group data. Population and race variables were available at the block level, but variables identifying financial status (population in poverty, Medium Family Income) and age (65 and older, and 5 and under) were interpreted during the block resolution assessments. The interpretation value was created to estimate the distribution of the block population within the block group parent boundary. The proportion of the block's population to the parent block group formed the interpretation value. For the variables not collected at the block resolution, the interpretation value was applied to the value of the parent block group's variable, making them an estimate based off the population proportion. An additional supplementary study used the parent block group value for all of the missing financial variables, rather than interpret them. This was used to see if there were complications to using the proportional interpretation variables for Median Family Income, Per Capita Income, and Medium Household Income. The missing variables were necessary to conduct this analysis so the study would be comparable to the variable set used in previous projects.

The dependent variable consisted of mortality data provided by the Illinois State Vital Records Department in Springfield, IL. Death certificate data were digitized and encoded into a database during the original EHVI study, including cause of death, age, and residential address (Daniel P Johnson et al., 2012). The database was geocoded using 1990 census street centerline files for Chicago with a success rate of 98% (Stanforth, 2011). In total, 720 death certificates

were collected during the heat wave, but only 586 were attributed to the weather in this study. These cases were not identified as accidental, or violent, deaths. It should be noted this study included contributory causes of death, such as cardiovascular or respiratory diseases, as previous studies have demonstrated this does not overestimate the influence of heat on mortalities (Donoghue et al., 1997; Semenza et al., 1996; Shen et al., 1998; Whitman et al., 1997).

To coincide with the 1995 Heat wave, Landsat 5 TM imagery acquired July 1, 1995 was used to quantify the environmental variables. An image more concurrent with the heat wave was not available due to cloud cover and the satellite's temporal resolution of sixteen days. Since the imagery was collected during the same season, similar landscape patterns would be present. Surface temperature values were derived from the satellite imagery using the procedures outlined by Chander, Markham et al. (2009), calculating the at sensor brightness temperature by correcting for sensor calibration. The NDVI and NDBI were similarly calculated using equations provided by Jensen (2005). Z-scores of all environmental and socioeconomic variables were used during the analysis, so the minimal difference in date and weather is not expected to cause any significant impact to the results (Daniel P Johnson et al., 2012; Stanforth, 2011). Satellite data was collected from the publicly available USGS Earth Explorer database. All environmental variables were developed during a previous project which used ERDAS Imagine to derive the raster values, and Esri ArcGIS 10.2 to conduct zonal statistics within the census boundaries.

Table 1-1: Remote Sensing Equations

Variable	Equation		Source
NDVI	$\frac{\text{NIR - Red}}{\text{NIR + Red}}$		Jensen, JR. 2005
NDBI	$\frac{\text{NIR - MidIR}}{\text{NIR + MidIR}}$		Jensen, JR. 2005
Temperature	$\frac{(LMAX_\lambda - LMIN_\lambda)}{Q_{cal\ max} - Q_{cal\ min}}$	$(Q_{cal} - Q_{cal\ min}) + LMIN_\lambda$	Chander, Markham et al. 2009

Principal component analysis (PCA) was conducted using IBM SPSS 23, similar to the methods of the previous EHVI study (Daniel P Johnson et al., 2012). A regression variable output was created for each component based on the independent variables' values within the census boundary. These regression values were combined into a sum factoral value for the boundary (UCLA). The factorals were used to compare the vulnerability level of a boundary to its neighbors by mapping the values, as has been done in previous vulnerability studies (S. L. Cutter et al., 2008; Susan L. Cutter et al., 2003; S. L. Cutter et al., 2010; Daniel P Johnson et al., 2012; C. E. Reid et al., 2009).

Not all of the 2000 census units used in this study contained NLCD class 21 or 22, low and high residential space, respectively (Vogelmann et al., 2001). In total, 9250 of 64201 blocks and 27 of 4241 block groups did not contain NLCD classes 21 or 22. To circumvent this limitation, census boundaries without NLCD residential spaces were separated from the full dataset and a separate zonal statistics was conducted within the full census boundary, so no NLCD classification was used to extract the satellite imagery in these areas. Between the small spatial area covered and the limited number of residents living within these boundaries (68,000 for block and 2,000 block group out of the 5,400,000 total population), this step was not anticipated to impact the study to any large degree. Upon completion of this stage, the boundaries were merged back into their respective datasets to reduce the impact of an error of

omission, which could have occurred if the features without NLCD classes 21 or 22 were excluded from the analysis.

To assess the difference in the EHVI analysis at the block group and block resolutions, the vulnerability classification of the EHVI blocks were documented against their parent block group vulnerability classification. If cohesion in vulnerability classification occurred between the two resolutions, aka the blocks represent similar vulnerability levels as their parent boundary, it would indicate there was little advantage to using the higher resolution. However, if the block vulnerability classification is different from its parent classification, it demonstrates the potential benefits of using the smaller block resolution. These results are considered a test of the Modifiable Aerial Unit Problem (MAUP) (Kwan, 2012a, 2012b). An additional comparison between the model results was conducted through the calculation of the Receiver Operating Characteristic (ROC) Area Under the Curve (AUC) for each resolution (Hanley & McNeil, 1982). The AUC identifies whether the results are able to appropriately identify positive reports of increased vulnerability, where mortalities occurred, or if the model created false positives, being unable to properly identify the distribution of mortality. Higher heat vulnerability is quantified by a higher sum factoral components from the PCA output, if these correspond to areas with reported heat mortality, the results would indicate a true positive result. Comparatively, false positives are identified as expected high vulnerability results in areas without reported mortalities (Hajian-Tilaki, 2013). This method can demonstrate the ability of the PCA regression output to correctly identify increased vulnerability rates between models (Kraemer et al., 2015).

1.4 Results

Both the block group and block resolutions yielded about 80% explanation of variance, demonstrating the principal components represent a majority of the input data's variability. The block group resolution contained 7 principal components and both of the block models (with and without the interpreted financial variables) contained 9. The breakdown of the PCA results

and variable loading can be viewed in the following tables. The variable loadings for both the block group and block resolutions indicate age, educational attainment, and financial status are critical variables to studying extreme heat hazards. Total population is listed in the block group resolution's first component, but is less influential in both block models, suggesting its usefulness is resolution dependent.

Table 1-2: Block Group Resolution PCA results.

	Component	Eigenvalue	% Variance	
Block Group Analysis	1	8.174	28.186	
	2	5.13	17.69	
	3	2.863	9.874	
	4	2.122	7.316	
	5	1.624	5.601	
	6	1.326	4.573	
	7	1.124	3.876	**Cumulative = 77.116%**

	Component						
	1	2	3	4	5	6	7
Male High School Degree	**0.849**	0.152	0.167	0.09	-0.137	0.001	0.071
Female High School Degree	**0.835**	0.035	0.355	0.02	-0.078	-0.091	0.1
Total Population	**0.771**	0.332	0.308	0.277	0.206	0.031	0.077
White Race	**0.672**	0.243	0.346	-0.267	0.404	-0.044	0.049
Male Under 5 years of age	**0.623**	0.542	0.021	0.46	0.035	-0.057	-0.016
Female Under 5 years of Age	**0.617**	0.54	0.022	0.47	0.035	-0.055	-0.016
Male No High School Degree	0.557	0.188	0.047	0.334	-0.202	0.141	0.068
Asian Race	0.418	0.092	0.225	0.021	0.286	0.29	0.054
Hispanic Race	0.182	**0.931**	-0.056	0.049	-0.13	0.101	-0.01
Other Race	0.1	**0.925**	-0.068	0.05	-0.149	0.083	-0.012
Native American Race	0.209	0.752	0.054	0.106	-0.066	0.107	0.047
Hawaiian and Pacific Islander Race	0.107	0.283	0.169	0.118	0.14	0.204	-0.027
Female 65 and Older Living Alone	0.245	-0.033	0.876	-0.034	0.033	-0.022	0.079
Male 65 and Older Living Alone	0.171	0.038	**0.815**	0.064	0.023	0.052	0.009
Female 65 and Older	0.453	-0.059	**0.745**	-0.059	0.08	-0.112	0.367
Poverty Age 65 and Older	0.027	0.058	**0.704**	0.404	-0.12	0.117	0.028
Male Age 65 and Older	0.532	-0.045	0.672	-0.08	0.176	-0.14	0.263
Population Living Below Poverty	0.119	0.345	0.174	**0.822**	-0.194	0.143	-0.006
Black Race	0.137	-0.308	-0.016	**0.797**	-0.236	-0.022	0.047
Poverty Age Under 5	0.05	0.345	-0.006	**0.784**	-0.195	0.055	-0.023
Female No High School Degree	0.459	0.105	0.161	0.491	-0.22	0.101	0.052
Median Family Income 1999	-0.003	-0.099	0.029	-0.225	**0.89**	-0.196	0.03
Per Capita Income 1999	-0.08	-0.116	0.13	-0.155	**0.884**	-0.045	0.008
Median Household Income	0.041	-0.078	-0.117	-0.218	**0.871**	-0.256	0.014
Normalized Difference Built-up Index	-0.07	0.131	0.085	0.037	-0.225	**0.886**	0.021
Normalized Difference Vegetation Index	0.125	-0.219	-0.081	-0.029	0.244	**-0.856**	0.006
Temperature	0.095	-0.016	-0.113	0.047	-0.02	0.484	0
Male 65 and Older in Group Housing	0.11	0.008	0.114	0.044	0.013	0.037	0.946
Female 65 and Older in Group Housing	0.095	0.003	0.153	-0.003	0.025	-0.012	0.945

Table 1-3: Block Resolution PCA results with interpretation values

Component	Eigenvalue	% Variance	
1	10.29	35.481	
2	3.268	11.267	
Block Analysis 3	2.251	8.108	
With Interpretation value for 4	1.857	6.405	
MHI, MFI, PCI 5	1.676	5.779	
6	1.404	4.843	
7	1.103	3.804	
8	1.09	3.759	
9	1.04	3.587	**Cumulative = 83.032%**

	Component								
	1	2	3	4	5	6	7	8	9
Female 65 and Older Living Alone	**0.887**	0.093	0.038	0.142	0.093	0.018	0.011	0.018	0.065
Male 65 and Older Living Alone	**0.813**	0.155	0.063	0.147	0.098	-0.028	-0.009	0.042	0.054
Male Age 65 and Older	**0.748**	0.047	0.085	0.254	0.165	0.37	0.091	-0.001	0.034
Female 65 and Older	**0.742**	0.045	0.043	0.179	0.118	0.488	0.055	0.012	0.011
Poverty Age 65 and Older	**0.693**	0.445	0.025	-0.009	0.089	-0.05	-0.02	0.049	0.078
Female High School Degree	**0.552**	0.287	0.296	0.277	0.406	0.101	0.135	0.03	0.13
White Race	0.511	-0.108	0.433	0.469	0.239	0.071	0.099	-0.007	0.229
Poverty Age Under 5	0.048	**0.886**	0.249	0.002	0.066	0.028	0.011	0.034	0.01
Population Living Below Poverty	0.244	**0.871**	0.264	0.051	0.148	0.012	0.003	0.059	0.075
Black Race	0.091	**0.668**	-0.216	0.06	0.612	0.024	0.018	0.016	-0.041
Female Under 5 years of Age	0.159	**0.605**	0.543	0.243	0.245	-0.026	0.081	0.001	0.112
Male Under 5 years of age	0.158	0.6	0.549	0.244	0.245	-0.031	0.082	0	0.117
Female No High School Degree	0.351	0.594	0.22	0.109	0.376	0.055	0.051	0.061	0.053
Hispanic Race	0.037	0.189	**0.919**	0.075	0.142	0.005	0.006	0.061	0.031
Other Race	0.004	0.191	**0.9**	0.045	0.075	0.01	0.004	0.061	-0.001
Native American Race	0.096	0.099	0.57	0.044	0.084	0.001	-0.009	0.018	0.077
Median Household Income	0.121	0.098	0.126	**0.937**	0.125	0.043	0.095	-0.022	0.057
Median Family Income 1999	0.208	0.076	0.1	**0.911**	0.105	0.041	0.086	-0.006	0.062
Per Capita Income 1999	0.21	0.072	0.026	**0.91**	0.035	0.017	0.01	0.014	0.039
Male No High School Degree	0.157	0.263	0.24	0.069	**0.824**	0.034	0.004	0.039	0.035
Male High School Degree	0.282	0.191	0.314	0.183	**0.801**	0.058	0.07	0.032	0.094
Total Population	0.42	0.354	0.409	0.368	0.531	0.055	0.069	0.028	0.228
Female 65 and Older in Group Housing	0.128	0.005	0	0.027	0.003	**0.941**	-0.005	0.006	0.007
Male 65 and Older in Group Housing	0.095	0.019	-0.003	0.023	0.053	**0.933**	-0.01	0.013	0.012
Temperature	0.093	0.085	0.076	0.137	0.079	0.002	**0.947**	0.175	0.01
Normalized Difference Vegetation Index	-0.006	-0.023	-0.077	0.069	-0.007	-0.009	**0.785**	-0.584	-0.013
Normalized Difference Built-up Index	0.072	0.081	0.094	0.001	0.055	0.018	0.02	0.978	0.036
Hawaiian and Pacific Islander Race	-0.016	0.083	0.04	0.005	-0.008	0.033	0.013	0.005	**0.861**
Asian Race	0.297	0.031	0.143	0.192	0.15	-0.023	-0.019	0.044	0.638

Table 1-4: Block Resolution PCA results without the Interpretation values

	Component	Eigenvalue	% of Variance	Cumulative %
	1	9.675	33.363	
	2	3.5	12.07	
Block Analysis	3	2.608	8.992	
Without Interpretation value for	4	1.865	6.429	
MHI, MFI, PCI	5	1.669	5.757	
	6	1.459	5.031	
	7	1.128	3.889	
	8	1.046	3.608	
	9	1.03	3.551	**Cumulative = 82.69%**

	Component								
	1	2	3	4	5	6	7	8	9
Female 65 and Older Living Alone	**0.896**	0.034	0.102	-0.005	0.088	0.005	0.005	0.056	0.017
Male 65 and Older Living Alone	**0.822**	0.063	0.165	-0.023	0.093	-0.038	-0.016	0.046	0.035
Male Age 65 and Older	**0.789**	0.096	0.04	0.018	0.187	0.357	0.103	0.038	0.001
Female 65 and Older	**0.771**	0.046	0.04	-0.011	0.125	0.475	0.061	0.01	0.007
Poverty Age 65 and Older	0.67	0.014	0.463	-0.105	0.054	-0.055	-0.044	0.06	0.028
Female High School Degree	0.599	0.312	0.247	-0.088	0.444	0.089	0.165	0.139	0.03
White Race	0.593	0.469	-0.127	0.182	0.322	0.059	0.135	0.245	0.049
Hispanic Race	0.048	**0.919**	0.16	-0.091	0.147	0.005	0.007	0.026	0.053
Other Race	0.009	**0.898**	0.164	-0.103	0.073	0.011	0.001	-0.005	0.045
Male Under 5 years of age	0.199	0.584	0.564	-0.033	0.302	-0.034	0.11	0.129	0.02
Female Under 5 years of Age	0.201	0.578	0.569	-0.033	0.301	-0.029	0.109	0.124	0.021
Native American Race	0.099	0.567	0.093	-0.048	0.074	0.003	-0.019	0.07	0.007
Poverty Age Under 5	0.043	0.27	**0.868**	-0.118	0.083	0.03	0.018	0.011	0.033
Population Living Below Poverty	0.244	0.283	**0.854**	-0.136	0.162	0.011	0.008	0.074	0.053
Black Race	0.096	-0.205	0.657	-0.143	0.611	0.027	0.02	-0.039	-0.012
Female No High School Degree	0.363	0.229	0.57	-0.146	0.383	0.051	0.059	0.053	0.043
Median Family Income 1999	-0.01	-0.08	-0.127	**0.947**	-0.066	0.003	0.065	0.001	-0.084
Median Household Income	-0.082	-0.059	-0.112	**0.935**	-0.049	0.008	0.078	-0.004	-0.117
Per Capita Income 1999	0.04	-0.092	-0.078	**0.92**	-0.073	-0.008	-0.031	0.004	-0.033
Male High School Degree	0.309	0.31	0.166	-0.086	**0.812**	0.053	0.076	0.088	0.028
Male No High School Degree	0.159	0.225	0.257	-0.097	**0.808**	0.035	-0.01	0.019	0.026
Total Population	0.48	0.439	0.328	0.018	0.588	0.049	0.095	0.237	0.05
Female 65 and Older in Group Housing	0.145	0	0.004	0.004	0.005	**0.94**	-0.005	0.006	0.007
Male 65 and Older in Group Housing	0.11	-0.003	0.019	0	0.053	**0.932**	-0.011	0.01	0.014
Temperature	0.114	0.078	0.084	0.001	0.087	0.001	**0.95**	0.011	0.172
Normalized Difference Vegetation Index	0.003	-0.069	-0.004	0.196	0.003	-0.01	0.778	-0.015	-0.563
Hawaiian and Pacific Islander Race	-0.011	0.042	0.084	-0.044	-0.022	0.035	0.002	**0.862**	-0.03
Asian Race	0.326	0.152	0.024	0.065	0.191	-0.03	-0.002	0.635	**0.091**
Normalized Difference Built-up Index	0.069	0.078	0.073	-0.177	0.046	0.02	0.022	0.032	**0.964**

Comparing the quantity of observed mortalities within the EHVI classification groups, distinguished through the quantile distribution of the PCA's sum factoral, can help demonstrate whether the methods produced viable results. Table 1-5 and Table 1-6 demonstrate a higher quantity of observed mortalities in the higher EHVI risk categories, suggesting these results, and those from previous EHVI studies conducted by Johnson et al. (2012), are able to identify vulnerability trends. This is particularly noticeable at the block resolution using the interpreted financial variables. The block group EHVI analysis depicts a less obvious trend on declining

mortalities between high and low risk areas. Results demonstrated in Table 1-5 suggests the population data alone may provide a strong indicator of risk, when compared to the block group resolution. The highest risk categories in both the block group and block analysis contain a higher quantity of people than the lower risk categories, suggesting a relationship between EHVI and population density which should continue to be studied. The ROC AUC results (Table 1-7) demonstrate all three models perform better than statistical random chance, but the block group resolution had the lowest value at 0.59 which is barely better than a statistical coin flip. An AUC value above .70 can be considered a fair model (Tape). With the block non-interpreted financial variables model at 0.72 and the interpreted variables at 0.82, the superiority of the block resolution is demonstrated.

Table 1-5: Expected and Observed Mortality data based on EHVI category Population estimates and the Observed Mortalities within EHVI categories

BLOCK Interpreted Financial Variables Mortality Assessment

	Population	Expected	EHVI
BlkHH	2770964	302.00	342
BlkHL	1072006	116.84	130
BlkMH	642508	70.03	60
BlkML	490709	53.48	30
BlkLH	349120	38.05	16
BlkLL	51434	5.61	8
Total	5,376,741	586	586

BLOCK GROUP Mortality Assessment

	Population	Expected	EHVI
BlkGrpHH	1627420	177.37	149
BlkGrpHL	1036790	113.00	125
BLkGrpMH	863513	94.11	97
BlkGrpML	729513	79.51	69
BlkGrpLH	640194	69.77	83
BlkGrpLL	479311	52.24	63
Total	5,376,741	586	586

Table 1-6: Expected and Observed Mortality data based on EHVI category Population estimates and the Observed Mortalities within EHVI categories without the use of interpreted financial variables in the block resolution

	BLOCK Non-Interpreted Financial Variable Mortality Assessment		
	Population	Expected	EHVI
BlkHH	2496223	272.06	260
BlkHL	1071451	116.78	123
BlkMH	754791	82.26	83
BlkML	586078	63.88	60
BlkLH	399054	43.49	47
BlkLL	69144	7.54	13
Total	5,376,741	586	586

	BLOCK GROUP Mortality Assessment		
	Population	Expected	EHVI
BlkGrpHH	1627420	177.37	149
BlkGrpHL	1036790	113.00	125
BLkGrpMH	863513	94.11	97
BlkGrpML	729513	79.51	69
BlkGrpLH	640194	69.77	83
BlkGrpLL	479311	52.24	63
	5,376,741	586.00	586

Table 1-7: Area Under the Curve results for the three models

	AUC
Block Group	0.59
Block Interpreted variables	0.80
Block Non-Interpreted Variables	0.72

The assessment of the Modifiable Aerial Unit Problem (MAUP) suggests the presence of an aggregation differences between the block group and block resolution analysis (Kwan, 2012a, 2012b). Table 1-8 and Table 1-9 demonstrate the majority of the block risk categories were distributed within the general risk level of the block group (general identified as High, Medium, or Low vs. the sub-categories High-High and High-Low) for several categories. The sub-categories of the High-High (HH) block group risk, however, contained a large quantity of as Low-Low block boundaries for both the interpretation and non-interpretation assessments. This demonstrated a high level of risk variability is present even within block groups. It is of note that the block categories without the use of the financial interpretation values appear to be more consistent to the block group results. This suggests the data and results may have been impacted by resolution. The non-interpreted financial variables for the block assessment leave several neighboring block features with the same value, simulating low resolution data similar to the environmental variables. This reduces the difference between the two resolutions, and

implies a MAUP within the block non-interpreted assessment. Both the block and block group resolutions appear to provide an opportunity to improve mitigation practices over the current weather methods. However, the block resolution better identified population vulnerable spaces, and could improve resource allocation or identification of risk levels.

Table 1-8: EHVI block with interpretation values results spatial location with the EHVI parent block group boundary. Chart numbers represent the quantity of block EHVI designation feature centroids occur within each block group category. Red identifies the highest quantities in the block group boundary, while Blue represents the fewest

Block	BlkGrpHH	BlkGrpHL	BLkGrpMH	BlkGrpML	BlkGrpLH	BlkGrpLL
BlkHH	3659	2699	1977	1210	779	376
BlkHL	1663	1846	2089	2259	1755	1088
BlkMH	1736	1597	1668	1999	2082	1618
BlkML	1677	1498	1599	1797	2053	2076
BlkLH	1999	1620	1464	1453	1772	2392
BlkLL	2725	1873	1164	1344	1292	2303

Block group

Table 1-9: EHVI block without interpretation values results spatial location with the EHVI block group boundary. Red identifies the highest quantities in the block group boundary, while Blue represents the fewest

Block	BlkGrpHH	BlkGrpHL	BLkGrpMH	BlkGrpML	BlkGrpLH	BlkGrpLL
BlkHH	4983	2942	1496	759	396	124
BlkHL	1898	2462	2507	2058	1338	437
BlkMH	1515	1533	1973	2492	2206	981
BlkML	1390	1380	1557	1953	2402	2018
BlkLH	1658	1196	1215	1349	1962	3320
BlkLL	2015	1620	1213	1451	1429	2973

Block Group

The difference aesthetic or vulnerable visualization capabilities between the two resolutions is demonstrated by Figure 1-1 and Figure 1-2. The block resolution analysis provides a map that is visually much cleaner and easier to read, as it removes many of the high industrial, heavy transportation routes, and park spaces by labeling them as low risk areas. This allows for a better indication of where mitigation resources could be distributed within populated space. This ability was not available at the larger block group or census tract resolutions studied by the

EHVI, as areas without residential space are aggregated into the larger boundaries, and remain hidden in the map (Daniel P Johnson et al., 2012; Stanforth, 2011). The block resolution therefore provided the ability to efficiently remove areas of industry and infrastructure from the risk map, and could better guide the allocation of resources during inclement conditions. The difference between the interpreted financial variables and non-interpreted values at the block resolution, demonstrated by Figure 1-2, are minimal in the downtown urban Chicago area where this methodology is primarily focused. Additionally, a population distribution map was paired with the EHVI model at the block resolution. Both maps in Figure 1-3 identify unpopulated space, but the EHVI appears to provide a better visual of the discontinuous vulnerability rates in the downtown Chicago area than the simpler population map. This could be advantageous to public health officials as they work to distribute aid during inclement conditions.

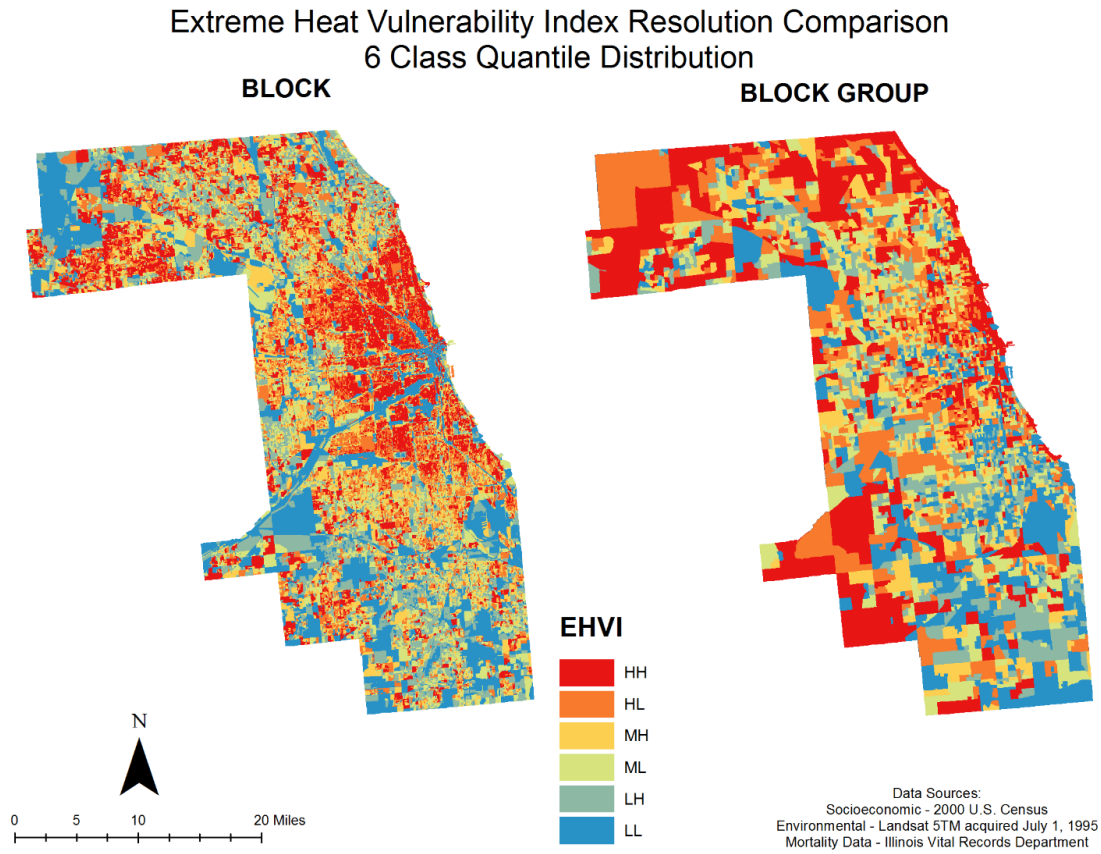

Figure 1-1 Block and block group EHVI Risk map

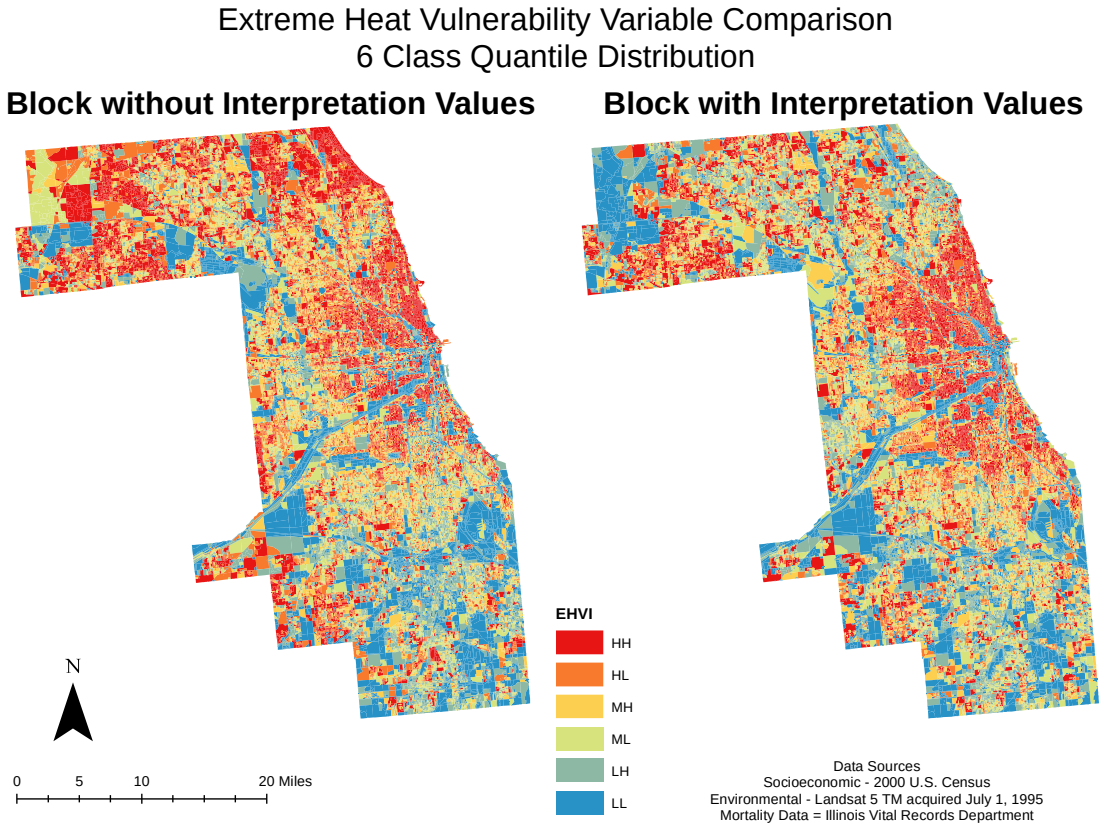

Figure 1-2: Comparison between the interpreted and non-interpreted financial variable block models

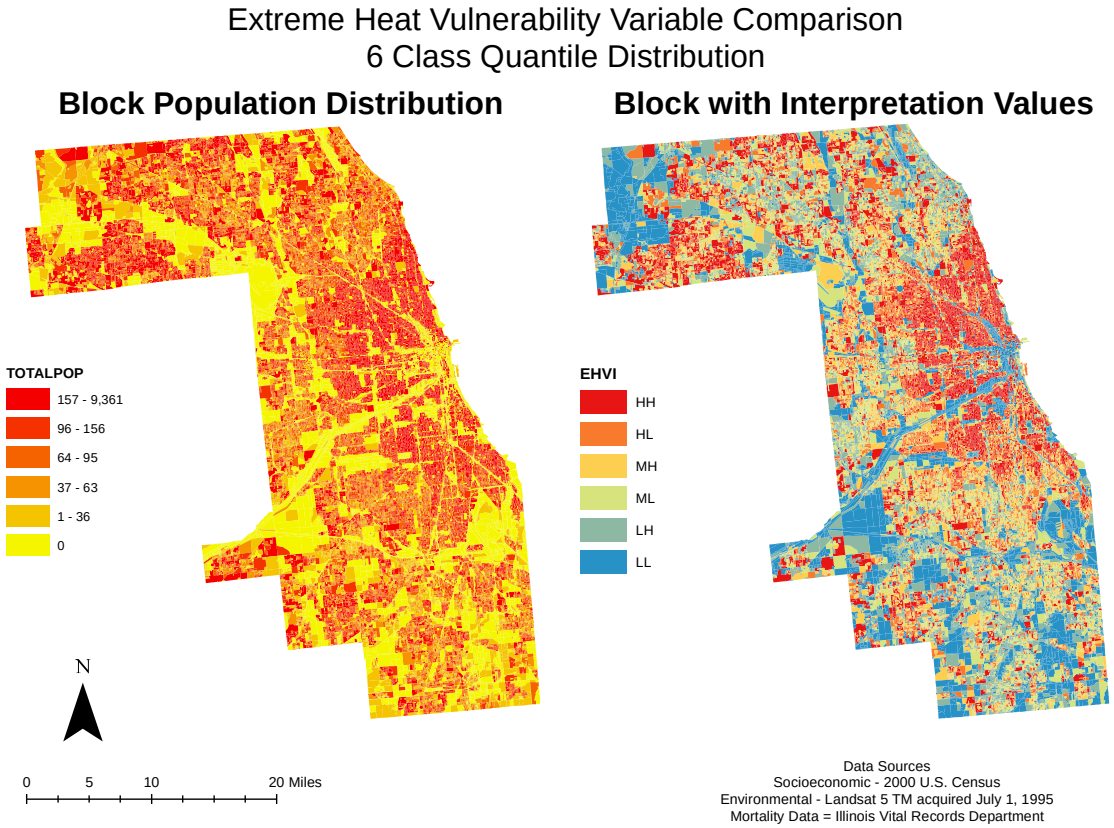

Figure 1-3: Comparison of the Population and EHVI distribution at the block resolution

1.5 Discussion

The EHVI analysis suggests the block resolution is more adept at identifying areas of heat vulnerability. This is demonstrated by the interpolated variable block model's improved documentation of more observed mortality rates in high risk areas than would be calculated by population alone, and it contained the strongest AUC results. The finer spatial boundary also does not appear to be as impacted by aggregation to the same extent as the block group analysis, except potentially by the lower resolution physical environment variables, demonstrated by the identification of low risk block units within the high risk block groups in Table 1-5. The finer spatial model could allow for improved mitigation strategies, or prioritized distribution of resources, during mitigation actions. It should be noted that the block resolution required additional preparation, due to the need to interpolate some variables, which could

reduce its usefulness for less experienced or time-constrained analysts. Although these methodologies could be reproduced at either resolution within the United States, comparable studies may not be as efficient in other countries with lower quality population records.

The PCA results are similar to what has been demonstrated in previous studies. The 1990 Census data used in the Johnson et al. (2012) paper had a comparable explanation of variance, and documented additional components within the smaller boundary model when comparing the census tract vs. block group resolutions (Daniel P Johnson et al., 2012; Stanforth, 2011). The similarity of the results between studies confirms the procedures were performed accurately (Daniel P Johnson et al., 2014). The results also support previous variable relationships to heat vulnerability, and are similar to those identified during this study are similar to those identified during the 2012 EHVI assessment using 1990 data, using the census tract and block group resolutions (Daniel P Johnson et al., 2012; Stanforth, 2011). Age, educational attainment, and financial stability continued to identify patterns of vulnerability regardless of the changes in spatial components of this study, or when using variables from different time periods (S. L. Cutter et al., 2008; S. L. Harlan et al., 2006; Daniel Johnson et al., 2009; Daniel P Johnson et al., 2014; D. P. Johnson & Wilson, 2009). Since these variables have consistently shown up in a variety of vulnerability publications, regardless of the study's spatial resolution, it demonstrates their superiority in heat vulnerability modeling and supports their inclusion in future studies (Changnon et al., 1996; Susan L. Cutter et al., 2003; Davis, 1997; S. L. Harlan et al., 2006; Daniel Johnson et al., 2009; Naughton et al., 2002; C. E. Reid et al., 2009; Whitman et al., 1997). The results also demonstrate the variables' ability to be used through time, as they have been consistently identified during multiple studies using different Census collection periods (Daniel P Johnson et al., 2012; Stanforth, 2011), a variety of study areas (Daniel P Johnson et al., 2014), and methodologies (S. L. Cutter et al., 2008; Susan L. Cutter et

al., 2003; Ebi et al., 2003; S. L. Harlan et al., 2006; Hattis, Ogneva-Himmelberger, & Ratick, 2012; Hondula et al., 2012; Luber & McGeehin, 2008).

Race variables were included in the EHVI studies for comparative purposes with previous studies (Susan L. Cutter et al., 2003; Davis, 1997; O'Neill et al., 2005). Some race categories loaded higher in the PCA component results during this and previous studies, but were not explored extensively in the discussion. Previous research indicated race variables were highly correlated with other variables, and current results support that suggestion (Stanforth, 2011). This may indicate specific race demographics are associated with more influential socioeconomic variables, such as financial status or educational attainment, rather than representing a genetic predisposition to heat stress.

Environmental variable results were also comparable to previous studies, by loading in less prominent components (Daniel P Johnson et al., 2012; Daniel P Johnson et al., 2014; Stanforth, 2011). The block analysis had the NDVI, NDBI, and temperature all in a lower component (component 7) than even the block group (component 6) analysis. The consistently lower results is potentially due to a bias caused by the difference in size between the large remote sensing pixels and smaller political boundaries (Daniel P Johnson et al., 2012; Stanforth, 2011). Due to the lower resolution of the satellite imagery, the environmental variables cover multiple neighboring boundaries with a single value. This limitation would be greatest in the smaller block resolution, demonstrated by the weaker component loading in that study. Previous studies have demonstrated model superiority when the environmental variables are included vs. when they are removed, so continued exploration of these variables is encouraged (Dan Johnson et al., 2011; Daniel Johnson et al., 2009; D. P. Johnson & Wilson, 2009). Future improvements in data acquisition, such as through the use of aerial or UAV data collection, could improve the application of these variables.

The variables used in this study did have limitations which should be addressed. As previously stated, the U.S. Census did not collect all variables at the block resolution. Basic demographic data, age and race, were available at the finer resolution, but financial and educational data were only available at the block group or larger boundaries (U.S.Census, 2000). Therefore, the missing information used in the block analysis was interpreted through the proportion of the block's population to its parent block group. This may have caused some confounding errors as the value is based off a percent of the total population rather than the recorded demographic values. This procedure is similar to the process used by the American Community Survey (ACS), the successive U.S. Census product which began in 2010. The ACS is a survey based data collection process, whereby a portion of the population is surveyed and the results are extrapolated to estimate the demographics of the whole population. This new method could complicate future population based assessments that require spatially-detailed socioeconomic information, but it is the best dataset currently available for public use (U.S.Census, 2015). An unforeseen characteristic of the block resolution was found in the documentation from the U.S. Census Bureau, which indicated several political census boundaries were drawn where populations did not reside. Since no population resides in these areas, it suggests the block resolution analysis could remove these features from mitigation plans, reducing the area under surveillance, and improve mitigation direction.

Additionally, the residential mask used to quantify environmental variable impact on the local population was not present in all of the political boundaries. Therefore, the full political boundary was used to quantify the environmental variables for those features. It is unclear how this may have impacted the results, but would be comparable to previous studies which did not incorporate residential space in their design. It also suggests additional resources for identifying residential space, such as building codes, need to be considered in future projects. This

limitation further supports the need for improved resolution data for both environmental variables and socioeconomics, to better address their local or micro-relationship.

Additionally, it should be noted that this study, and its predecessors, used data from a single study site and event. Although these studies incorporated data obtained from multiple Census datasets, the results may not be directly attributable to other study locations. The consistent prevalence of specific socioeconomic and environmental variables through multiple studies demonstrate their usefulness in vulnerability studies (S. L. Cutter et al., 2010; Dolney & Sheridan, 2006; Sharon L Harlan, Declet-Barreto, Stefanov, & Petitti, 2013; Hondula et al., 2012; Daniel P Johnson et al., 2012; Schwartz, 2005). Application of these variables and results are expected to fluctuate based on the local community attributes and time to which they are applied (Daniel P Johnson et al., 2014; Patz et al., 2000; Stanforth, 2011).

1.6 Conclusion

This study was designed as an assessment of the EHVI, a model of vulnerability to extreme heat, using both socioeconomic and environmental data. The model was able to appropriately identify areas of increased vulnerability, and the variables which contribute to those higher values (Daniel P Johnson et al., 2012; Daniel P Johnson et al., 2014; Stanforth, 2011). Age, education, and financial status continue to be identified as strong indicators of vulnerability, supporting previous work. Satellite derived environmental variables have been identified as useful for modeling heat vulnerability, but are not readily available at a resolution capable of making them load within the top principal components using the parameters of this study.

This type of analysis is not only important for academics, but for the broader emergency management and health community. Mitigation practices could be implemented to reduce risk during oppressive events through intelligence lead decision making, such as is provided by the

EHVI. This type of modeling could help management better identify mitigation methods which would best aid their community. This style of analysis could be used to identify areas of high risk in the Midwest which could benefit from a reduction in UHI through prioritized tree canopy planting. This practice would be less efficient in the desert climate of the American Southwest without the allocation of additional water resources to sustain the vegetation. Similarly, adding education or job opportunities to areas of high risk may not assist a community of retirement aged residents. These examples demonstrate the need for local vulnerability modeling, where mitigation practices could be tailored to meet the needs of the community.

Chapter 2 Exploratory Analysis of Dengue Fever Niche Variables within the Río Magdalena Watershed

2.1 Abstract

Previous research on Dengue Fever have involved laboratory tests or study areas with less diverse temperature and elevation ranges than is found in Colombia, therefore research was needed to identify local attributes of Dengue Fever transmission. Principal component analysis (PCA) was conducted on variables derived from Geographical Information Science (GIS) and Remote Sensing (RS) platforms for use in a preliminary Dengue Fever study to identify environmental and socioeconomic variables to incorporate in future Dengue Fever outbreak modelling projects in the Río Magdalena watershed in Colombia. Environmental variables derived from the Moderate Resolution Imaging Spectroradiometer (MODIS), the Shuttle Radar Topographic Mission (SRTM), and Tropical Rainfall Measuring Mission (TRMM) platforms were combined with population variables to be statistically compared against reported cases of Dengue Fever from the watershed. Three factor analysis models were investigated to analyze variable patterns, including population, population density, and Empirical Bayesian Estimation models. Results identified varying levels of Dengue Fever transmission risk, and environmental characteristics which support, and advance, the research literature. Multiple temperature metrics, elevation, and vegetation composition were among the more contributory variables found to identify future potential outbreak locations.

2.2 Introduction

The resurgence of tropical diseases in the Americas, including Dengue Fever or Zika, have received increased interest due to their advancing transmission rates and zone of infection. The resurgence of these diseases is partially due to the suspension of vector eradication programs throughout the Americans, such as those conducted by the Pan American Health Organization (PAHO) programs until the 1970s (Gubler, 1998). The prohibitive costs of *in*

situ field data collection to monitor such vectors, and their expanding habitats, has justified the use of remote sensing (RS) and Geographic Information Science (GIS) applications to estimate disease spread and population vulnerability (Ceccato et al., 2005; Moreno-Madriñán et al., 2014). Public health agencies can use such tools to implement monitoring and surveillance programs to estimate vector abundance, implement prevention strategies, and control further dispersion of the disease or vector (Cleckner & Allen, 2014). However, local metrics and research are still needed to identify site specific attributes of vector dispersion before predictive or forewarning systems can be cost-effectively implemented (Beebe et al., 2009; Liu et al., 2008; Pinto, Coelho, Oliver, & Massad, 2011; S. Zhang & Zhao, 2015).

One of the more commonly implemented remote sensing tools for vector research has been the Moderate Resolution Imaging Spectroradiometer (MODIS). MODIS is a dual satellite, Aqua and Terra, platform of NASA's Earth Observation System with a sun synchronous orbit, making it capable of collecting data over the entire earth multiple times a day, 10:30AM and 10:30PM for Terra, and 1:30PM and 1:30AM for Aqua (S. L. Harlan et al., 2006). The high temporal resolution reduces the sensor's spatial capabilities, 250m for multispectral data and 1km for land surface temperature (LST), but its continuous ability to collect a variety of spectral wavelengths has proven useful for a wide variety of environmental assessments (NASA, 2016a). The MODIS satellite is currently being used for the world famine prediction index, a process which could also incorporate vector warning products (Ceccato et al., 2005; Fuller et al., 2009). NASA scientists have developed such a Malaria transmission model though the Goddard Space Flight Center's Health Planet program, which incorporates input variables for parasites, hosts, vectors, and environmental and human factors (Cleckner & Allen, 2014).

The *Aedes aegypti* mosquito vector is capable of spreading a variety of diseases, including: Yellow Fever, Chikungunya, Zika, and Dengue Fever. Both cultural and climatic

conditions are considered to influence the transmission rates of these diseases, through this vector. Although commonly associated with tropic and subtropical climates, due to the habitable zone of the mosquito, changing climate may expand the dispersion potential of these diseases (CDC; Gubler, 1998; Lozano-Fuentes et al., 2012; Ruiz-López et al., 2016; WHO, 2016). This demonstrates why improved methods to identify habitable zones of this type of vector will become increasingly important to improve future disease mitigation and prevention practices.

The onset of Dengue Fever includes symptomatic fever, rash, and orthopedic discomfort. Severe cases, known as Hemorrhagic Fever, can lead to bleeding, blood pressure complications, and even death (CDC, 2016; WHO, 2016). Since no vaccine is currently available, mosquito control through proper hygiene, by use of repellant and screen/netting, are critical practices (Gubler, 1998). It is also important to identify geographical areas to implement mosquito control/preventative practices to decrease contact between the vector and people. Previous research has identified flower pots, water storage containers, impervious surface catchments, and other anthropogenic water sources as opportunities for *Aedes aegypti* larval development (Beebe et al., 2009; Gubler, 1998). These features are common to domesticated or developed areas, perpetuating the cause for the *Aedes aegypti* mosquito to be found in urban/populated environments, increasing the opportunity for mature specimens to feed on human populations, and disperse the virus (Beebe et al., 2009; Cleckner & Allen, 2014; Delmelle et al., 2014).

The vector's habitat and disease transmission has also been found to contain meteorological and elevation parameters (Brunkard et al., 2008; Moreno-Madriñán et al., 2014). These variables can be estimated through RS data to derive the geographic extent of the mosquito's habitat, and could become increasingly important to accurately monitor future transmission risk of Dengue Fever. Similar habitat and dispersion studies are already being

conducted for other vector diseases, including West Nile Virus and Malaria (Ceccato et al., 2005; Cleckner & Allen, 2014; Liu et al., 2008). This chapter's primary focus is the Dengue Fever virus, although some of the incorporated variables and discussion directly address the mosquito vector. This is in reference to the vector's ability, or limitations, to disperse the disease and will thus be assumed to be synonymous to the discussion of the disease's dispersion.

A study of Dengue Fever in Singapore described an increase of 22-184% in reported cases given a 2-10°C rise in maximum or minimum temperature (Pinto et al., 2011). Research has demonstrated the vector is impacted by lower temperatures (Miron, Montero, Criado-Alvarez, Diaz, & Linares, 2010). Under lab conditions, Brady et al. (2013) reported low survival of the mosquito vector below 14-15°C. These temperature limits may also impact the latitude and altitude of the mosquito's survivability (Brunkard et al., 2008). In Colombia, the mosquito is normally not abundant above 1,800 meters elevation (Kanaroglou & Delmelle, 2015). Several studies have documented a lower limit of 1,700 meters above sea level in Central Mexico (Lozano-Fuentes et al., 2012; Moreno-Madriñán et al., 2014). Although these levels may be impacted by climate change of land use, as a recent study has identified habitable space at higher elevations (Ruiz-López et al., 2016). Additional habitat parameters can involve the presence of anthropogenic water storing activities, which may be responsible or up to 79% of the reported cases (Moreno-Madriñán et al., 2014; Pinto et al., 2011). Water storage tanks, potted plants, and similar peridomestic related containers can provide larval rearing sites proximal to populated areas. Combining the habitable zone of the mosquito and propensity for anthropogenic water container use, the locations of increased mosquito populations and disease transmission can be extrapolated.

Exploratory factor analysis has previously been used to investigate the relationship between a dependent variable and potential contributing, independent variables (Stanforth,

2011; StatSoft, 2015). Principal component analysis (PCA) is a factor analysis method used to identify trends within a multivariate dataset to reduce or derive new, efficient, component variables (DeCoster, 2004). Independent variables are plotted against a dependent variable for the PCA process to identify linear combinations, or clusters, within the dataset. The PCA identifies a new axis origin within the core of the variable scatter plot and realigns the axes so the majority of the data's variance exists along the new x/y plane. The new x and y axes will identify the majority of the variance and create new composite components 1 and 2, respectfully. The new composite variables incorporate the relationship of the original variables, but the realignment requires fewer composite variables to explain the majority of the dataset variability; these can be known as principal components. The amount of the original dataset variability a component variable encompasses is represented by its eigenvalue; the larger the eigenvalues the stronger the composite component. Following the Kaiser Criteria, composite components with an eigenvalue of one (1) or greater remain in the analysis as principal components since they provide at least as much information as one of the original variables (Kaiser, 1960; StatSoft, 2015). Components which do not reach the Kaiser Criteria can be considered 'noise' and not implemented in future stages, therefore fewer variables are used to describe the data, simplifying the processing (idre, 2016; Kaiser, 1960). These methods have been shown to reduce the amount of data used and identify trends in the dataset (DeCoster, 2004; Pinto et al., 2011). The results demonstrate which of the original input variables are better related to the dependent variable based on how they load into the composite variables.

This project was designed as an exploratory analysis to identify variables which may explain the annual incidence of Dengue Fever within Colombia's Magdalena River watershed during the 2012-2014 study period. This study site is characterized by climate conditions ranging from tropical sea level to permanent mountain cap snow, at the top of the Andes, and from

dessert to tropical rainforest. The wide range of climate conditions and relatively consistent cultural populations make this an ideal site for this type of exploratory analysis. The borders contain the upper and lower climate limits of the *Aedes aegypti* vector, providing an isolated study area. This analysis was designed to investigate whether the study site supports previous habitat and transmission information obtained about the vector during previous laboratory and environmental studies in other geographical areas. The intended use of the PCA results is to guide the development of supplemental research projects on predictive models or early warning systems for the disease and vector.

2.3 Materials and Methods

2.3.1 Study Site

Analysis was conducted in the Río Magdalena watershed in the South American country of Colombia. The approximately 273,000 km^2 area is bounded by a split in the Andes Mountain range on three sides and the Caribbean Sea coast to the north, as can be seen in the site map in Figure 2-1. The 2014 projected population of the watershed, including the capital city of Bogota, was calculated to be over thirty-six million. The coordinates of the watershed's approximate center is 6.08 N, -74.58 W, which is well within the currently documented latitude range of the *Aedes aegypti* mosquito (Pinto et al., 2011). The elevation of the watershed's mountain borders should naturally isolates the *Aedes aegypti* population from external areas, providing a study area with limited external influence (Morin, Comrie, & Ernst, 2013; Venkatesan & Rasgon, 2010). Due to its location on the western side of South America, it is also influenced by El Niño/La Niña seasonal variations. Dengue Fever has been present in this region for some time; reports indicate the disease prevalence will continue to rise and could eventually reach epidemic proportions (Kanaroglou & Delmelle, 2015).

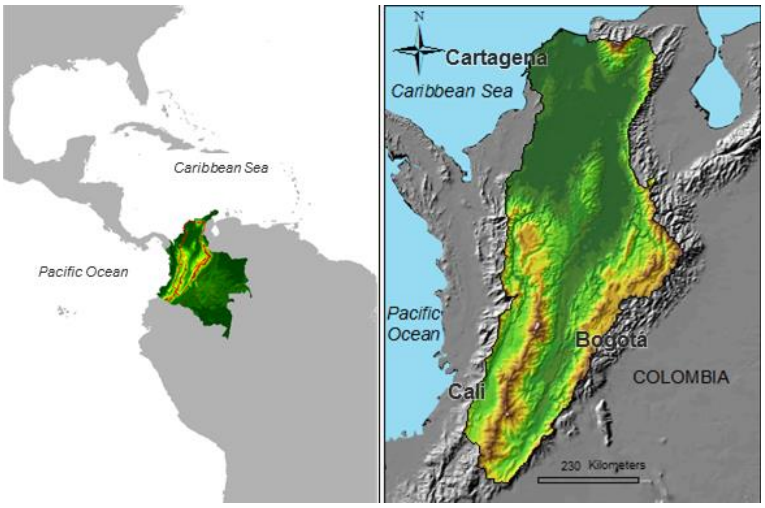

Figure 2-1: Study site identification of the Magdalena Watershed in Colombia, demonstrating its geographical location, elevation diversity, and presence of larger populated urban environments

2.3.2 Population Data

Population variables incorporated into the analysis included socioeconomic data collected during the 2005 General Census administered by El Departamento Administrativo Nacional de Estadística (The National Administrative Department of Statistics). Projected populations for 2012-2014 were created by the data's proprietors, and consisted of tabulated total population, rural and urban populations distinguished by gender (DANE, 2010). Population variables were collected at the municipality resolution, which became the assessment size for this study. Colombian municipality boundaries were distinguished through an open access Esri shapefile, which was clipped to the watershed boundary.

2.3.3 Remote Sensing Data

Environmental variables were obtained from the U.S. Geological Survey's (USGS) (http://e4ftl01.cr.usgs.gov) and NASA (http://disc2.nascom.nasa.gov/) open access database servers (USGS, 2016). Daily MODIS data was collected using an automated MS-DOS batch file operation covering 2012-2014, independently. Due to the high volume of files, a Python 2.7 script was composed for ArcGIS 10.3's Model builder to automatically process the satellite imagery to use a consistent projection file (WGS 1984 UTM Zone 18N), mosaic, and apply

algorithms as the products required. Rather than use the eight day MODIS composite product available through the database, the daily imagery was composited into seven day composite raster files to coincide with the weekly calendar year. The composite data was necessary to reduce the impact of cloud cover on the analysis (NASA, 2016a). All imagery was clipped within the watershed boundary.

ArcMap's raster calculator tool was used to convert MODIS MYD11A1 composited data to LST through additional scripting. Both daytime and nighttime thermal wavelength raster cells were multiplied by a gain factor of 0.02, then converted to Celsius by subtracting 273.15 (Salano, Didan, Jacobson, & Huete, 2010). Both day and night temperatures were included in this study to investigate previously reported low temperature (Brunkard et al., 2008; Moreno-Madriñán et al., 2014) and high/low temperature range (Brady et al., 2013) impact on vector survivability. An additional MODIS product, the MYD09GQ, was used to calculate the two band enhanced vegetation index (EVI) (2.5 * [(NIR-Red) / (NIR+Red+1)]) (Salano et al., 2010). The two band EVI was used due to the MYD09GQ's 250m spatial resolution, the prevalence of high biomass / cloud cover in the study region, the index's ability to differentiate vegetation canopy from background noise, and its proven usefulness in previous Dengue Fever research (Fuller et al., 2009; Jensen, 2005).

Precipitation data was collected through the Tropical Rainfall Measuring Mission (TRMM); a joint venture between NASA and the Japanese Aerospace Exploration Agency launched in 1997 (NASA, 2016b). TRMM's daily temporal resolution has been proven useful for vector research (Pinto et al., 2011). One of TRMM's three onboard sensors, the Visible Infrared Scanner (VIRS), provided the 3B42 V7 dataset, which reports global precipitation estimates at a resolution of 0.25 degree x 0.25 degree (NASA, 2015). Although the TRMM was decommissioned

in 2015, its flight time sufficiently covers the 2012-2014 study period, the product name is still used for the composite precipitation data created from multiple platforms (NASA, 2016b).

Additional annual products for land use land cover (LULC) and elevation were incorporated into the analysis. The pre-processed MODIS MCD12Q1 product identified patterns of LULC at a resolution of 500m. The data identifies major landscape elements, including: barren, cropland, forest, savanna, shrub land, snow and ice, urban, and wetland features (NASA, 2016a). The Shuttle Radar Topography Mission (SRTM) provided elevation data at a spatial resolution of 30m. These resolutions are commonly considered to be sufficient to quantify variations across such a watershed feature.

This project consisted of three stages consisting of data acquisition and analysis, which is graphically represented in Figure 2-2. Stage 1 was a multistage preprocessing, or data preparation, phase. The acquisition and development of the weekly composite satellite data represent the preprocessing part of Stage 1. Composited data were further compounded into yearly values using Esri ArcMap's Cell Statistics tool. Cell statistics calculated an input variable's maximum, minimum, and average value at a pixel's location over the year. The final part of Stage 1 consisted of zonal statistics being applied to the annual composited cell values, using the Colombian municipality boundaries. Variables acquired as yearly variables, elevation and LULC, did not require a cell statistics step, but were still processed through the zonal statistics tool. The processing order demonstrates the naming structure of the variables used throughout the rest of this project. Nomenclature corresponds to the variable type (example LST or VEG), followed by the cell statistic (annual longitudinal: Min, Max, Mean), and concludes with the zonal statistic calculation (Cross-sectional per municipality: Min, Max, Range, Mean, Standard Deviation, Sum). The zonal statistic step also provided a unique identification code per

municipality to improve database organization and allow for the municipality population data to be joining to the environmental variables.

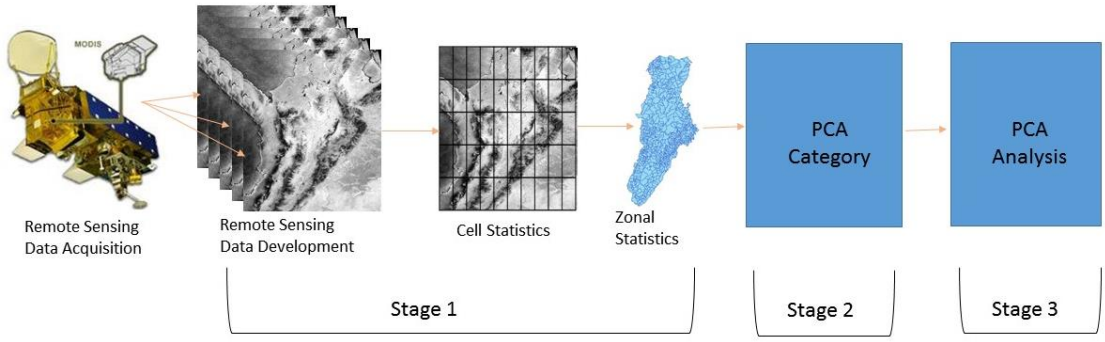

Figure 2-2: Graphic depicting stages of analysis

2.3.4 Health Data

Municipality records of Dengue Fever were provided by the Colombian Instituto National De Salud (National Institute of Health) for the three study years (Salud, 2016). Confirmed cases were identified by the proprietors through laboratory tests of patients exhibiting symptoms of: headache, ocular pain, myalgia, arthralgia, and/or rash. Specific virus specimen antibodies were identified by Polymerase Chain Reaction (PCR) tests for fever patients who had symptoms for less than five days, while an ELISA (enzyme-linked immunosorbent assay) test was used for patients who had symptoms for more than five days (CDC, 2016). The annual municipality records were merged to the shapefile database to be used as the dependent variable in the statistical analysis.

2.3.5 Principal Component Analysis

Following the creation of the complete database (incorporating health data, population demographics, and yearly zonal environmental variables), Stage 2 conducted a PCA independently on each variable category using the IBM SPSS 23 statistical software's Dimension Reduction tool. A category consisted of the zonal statistics derived from a single data source, such as the nighttime LST or precipitation data. The dependent variable was comprised of the

number of confirmed cases of Dengue Fever per municipality. This stage was used to identify the strongest variables from each variable category to include in Stage 3. A variable was considered for Stage 3 if it strongly loaded in component 1 or 2 in both 2012 and 2014. As these years had different El Niño/La Niña seasonal characteristics, they provided assurance the variables were not identified due to a meteorological confounding relationship. Stage 2 was implemented to remove variables which were highly correlated to one another and eliminate noise within the dataset. One variable from component 1 and component 2 was added to Stage 3 to represent the largest diversity of information provided by the category. Multiple variables from a single component are commonly highly correlated to one another, and therefore contribute little more than noise to the results. LULC variables were the exception to this rule, as four variables were included in Stage 3 due to the literature's consistent documentation of their value during this type of analysis (Beebe et al., 2009; Gubler, 1998; Z. Li et al., 2016; Lin & Wen, 2011; Moreno-Madriñán et al., 2014; Mylne et al., 2015). The purpose of Stage 2 was to reduce the amount of input variables, from the original 150 zonal statistic derived variables, to 14 of the strongest variables selected for Stage 3. This step allowed for a more robust relationship to be built for the derived model.

Stage 3 also incorporated a PCA, which identified the relationship between the top variables identified in Stage 2. Three models were analyzed over each of the three years, 2012-2014, identified as: population, population density, and an Empirical Bayesian Estimation (EBE) of Dengue Fever counts. The population model included the yearly projected population per municipality as one of the 14 independent variables, while the population density model used the municipality's projected population divided by its area (population / km^2). The municipalities are not consistent in size, but it was anticipated the higher populated urban areas would increase the density value of the municipality enough to distinguish them from less developed

areas regardless of the municipality's size. Both population models used Dengue Fever cases as the dependent variable.

The EBE variable was developed following the procedures of Brandel (2004) using a Standardized Event Ratio for each municipality and an adjacency matrix to produce an estimate of vector presence without political boundary influence (Brandel, 2004; Woodward, 2013). The adjacency matrix was created through the ArcGIS plug-in 'Adjacency for WinBUGS Tool' (available through the USGS http://www.umesc.usgs.gov/management/dss/adjacency_tool.htm). Since population of the municipality was incorporated into the EBE dependent variable, no population independent variable was required for the model, so only 13 variables were used for this model.

Normalization of all variables was conducted prior to conducting Stage 3 to reduce distribution or range bias in the results. Independent variables were normalized to Z-scores using SPSS's descriptive analysis tool. Similarly, the Poisson distribution of the dependent variable (reported Dengue Fever cases) was transformed through a natural log calculation (Ln [cases+1]). The original Poisson distribution of Dengue Fever cases was, in part, due to the large number of municipalities which reported zero cases in a given year. The EBE dependent variable includes a normalizing step in its development, so no further preparation was required. The Stage 3 PCA was processed by the SPSS dimension reduction tool, using a varimax rotation for an oblique transformation method, and included a regression variable output (Costello & Osborne, 2005). The rotation technique provides a stronger representation of the independent variables by transforming, or rotating, the axis to simplify relationship patterns across multiple dimensions. The regression calculation assesses the relationship an independent variable has to the dependent variable, than calculates an output based on the unique value of the municipality's independent variable. As with previous vulnerability studies, these component

values can be added together to derive a sum factoral value for the municipality to provide a comparison and output for mapping (S. L. Cutter et al., 2008; Susan L. Cutter et al., 2003; S. L. Cutter et al., 2010; S. L. Cutter & Finch, 2008; Daniel P Johnson et al., 2012; O'Neill et al., 2005; Stanforth, 2011).

Comparative assessment of the Stage 3 results was conducted for the three models through a Receiver Operating Characteristic (ROC) Area Under the Curve (AUC). The AUC demonstrates the positive reporting of Dengue Fever cases in high sum factoral areas compared to modeled false positives in areas without reported Dengue Fever cases. This can demonstrate the ability of the PCA regression output to correctly identify areas of increased risk of transmission and provide a comparison between models and years (Kraemer et al., 2015).

2.4 Results

The Stage 2 categorical PCA successfully reduced the list of 150 independent variables to 14 variables which were able to more efficiently explain the variance in the dataset without extraneous variable noise. Among the more prominent variables found from Stage 3 were nighttime LST Minimum cell-Max zone value, Elevation minimum value, Vegetation Min cell-Max zone, and daytime LST Max cell-Mean zone. Four LULC variables were used in the Stage 3 assessment based on Stage 2 results and a priori knowledge from the literature, such as the relationship between the mosquito and populated areas, with limited results (Beebe et al., 2009; Gubler, 1998). A complete list of the Stage 3 independent variables can be viewed, along with their component scores, in Table 2-1; negative numbers in the table indicate an inverse relationship, and the strongest component loading was highlighted for each variable. The variables which load heavier in the earlier components (example 1 or 2) demonstrate a stronger relationship to the dependent variable.

Table 2-1: Model output of principal component analysis

	Population vs. Ln Cases						Population Density vs. Ln cases						EBE				
		Component						Component						Component			
		1	2	3	4	5		1	2	3	4	5		1	2	3	4
2012	Elevation Min	**-0.952**	-0.088	0.071	0.064	-0.004	Nighttime LST Min Max	**0.954**	0.082	-0.038	-0.035	-0.03	Nighttime LST Min Max	**0.954**	0.096	0.014	-0.039
	Nighttime LST Min Max	**0.952**	0.088	-0.073	-0.036	-0.027	Elevation Min	**-0.95**	-0.075	0.046	0.064	0.072	Elevation Min	**-953**	-0.093	-0.013	0.069
	Vegetation Min Max	**0.836**	-0.045	**0.18**	0.201	-0.067	Vegetation Min Max	**0.826**	-0.043	0.198	**0.21**	-**0.11**	Vegetation Min Max	**0.82**	-0.058	0.241	0.207
	Daytime LST Max Mean	**0.619**	-0.267	-0.39	0.468	0.056	Daytime LST Max Mean	**0.625**	-0.279	-0.387	0.463	0.002	Daytime LST Max Mean	**0.654**	-0.258	-0.38	0.436
	Nighttime LST Mean Range	-0.001	**0.838**	**0.18**	-0.344	-0.059	Nighttime LST Mean Range	-0.009	**0.83**	0.164	-0.349	-0.122	Daytime LST Max Range	0.175	**0.841**	-0.09	-0.113
	Daytime LST Max Range	0.169	**0.834**	-0.149	-0.127	0.038	Daytime LST Max Range	0.168	**0.813**	-0.18	-0.141	-0.106	Nighttime LST Mean Range	-0.023	**0.836**	0.247	-0.31
	Vegetation Mean Min	0.085	**-0.65**	0.479	0.262	-0.096	Vegetation Mean Min	0.06	**-0.638**	0.492	0.292	-0.073	Vegetation Mean Min	0.051	**-0.683**	0.449	0.267
	Precipitation Min Max	-0.148	0.556	**0.43**	0.318	0.023	Precipitation Min Max	-0.149	0.595	0.383	0.315	0.037	Precipitation Min Max	-0.187	0.533	0.397	0.361
	Precipitation Sum Min	0.386	-0.143	**0.786**	-0.092	-0.49	Precipitation Sum Min	0.371	-0.1	**0.809**	-0.066	0.052	Precipitation Sum Min	0.316	-0.174	**0.788**	-0.053
	Barren Area	0.171	-0.012	-0.437	0.019	0.004	Barren Area	0.204	-0.012	-0.394	-0.005	0.143	Barren Area	0.208	0.011	-0.406	-0.002
	Crop Area	0.075	-0.196	0.127	**0.769**	-0.054	Crop Area	0.074	-0.174	0.132	**0.776**	0.071	Urban Area	-0.076	-0.054	-0.366	0.008
	Forest Area	-0.044	0.187	0.537	**-0.735**	-0.122	Forest Area	-0.058	0.195	0.573	**-0.718**	-0.041	Crop Area	0.072	-0.224	0.094	**0.767**
	Population	0.011	0.123	0.051	0.002	**0.832**	Urban Area	-0.005	0.009	-0.214	-0.025	**0.772**	Forest Area	-0.095	0.18	0.591	-0.698
	Urban Area	-0.06	-0.089	-0.113	0.009	**0.822**	Population Density	-0.136	-0.116	0.097	0.111	**0.739**					
		1	2	3	4	5		1	2	3	4	5		1	2	3	4
2013	Nighttime LST Min Max	**0.946**	0.073	-0.092	-0.031	-0.098	Nighttime LST Min Max	**0.949**	0.073	-0.09	-0.009	-0.092	Nighttime LST Min Max	**0.953**	0.065	0.009	-0.082
	Elevation Min	**-0.94**	-0.086	-0.107	-0.004	0.127	Elevation Min	**-0.938**	-0.086	0.112	0.029	0.126	Elevation Min	**-0.948**	-0.076	0.011	0.113
	Vegetation Min Max	**0.804**	-0.207	0.008	-0.072	0.215	Vegetation Min Max	**0.802**	-0.211	0.006	-0.095	0.213	Vegetation Min Max	**0.793**	-0.212	0.145	0.216
	Daytime LST Max Range	0.053	**0.852**	-0.034	0.04	0.203	Daytime LST Max Range	0.05	**0.851**	-0.038	-0.006	0.202	Daytime LST Max Range	0.063	**0.842**	-0.08	0.212
	Nighttime LST Mean Range	0.063	**0.821**	0.359	-0.076	0.206	Nighttime LST Mean Range	0.062	**0.818**	0.358	-0.096	0.205	Nighttime LST Mean Range	0.03	**0.842**	0.325	0.19
	Vegetation Mean Min	0.129	**-0.792**	0.077	-0.106	0.112	Vegetation Mean Min	0.126	**-0.798**	0.08	-0.094	0.1	Vegetation Mean Min	0.11	**-0.791**	0.183	0.091
	Crop Area	0.079	-0.511	-0.264	-0.038	0.48	Crop Area	0.076	-0.518	-0.261	-0.037	0.472	Crop Area	0.098	-0.541	-0.163	0.486
	Forest Area	0.03	0.29	**0.871**	-0.1	-0.21	Forest Area	0.033	0.292	**0.871**	-0.088	-0.211	Forest Area	-0.066	0.364	**0.826**	-0.264
	Daytime LST Max Mean	0.432	-0.309	**-0.746**	0.038	0.001	Daytime LST Max Mean	0.432	-0.31	**-0.745**	0.052	0.004	Precipitation Sum Min	0.359	-0.262	**0.739**	0.154
	Precipitation Sum Min	457	-0.329	**0.684**	0.008	0.199	Precipitation Sum Min	0.46	-0.324	**0.686**	0.023	0.203	Daytime LST Max Mean	0.515	-0.372	**-0.656**	0.052
	Barren Area	0.157	0.078	-0.376	0.022	-0.206	Barren Area	0.158	0.079	-0.377	**0.03**	-0.189	Barren Area	0.205	0.049	-0.374	-0.163
	Population	0.017	0.11	0.019	**0.825**	0.057	Population Density	-0.035	0.026	-0.011	**0.983**	0.006	Urban Area	-0.106	0.025	-0.331	-0.028
	Urban Area	-0.08	-0.034	-0.113	**0.825**	-0.052	Urban Area	-0.057	0.003	-0.097	**0.979**	-0.005	Precipitation Min Max	-0.054	0.254	-0.139	**0.759**
	Precipitation Min Max	-0.028	0.243	0.127	0.025	**0.756**	Precipitation Min Max	-0.028	0.243	0.131	0.019	**0.769**					
		1	2	3	4	5		1	2	3	4	5		1	2	3	4
2014	Nighttime LST Min Max	**0.957**	0.055	-0.023	-0.068	-0.024	Nighttime LST Min Max	**0.954**	0.07	-0.047	-0.059	-0.034	Nighttime LST Min Max	**0.957**	0.69	-0.014	-0.064
	Elevation Min	**-0.954**	-0.052	0.052	0.071	-0.008	Elevation Min	**-0.946**	-0.066	0.054	0.096	0.073	Elevation Min	**-0.953**	-0.071	-0.005	0.078
	Vegetation Min Max	**0.827**	-0.084	0.181	0.186	-0.083	Vegetation Min Max	**0.831**	-0.122	0.121	0.185	-0.112	Vegetation Min Max	**0.829**	-0.133	-0.077	0.223
	Daytime LST Max Mean	0.566	**-0.291**	0.469	-0.472	0.044	Daytime LST Max Range	0.129	**0.852**	-0.011	-0.07	-0.061	Daytime LST Max Range	0.14	**-0.849**	0.118	-0.011
	Daytime LST Max Range	0.144	**0.829**	-0.2	-0.078	0.05	Nighttime LST Mean Range	0.009	**0.848**	0.335	-0.121	-0.113	Nighttime LST Mean Range	0.029	**0.809**	0.403	0.2
	Nighttime LST Mean Range	0.027	**0.821**	-0.363	0.223	-0.052	Vegetation Mean Min	0.14	**-0.75**	0.21	0.267	-0.105	Vegetation Mean Min	0.137	**-0.748**	-0.119	0.335
	Vegetation Mean Min	0.133	**-0.67**	0.319	0.337	-0.128	Precipitation Min Max	-0.141	0.543	0.055	0.488	-0.071	Precipitation Min Max	-0.15	0.535	-0.232	0.376
	Precipitation Min Max	-0.149	**0.623**	0.348	0.172	-0.035	Precipitation Sum Min	0.371	-0.175	**0.795**	0.116	0.05	Forest Area	-0.016	0.131	**0.91**	0.233
	Forest Area	-0.025	0.144	**-0.737**	0.567	-0.117	Forest Area	-0.046	0.224	**0.786**	-0.475	-0.066	Daytime LST Max Mean	0.562	-0.258	**-0.628**	-0.269
	Crop Area	0.083	-0.236	**0.722**	0.174	-0.023	Daytime LST Max Mean	0.577	-0.333	**-0.6**	0.215	0.027	Crop Area	0.079	-0.344	-0.555	0.38
	Precipitation Sum Min	0.358	-0.169	-0.103	**0.797**	-0.038	Barren Area	0.198	-0.004	-0.399	-0.203	0.13	Precipitation Sum Min	0.361	-0.292	0.469	0.548
	Barren Area	0.189	-0.079	-0.1	-0.504	0.004	Crop Area	0.118	-0.342	-0.04	**0.712**	0.113	Urban Area	-0.1	-0.069	-0.021	-0.537
	Population	0.012	0.114	0.005	0.066	**0.836**	Urban Area	-0.025	0.047	-0.203	-0.098	**0.766**	Barren Area	0.19	0.019	-0.142	-0.449
	Urban Area	-0.071	-0.071	0.02	-0.123	**0.817**	Population Density	-0.134	-0.132	0.105	0.173	**0.748**					

The Stage 3 PCA descriptive results in Table 2-2 indicate the methods were fairly consistent between the three models. The results show both population based models contain 5 components, and explain over 73% of the variance within the dataset. The EBE dependent model contained one fewer component, with 4, than the other models. This is probably due to the reduced number of input variables. The lower explanation of variance in the EBE model, a difference of 4-9%, suggests the lack of a population variable is responsible for this difference, but further analysis would be needed to confirm such a theory. The three models also had similar results when compared through the AUC, Table 2-3, demonstrating limited statistical difference between them. The ROC AUC also demonstrated the exploratory methods used in this study were able to provide results better than statistical chance, but they could be improved upon. The ROC AUC results represent a ranking of "Poor – Fair" statistical methods, which can be considered acceptable for an exploratory study (Tape). The similarity between all three PCA results suggest any proposed model could provide an accurate method to identify

patterns of Dengue Fever in the watershed, however the different models may be suited for different end-user purposes (Hajian-Tilaki, 2013; Hanley & McNeil, 1982).

Table 2-2: Principal Components Analysis Results

		2012	2013	2014
Population	# Components	5	5	5
	Cumulative Variance	74.22	73.40	74.07
Population Density	# Components	5	5	5
	Cumulative Variance	72.28	77.35	77.97
EBE	# Components	4	4	4
	Cumulative Variance	69.62	68.67	69.60

Table 2-3: ROC AUC result. There was no difference between the inputs in the Population and Population Density models for 2012-2013 when only the Top 2 component variables were used

	Model	Primary results AUC	TOP 2 Component AUC
2012	Population	0.69	0.67
	Density	0.69	N/A
	EBE	0.67	0.67
2013	Population	0.66	0.70
	Density	0.66	N/A
	EBE	0.67	0.70
2014	Population	0.66	**0.**74
	Density	0.71	0.73
	EBE	0.69	0.73

Night LST Min cell-Max zone consistently ranked high across all three models. Night LST Mean cell-Min zone and Daytime LST Max cell-Range zone were also highly loaded in the first or second component, supporting previous documentation on the influence of temperature range on mosquito survivability (Brady et al., 2013). Since both warmer daytime and cooler nighttime temperatures were found to be highly influential in the results, it demonstrates how the variability of temperature within an area may influence Dengue Fever transmission. This is particularly relevant since the range zonal statistic was determinant in both day and night during Stage 2. The general term 'temperature range' may be used hereafter to refer to the influence temperature variability can have on the dependent variable. LST data documented an annual

mean nighttime pixel range of -8.218 – 21.88°C during the night, and 8.1 – 42.73°C during the day for 2012 and 2014. The importance of the temperature range on Dengue Fever is further supported by the consistently strong inverse loading of elevation in the first component; high elevations contain colder temperatures and have been documented to impede mosquito habitation (Brady et al., 2013; Kanaroglou & Delmelle, 2015; Moreno-Madriñán et al., 2014). Another influential explanatory variable was identified as Vegetation Min cell-Max zone, which loaded positive in the first component for all models and could represent areas capable of supporting flora and fauna (Gubler, 1998).

An additional assessment was conducted to ascertain whether a simpler model could be developed, by using only the variables identified from the top 2 components of the Stage 3 analysis. A PCA was conducted on the 6-7 variables which weighed heavier (0.7 rounded or higher) in the top two components of stage 3 (see Table 2-1). Variables meeting this criteria were consisted of temperature, elevation, and vegetation. Since Population count and Population Density did not load in the top 2 components, the Population and Population Density models contained the same input variables for both 2012 and 2013, producing identical results. The 2014 model contained a different number of input variables between the Population and Population Density models, and the results deviated slightly. The results of the top 2 component PCA test can be viewed in the Supplement 2-1 table, while the component loading can be viewed in Supplement 2-2. It should be noted the component loadings were very similar between the top 2 component model and the original 14 variable PCA assessments. Since the variables were not altered between the studies, these results were not unexpected. The top 2 study does have fewer components and a larger explanation of variance, probably a direct result of the fewer input variables. The increase in explained variance does not necessarily demonstrate a stronger relationship to the dependent variable, but implies less noise is present

within the input variables. The decrease in components and increase in explanation of variance in the top 2 study is probably due to the use of variables which had already been identified as the strongest. The AUC results in Table 2-3 also indicate the similarity between all the models. The top 2 component model may only be easier to process for a future end user due to the need to collect fewer variables.

The SPSS regression variable output was mapped to demonstrate the geographical distribution of the project results. Following the procedures of previous vulnerability studies, the sum of the factor values derived from the Stage 3 PCA were mapped by municipality (S. L. Cutter et al., 2010; S. L. Cutter & Finch, 2008; Daniel P Johnson et al., 2012; C. E. Reid et al., 2009; Stanforth, 2011). Figure 2-3 demonstrates the visual comparison of the Population Density and EBE models to the distribution of reported Dengue Fever cases per 10,000. All three of the models follow a similar spatial pattern across all years by indicating low risk in higher elevated municipalities and increased risk in the more moderate temperature valleys, which coincide with areas more suitable for vector survival (Brady et al., 2013; Cleckner & Allen, 2014; Lozano-Fuentes et al., 2012; WHO, 2016). The maps suggest the population based models have a propensity to indicate higher risk in more densely populated areas, even if they exist above the elevation limit of the mosquito, such as Bogota. The EBE model does not indicate as high of a risk for densely populated areas, even those within the mosquito's habitable elevation. Therefore, it is anticipated that the population models weighed large urbanized areas stronger, whereas the EBE model did not. While all model types remain valid, according to the results demonstrated in Table 2-2, this could suggest different applications of the models depending on a future user's need. The results demonstrated between Table 2-1, Table 2-2, and Table 2-3 demonstrate similarities between the models outputs, however the mapped results may indicate different future applications for mitigation practices.

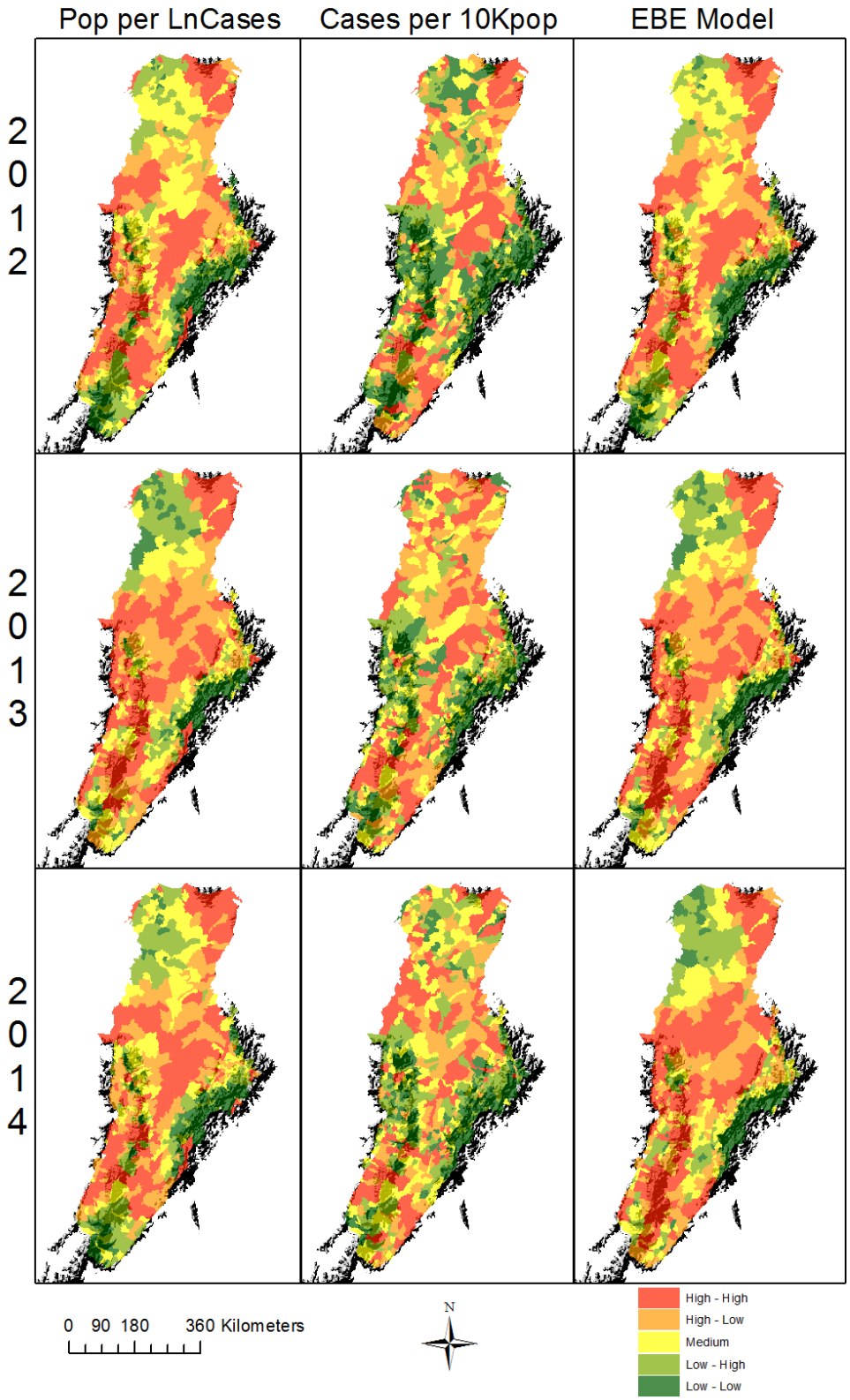

Figure 2-3: Population Density and EBE model graphical comparisons to the Reported cases per 10,000 populations

2.5 Discussion

Similar to what was found in the literature, this study supports the importance of temperature range to the identification and assessment of Dengue Fever modeling (Brady et al., 2013; Brunkard et al., 2008). Both nighttime LST Min cell-Max zone and daytime LST Max cell-Mean zone, were consistently ranked high in the top 2 components (Table 2-1). Previous laboratory studies on the *Aedes aegypti* mosquito indicated temperature range can impact the vector's ability to thrive (Brady et al., 2013), and field studies suggested increasing low temperatures improve vector resilience (Moreno-Madriñán et al., 2014; Pinto et al., 2011). Due predominantly to the mountainous terrain, the Río Magdalena watershed has a drastic temperature range which exceeds both the minimum and maximum temperature (15-35°C) for the mosquito's ability to thrive (Brady et al., 2013), impacting the extrinsic incubation period and duration of the gonotrophic cycle (Morin et al., 2013; Rueda, Patel, Axtell, & Stinner, 1990). Implication of temperature range were demonstrates by the incorporation of both nighttime and daytime determinant variables strong results in Stage 3. This aligns with documentation identifying temperature range's impact on the mosquito's ability to thrive (Brady et al., 2013; Brunkard et al., 2008; Pinto et al., 2011). Previous field experiments have focused on the minimum temperature limitation, due to their research in higher latitude locations, but this project supports other limiting factors which may be present in different geographical study areas (Moreno-Madriñán et al., 2014). As noted, the temperature range variables were based off three annual assessments, so the results do not necessarily document the actual temperature limits of the *Aedes aegypti.* Additionally, since temperature range is identified in this study as indicative of Dengue Fever, it supports the theory of vector habitat re-distribution is impacted by climate change (Brunkard et al., 2008; Ruiz-López et al., 2016).

The EVI, vegetation variables, loaded in the strongest components of Stages 2 and 3 (Table 2-1), which suggested vegetation is also useful in modeling Dengue Fever. The positive relationship between vegetation and dengue cases might be explained by the increased abundance of vegetation in wetter areas, which in turn are more likely to sustain larval rearing. This is particularly relevant near populated areas where surrounding vegetation may be more likely to be irrigated. Greener areas are often associated with increased moisture and biodiversity, leading to more fauna and flora, which may also provide opportunities for suitable breeding sites (Gubler, 1998). Although the LULC forest classification often contained an inverse relationship in Stage 3's results, the EVI variables had markedly better results. Dense forested areas may not have a high reported rate of Dengue Fever due to lower population rates, a possible explanation why the literature has mixed responses as to the influence of forest or large leaf areas on vector populations (Fuller et al., 2009). This may be further demonstrated as the literature indicates the *Aedes aegypti* preferentially breeds in peridomestic water sources, such as storage tanks or flower pots (Lozano-Fuentes et al., 2012). These water sources are commonly located in more populated or urbanized areas, which may have lower forest LULC classified space. Urban LULC similarly did not have strong results in the Stage 3 models, the highest being component 3 during the 2014 Population Density model. These results could be due to a low proportion of urbanized space within the municipalities, misclassification of suburban/rural communities in the LULC dataset, or due to several other anthropogenic causes.

Contrary to what was expected, precipitation, population, and LULC variables were less determinant within the Stage 3 models. Precipitation frequently loaded within the lowest 3-5 components across all models. This could represent a limitation of the data resolution or methodology used to measure rainfall, it could alternatively be explained by the dependence of this mosquito on peridomestic water sources for larval rearing, containers filled with water by

humans (i.e. water storage tanks, flower pots), suggesting the mosquito reproduction is less dependent on rainfall. LULC overall loaded in the weakest 4-5 components during this project, particularly the EBE model, even though the literature has indicated landscape composition to be important in modeling Dengue Fever (Beebe et al., 2009; Lozano-Fuentes et al., 2012; Moreno-Madriñán et al., 2014; WHO, 2016). This may indicate the 500m low resolution MODIS LULC product is less suitable for documenting environmental discontinuity in this type of study (Jensen, 2005; Moreno-Madriñán et al., 2014; Stanforth, 2011). The large pixel resolutions and municipality aggregation used in this study may reduce the ability to identify urban space, in particular (Kwan, 2012a, 2012b). It should be noted that barren LULC frequently contained stronger results than urban, which might be due to a misclassification of urban or agriculture space, or periphery/transitional landscapes (Jensen, 2005). The yearly assessment of this project, and the LULC data incorporated, may also fail to identify seasonal transitional landscapes (land with intermitted use) which may impact the rate of transmission by bringing people to areas with a higher mosquito breeding potential (Fuller et al., 2009; Lin & Wen, 2011). Seasonal transitional land could include agriculture or seasonal markets. Additionally, the majority of urbanized features in Colombia are built higher in elevation, where the mosquito is not anticipated to naturally thrive. This cultural dynamic may reduce the effect of urban or similar LULC results within this study area. Future studies should continue to explore varying LULC variables, even though they were less influential during this project.

All three of the mapped models locate higher risk areas within the mountain valleys, which is supported by the knowledge that the *Aedes aegypti* does not naturally thrive in high altitudes (Kanaroglou & Delmelle, 2015; Moreno-Madriñán et al., 2014). Table 2-1 indicates higher risk in lower altitudes due to the strong inverse relationship for elevation in every model. However, the population models represented some high populated area, such as the capital city

of Bogota, with an increased risk of Dengue Fever even though they reside above the 1,800m threshold as can be seen in Figure 2-3. This may represent a limitation of the models which used population as an independent variable, as indicate their results adds weight to larger populated municipalities. Other highly populated areas, such as: Cali and Cartagena, similarly displayed greater risk during the population model outputs, but are located within the mosquito's elevation threshold. The EBE model does not map high risk in any highly populated area, whether they are above or below the elevation threshold. This may suggest the models have different strengths, and may provide an opportunity for different applications based on the interest of the user. Since the EBE reduces the influence of political boundaries (due to inclusion of an adjacency matrix in its calculation and accounting for the influence of neighbor municipalities), it may provide a better indication of environmental suitability of the host mosquito. It could be used for dispersion models or to estimate mosquito exposure potential between areas, regardless of population size, and prove beneficial for mosquito control or pesticide intervention practices. The models using a population independent variable assign higher risk of outbreak to populated areas, which could prove beneficial for medical mitigation efforts. These are important distinctions to make, as public health departments may utilize different mitigation methods depending on whether an area is populated or not (Cleckner & Allen, 2014; Delmelle et al., 2014).

Since impervious surfaces increase local thermal conditions in a phenomenon known as the urban heat island, it should be considered that urban environments hold the potential to be oasis breeding habitats. The increased temperature and water storage capacities of urban impervious features may provide a haven for mosquito habitats in higher elevations (Moreno-Madriñán et al., 2014). A recent study has indicated the mosquito has the ability to survive above the previously measured elevation limit (Ruiz-López et al., 2016). Human activities can

also increase the promotion of vector breeding or host-vector contact through the establishment of gathering spaces in high vector breeding sites, such as seasonal markets (Lin & Wen, 2011). This suggests a need to strongly consider use of population-based models. However, it seems appropriate to mention a priori knowledge of the areas would be required to take advantage of the information provided by either type of model due to seasonal changes.

The annual resolution of this study was intended as an exploratory analysis. Future studies should consider a longitudinal design to assess the impact of the El Niño/La Niña seasons on the transmission of Dengue Fever. According to the literature, El Niño seasons are characterized by less precipitation and higher temperature in this region of Colombia, which could increase the incidence rate (Fuller et al., 2009). Studies on other mosquito vector diseases have found an increase in vector population when streams experience seasonally low water content, allowing for calm water breeding pools (Liu et al., 2008). Similarly, people may increase water storage practices during dry seasons and cause host amplification proximal to populated areas. Conversely, during La Niña seasons the additional precipitation might overflow water storage features, flushing *Aedes aegypti* larval specimens, and disrupt reproduction cycles. This could decrease, or disperse, the number of mature adults to reduce transmission. A longitudinal study using shorter time increments would better indicate any relationship between the rates of reported cases and seasonal weather patterns not explored in this study. Additionally, a temperature lag study could better identify specific conditions which increase the vector population, as the annual assessment in this study is unable to identify short term weather influence.

Future studies could also investigate the independent variables' impact over time. Since about 70% of the data's variability was explained across all models, there is still unexplained variance which could be investigated. The utilized independent variables could be modified to

potentially increase their use. For example, different elevation variables, such as: quantity of habitable area, percent of municipality space within the habitable altitude of the mosquito, or elevation slope could be used rather than the simple minimum elevation variable used here. This could provide a comparison of habitable space or precipitation accumulation, adding dimension to the variable. Additional metrics to quantify urban LULC should also be considered to improve the identification of habitable space for this vector.

Additionally, research should be conducted for the use of the independent variables as an early warning system. Remote sensing and GIS applications are already used for a number of automated early warning systems, including the Famine Early Warning System. Similar products could be implemented into a Dengue Fever, Malaria, or similar vector transmitted illnesses forecast tool (Ceccato et al., 2005). These could advance implementation procedures or assist medical personal in expediting the identification of an outbreak (Cleckner & Allen, 2014).

2.6 Conclusions

This analysis was able to appropriately identify variables identified as statistically relevant to the estimation of Dengue Fever transmission in the Colombian Río Magdalena watershed, as is demonstrated by Table 2-1. Stage 2 and 3 results indicate temperature is an important modeling variable within this study area. The use of nighttime Min cell and the Daytime Max cell variables designate that temperature range is a particularly relevant variable, presumably on how it relates to the comfort zone for the vector and virus replication. The inverse Elevation Min variable also loaded highly through all the results, supporting previously identified altitude limitations to the disease. Vegetation was the next most important variable to explain dengue risk, presumably due to a greater likelihood of larval rearing sites in wetter areas, which could also support increased quantities of vegetation. The most consistent results

suggest temperature and elevation strongly influence the mosquito vector's ability to thrive and transmit Dengue Fever.

The PCA, AUC, and mapped results suggest the variables used in this exploratory analysis should be considered in future Dengue Fever monitoring or prediction studies. Whether to adopt a population or EBE based model would be dependent on the individual's prior knowledge and focus. The population models were more likely to identify risk in higher populated areas, sometimes despite being located above the mosquito's expected habitable elevation. Alternatively, the EBE dependent model may better identify vector dispersion due to its use of an adjacency matrix, but does not appear to strongly weigh populated areas. Therefore, an individual could preferentially pick a model based on their background and need. A preference for identifying populated areas may be important for individuals focused on the distribution of medicine or medical information. However, if the individual is more interested in natural transmission rates, to conduct future field experiments or preventative mitigation at vector sources, they may prefer to use the EBE method.

As an exploratory study, the conditions and methods depicted within this report were able to appropriately quantify and model Dengue Fever reports through the use of environmental and population data. The results were able to explain approximately 70% of the dataset variability, across all three models used. Future studies should include the variables identified in Stage 2 of this study, but consider modifying them in an attempt to further improve model strength. Although Elevation Min proved to be a strong variable is this study, additional variations on the simple elevation variable used in this study may prove beneficial in the future. Other statistical methods, such as Geographically Weighted Regression or Boosted Regression Trees, could also be explored to interpret the relationship between the variables and Dengue Fever records. Still, this project was able to appropriately and accurately accomplish its goal of

identifying descriptive variables of increased *Aedes aegypti* and Dengue Fever cases. The results and models described by this exploratory analysis could be used to aid future mitigation practices or development of an early warning system for the transmission of Dengue Fever by the *Aedes aegypti* mosquito.

2.7 Supplemental Visuals

Supplement 2-1: PCA results from Top 2 Components

		2012	2013	2014
Population	# Components	2	2	2
	Cumulative Variance	76.48	80.15	77.14
	# input variables	7	6	7
Population Density	# Components	2	2	2
	Cumulative Variance	76.48	80.15	80.93
	# input variables	7	6	6
EBE	# Components	2	2	2
	Cumulative Variance	76.48	80.15	80.93
	# input variables	7	6	6

*Variable cut off for Component 2 was 0.7

** Population and Population Density independent variables were not in top 2 components, so the models were identical for 2012 and 2013; 2014 had a different number of independent variables

Supplement 2-2: Component loading from Top 2 Component review

2012

Population Vs. Ln cases

	Component	
	1	2
LST Min Max	**0.945**	0.145
Elevation Min	**-0.941**	-0.146
Vegetation Min Max	**0.832**	-0.089
Daytime LST Max Mean	**0.729**	-0.367
LST Mean Range	-0.071	**0.899**
Daytime LST Max Range	0.153	**0.874**
Vegetation Mean Min	0.092	**-0.747**

Population Density vs. Ln cases

	Component	
	1	2
LST Min Max	**no variable input**	
Elevation Min	**change from**	
Vegetation Min Max	**population study**	
Daytime LST Max Mean		
LST Mean Range		
Daytime LST Max Range		
Vegetation Mean Min		

EBE

	Component	
	1	2
LST Min Max	**0.945**	0.145
Elevation Min	**-0.941**	-0.146
Vegetation Min Max	**0.832**	-0.089
Daytime LST Max Mean	**0.729**	-0.367
LST Mean Range	-0.071	**0.899**
Daytime LST Max Range	0.153	**0.874**
Vegetation Mean Min	0.092	**-0.747**

2013

	Component	
	1	2
LST Min Max	**0.964**	-0.056
Elevation Min	**-0.963**	0.056
Vegetation Min Max	**0.768**	0.15
LST Mean Range	0.724	0.538
Daytime LST Max Range	0.025	**-0.905**
Vegetation Mean Min	0.084	**0.834**

	Component	
	1	2
LST Min Max	**no variable input**	
Elevation Min	**change from**	
Vegetation Min Max	**population study**	
LST Mean Range		
Daytime LST Max Range		
Vegetation Mean Min		

	Component	
	1	2
LST Min Max	0.964	-0.056
Elevation Min	**-0.963**	0.056
Vegetation Min Max	**0.768**	0.15
LST Mean Range	0.724	**0.538**
Daytime LST Max Range	0.025	**-0.905**
Vegetation Mean Min	0.084	**0.834**

2014

	Component	
	1	2
LST Min Max	**0.956**	0.097
Elevation Min	**-0.952**	-0.095
Vegetation Min Max	**0.82**	-0.112
Daytime LST Max Mean	**0.662**	-0.455
LST Mean Range	-0.028	**0.914**
Daytime LST Max Range	0.124	**0.888**
Vegetation Mean Min	0.148	**-0.754**

	Component	
	1	2
LST Min Max	**0.96**	0.072
Elevation Min	**-0.957**	-0.068
Vegetation Min Max	**0.849**	-0.159
Daytime LST Max Range	0.116	**0.91**
LST Mean Range	0.027	**0.886**
Vegetation Mean Min	0.155	**-0.781**

	Component	
	1	2
LST Min Max	**0.96**	0.072
Elevation Min	**-0.957**	-0.068
Vegetation Min Max	**0.849**	-0.159
Daytime LST Max Range	0.116	**0.91**
LST Mean Range	0.027	**0.886**
Vegetation Mean Min	0.155	**-0.781**

Chapter 3 Space and Time: Assessing Heat Mortality Distribution

3.1 Abstract

Extreme heat events are responsible for more weather fatalities than any other meteorological pattern. Previous vulnerability studies have investigated a historical heat wave event to correlate socioeconomic and environmental variables to increased rates of mortality. Research also needs to investigate the distribution of non-event specific mortality through space and time. This project was designed to investigate heat induced mortalities by comparing their geographic and temporal clustering. Geocoded heat mortality data was compared across temporal periods through cluster and hot spot analysis methods. The results demonstrate a clustering of mortalities, which could lead vulnerability studies and mitigation practices to anticipate the location and time of future events to improve preventive activities.

3.2 Introduction

Heat mortality and morbidity is easily preventable, but extreme heat can be associated with a variety of health complications (CDC, 2015). This is why it is currently considered the deadliest weather pattern throughout the world (Patz et al., 2000). During excessive heat events, it becomes difficult to naturally maintain normal core body temperatures. Recent climate studies anticipate normal temperature ranges will increase over the next decades, which could foster more frequent and severe extreme weather events. This indicates heat waves will increase in frequency and intensity in the future. Developing methods to improve mitigation practices and implement preventative aid will become increasingly important as these climate predictions come to fruition (Jackson & Shields, 2008).

The National Weather Service (NWS) offices are required to announce heat waves due to the potential health threats (Ebi et al., 2003). The NWS uses a standardized threshold of 105°F, for two or more days, to identify heat waves throughout the continental United States. This single criteria does not account for the regional climate differences which can be found

within the nation. The cool, moist Pacific Northwest experiences a drastically different climate than the hot, dry weather Arizona. Therefore, this single threshold may not appropriately identify inclement weather conditions. A local heat wave method is Robinson's Threshold, which tracks temperature conditions over the course of several years to record the distribution of a local climate. A Robinson's heat wave alert is issued when temperatures exceed the record's 95^{th} percentile over consecutive days (Robinson, 2001).

Heat can cause independent health complications, including heat stroke or dehydration, or can exasperate preexisting conditions, such as cardiovascular disease (CDC, 2002, 2006; Whitman et al., 1997). This causes individuals to succumb to inclement conditions in a variety of ways. Intervention practices become important risk prevention solutions, and should be designed to save lives through efficient low economic and organizational strategies (Ebi et al., 2003). Improving mitigation practices can be as simple as updating how vulnerable populations are identified or notified during inclement conditions. Heat wave warning studies have demonstrated the ability to prevent excess mortalities during heat events (Ebi et al., 2003; Laurence S. Kalkstein et al., 1996). However, simply increasing the number of warning messages has not been shown to improve resilience. In fact message fatigue, a phenomenon when people grow weary of and ignore notifications, has been found to reduce population adherence to warnings and mitigation practices if they are used inappropriately (Shen et al., 1998). This demonstrates the need for research to improve heat mitigation plans, to provide superior methods of monitoring and notification.

Previous studies have identified conditions, physiological and socio-environmental, which lead to increased rates of health ailments during excessive heat events. Temperature variations can be attributed to interactions between weather and the local environment, which can disproportionately influence resident health. Impervious surfaces and local building

practices, such as insulation or roofing materials, can increase local thermal properties through the urban heat island (UHI) (Q. Weng & Lu, 2008; Yuan & Bauer, 2007; X. Zhang et al., 2009). These features create discontinuous thermal strain and negatively impact human thermoregulation (Wang et al., 2004; Q. H. Weng et al., 2004; Q. H. Weng & Quattrochi, 2006; Zhou & Shepherd, 2010). Socioeconomic conditions can exasperate local heat health risks. Lower or fixed income households may be unable to afford artificial means of temperature control, such as air conditioning (Davis, 1997; S. L. Harlan et al., 2006; Naughton et al., 2002; O'Neill et al., 2005). Elderly populations have been found to be less able to regulate internal body temperatures due to preexisting health conditions. Reduced mobility in older or disabled individuals can also hinder their ability to relocate to safer environments or obtain fluids to maintain hydration (Changnon et al., 1996; Dolney & Sheridan, 2006; Pantavou, Theoharatos, Mavrakis, & Santamouris, 2011; Whitman et al., 1997). Dehydration is particularly dangerous for elderly men, who have even been shown to exhibit a reduction in thirst sensation (Semenza et al., 1996). As previous studies have demonstrated a propensity for these 'vulnerable populations' to maintain higher risks during extreme heat events, the question remains whether the location of these vulnerable areas are stable or dynamic across space and time.

Spatial analysis can be used to identify statistically significant patterns, or clusters, of vulnerable populations, as well as quantify trends or similarities within a dataset. Frequently, this entails measuring the distances from mean centers (the geographical proximity between extreme variables) of an area of interest. Measures of spatial autocorrelation, such as the Moran's I, are used to describe the local space correlation between variables, and produces an output to demonstrate the resulting statistical strength of the relationship (Esri, 2016a). Moran's I is designed to detect spatial autocorrelation by comparing the relationships between neighboring features through a spatially weighted correlation coefficient; it identifies local value

peaks, or discontinuities, across a study area (Greiling, Jacquez, Kaufmann, & Rommel). The method quantifies a feature's first-order queen neighbor weights as 1/# of regions, while all other values are 0, and compiles the study area's features into a final value. Output result values indicate the amount of dispersion in a dataset between -1 to 1; extremely dispersed data will be valued at -1, while complete spatial clustering is documented as 1. A value of '0' indicates the data is randomly distributed. The development of a significance level, or p value, is calculated in such models to identify whether the process is significantly different from the null hypothesis of random distribution. Pre-developed tools also may also report a Z score to further demonstrate whether the output would be naturally expected, where large values provide another measure of whether to reject the null hypothesis of random dispersion.

Another metric is the Getis-Ord General G analysis, which compares a feature and its neighboring values against the overall local population's sum. This metric is not designed to detect individual anomalies or outliers, as the Moran's I does, but rather to identify a group of values with a similarly elevated value within the study area (Greiling et al.). Therefore, the Getis-Ord is not designed to identify an individual's high value unless its neighbors also have relatively high values; this is accomplished by incorporating additional similar weighted neighbors the same as the center value, rather than only the queen neighbors used in the Moran's I. Thus the Getis-Ord is viewed as a population based assessment. Trend results in the dataset are documented as a single reported value. A positive score will identify clustering of positive values, while a lower (negative) will identify areas of negative or lower values. A Getis-Ord value near zero will demonstrate no statistical clustering. This statistic also requires Z scores and P values which, similar to the Moran's I, demonstrate the ability to reject the null hypothesis of random dispersion. Since a mortality dataset cannot contain a negative rate of death, a large number of boundaries within this study will contain a lower limit of a value of zero "0"

mortalities. This indicates the results are not expected to identify any negative valued clusters. The Esri ArcMap tool which calculates this statistic is found within the Spatial Statistics toolbox, and additionally provides a visual demonstration of mapping clusters through the Hot Spot Analysis tool (Esri, 2016a).

Previous heat wave studies would primarily focus on the geographic component of heat risk. The temporal variable consisting of the length, or severity, of an event or the seasonality of a phenomenon, example a specific heat wave or summer, respectively. Future improvements to heat vulnerability modeling may be dependent on identifying ways to quantify spatial clusters across both time and space. This would allow for the identification of vulnerable hot spots by documenting cluster locations, and the frequency or pattern at which they occur. This is different from the space aquarium method, which follows the transportation, or movement, and the time spent during the documentation. The aquarium method is also known as the Spatiotemporal Continuum, which is a geospatial feature (such as a set of points) configured in a continuous spatial arrangement organized by their temporal collection (Christakos, Bogaert, & Serre, 2012). Other methods of temporal Geospatial analysis typically branch into two branches, snap or span. Snap features refer to an endurant features, such as a river, which is ever present but whose attributes may change over time. A span or perdurant features have a value in time, such as a river's outflow at a specific point in time (Mathian & Sanders, 2014). When it comes to health or epidemiological studies, these methods get a variety of uses. Disease or vector studies may employ the perdurant method, when researchers analyze outbreak patterns or intensity of emergent diseases as stand-alone and unique features without a baseline infection rate; particularly if the disease was thought to have been eradicated (Cleckner & Allen, 2014; Curtis et al., 2014; Jerrett, Gale, & Kontgis, 2010). Other studies suggest disease is a more constant endurant variable, and examine the quantity or work to identify fluctuating prevalence rates,

such as seasonal flu or cancer rates (Delmelle et al., 2014; Greiling et al.; Kulldorff, Heffernan, Hartman, Assunçao, & Mostashari, 2005).

Heat mortality studies have had limited inclusion of the temporal aspect. Several early vulnerability studies focused on identifying the attributes of vulnerable populations, regardless of a specific event or adverse influencer, such as an endurant study (Susan L. Cutter et al., 2003; C. E. Reid et al., 2009). Other studies have examined isolated events, such as a single flood or heat wave, in perdurant fashion (Duneier, 2006; Ebi et al., 2003; Hondula et al., 2012; Daniel P Johnson et al., 2012; Laurence S. Kalkstein et al., 1996; Naughton et al., 2002; Semenza et al., 1996; Smoyer, 1998; Stanforth, 2011). Additional studies have attempted to investigate patterns of heat mortalities, but have not consistently used sufficient temporal or spatial resolution to identify the seasonality or micro patterns for long term studies (Daniel P Johnson et al., 2014; Miron et al., 2010). Methods to analyze heat mortality at a micro resolution require the use of a date attribute within a point shapefile, as both spatial and temporal coordinates are required to build the necessary relationships. Previous research has used the Moran's I or Getis-Ord statistics over defined study period intervals, and compared the results of independent study periods investigated at the same location (Rossen, Khan, & Warner, 2014; Ruddell, Harlan, Grossman-Clarke, & Buyantuyev, 2009). Newer statistical models, such as a Space-Time Kernel Density Estimation, have not been included in many publications but may be used to compare assessments through time (Delmelle et al., 2014). Input data build a virtual 3D or cube dataset to compare across space or time. Such a cube consists of an X/Y plane of the geographic location of the input points. The vertical, or Z, axis is created through the date attribute. Through this data structure, the spatial data can be explored through distinct layers of time, similar to a layer of cake. The data could also explore a single geographic location through the entire time span, by looking down the column of time (Delmelle et al., 2014; Esri, 2016a). One tool for such

analysis is the ArcGIS Space Time Cube, which identifies the spatial distance units either as the maximum spatial extent divided by 100, or by building an algorithm using a measure of the nearest neighbor principal on the spatial distribution of the input features. For small study areas, such as the single county used in this study, minimal difference would be expected between these two options (Esri, 2016b). Time bin resolutions use a Mann-Kendall style statistic to properly establish the data's trend, but this can influence the application to all study practices since it requires a minimum of ten time bins (Shimazaki & Shinomoto, 2007; Silverman, 1986).

This study was designed to assess whether reported cases of heat mortality cluster across space and/or time. If researchers can better recognize where and when mortalities occur, it may improve the collective understanding of why certain vulnerable populations are at risk and potentially lead to improved mitigation practices. This project undertook the approach of assessing heat mortality through multiple years, by assess the spatial clustering during both independent years and across the entire time period. The methodologies include previously used cluster analysis and more recently available space-time analysis tools.

3.3 Methods

3.3.1 Study Area

This study was conducted in Maricopa County, located in the continental southwest United States, including the city of Phoenix, Arizona. The 2010 American Community Survey estimated the county population is around 3,817,000. The area encompasses about 9,200 square miles, and boasts a population density of nearly 415 people per square mile (U.S.Census, 2015). This population density may appear to be low for an urban area, but is explained due to the county incorporating large unpopulated areas, including National Parks and a U.S. Air Force base. The average annual temperature for Phoenix, AZ is 75.05°F with an average range of 45-106°F. Being an arid environment, it has thirty-six anticipated days of precipitation a year (8.04

inches a year) (YourWeatherService, 2016). Considering the consistently high temperatures and small relief from clouds/precipitation, the region is poised to be influenced by frequent heat mortality events (CDC, 2005).

3.3.2 Dependent Variable

Mortality data was obtained through the Maricopa County Department of Public Health and consisted of 260 unique heat mortality records spanning 2009-2011. The database has been used for other successful studies, both internal and external to the department, due to their advanced monitoring techniques used to identify all heat or heat contributing cause of mortality (Sharon L Harlan et al., 2013; Health, 2015). The public health department has several resources available through their website to encourage the community to learn more about heat illnesses and take preventative action (http://www.maricopa.gov/publichealth/programs/heat/). The International Classification of Diseases (ICD) codes from the database include, but are not limited to: hyperthermia, cardiovascular complications, dehydration, and conditions identified as 'natural causes' or 'old age'. Previous studies demonstrated the inclusion of peripheral causes of heat mortalities, such as cardiovascular disease, does not overestimate weather's impact (Semenza et al., 1996; Shen et al., 1998; Whitman et al., 1997). Additionally, the dataset included intoxication deaths from alcohol and drug cases, which the substances can make individuals more susceptible to succumbing to heat (Dolney & Sheridan, 2006; Donoghue et al., 1997). The mortality data was first collected into Microsoft Excel and organized by date to identity the distribution of mortality spikes. The dataset also contained information pertaining to the date of death and age, which provide basic descriptive information about the study population.

The mortality data was paired with meteorological data obtained through the National Oceanic and Atmospheric Administration (NOAA). The National Centers for Environmental

Information, formerly the National Climatic Data Center, archived the meteorological data collected at the Phoenix Sky Harbor International Airport. Hourly data was downloaded from their online database from 1998-2012 in a text format to demonstrate weather trends during the study period. The weather data was converted into an Excel file and graphically compared to the mortality data. Heat waves for this area were identified using a derivative of the Robinson's Threshold. Consecutive days which reached the upper two standard deviations of the meteorological trend were identified as heat wave conditions. This criteria followed procedures similar to the Robinson's heatwave threshold, which uses the 95th percentile of yearly temperature distributions (Robinson, 2001).

3.3.3 Geocoding

The mortality data was geocoded using a Census TIGER/Line street centerline shapefile from 2011 (U.S.Census, 2011). The mortality dataset contained optional entries for both residential and injury locations, but the records were not complete for either section. Preference was given to geocode the residential address, as it was assumed to be the location where the individual would have spent the majority of their time and provide the greatest thermal influence. Several of the residential addresses indicated the individuals were seasonal or visitors to the area. Since the focus was on Maricopa County, the injury location was used for non-residents or for residential records which did not have a viable Maricopa County residential address. Several of the injury locations were coordinate points or indicated the individuals were injured during extracurricular activities, as they were not located near any built features. Coordinate locations were identified and a point was manually added to the geocoded dataset proximal to a nearby road. In all, 230 of the original 260 mortality records had viable location information, with 226 of those being located within the Maricopa County boundary and eligible for use in this study.

3.3.4 Spatial Analysis

To correspond with the 2009-2011 mortality data, Census block group resolution shapefiles were obtained through the American Factfinder website, using the 2010 TIGER polygon files for Maricopa County, AZ (U.S.Census, 2015). All spatial Analysis and data developmental operations were conducted in Esri ArcGIS 10.2/10.3. Since Maricopa County census boundaries had irregular area and shape formations, a standardized grid was created for computational comparative purposes. Esri documentation indicated hot spot analysis can benefit from the use of either a standardized polygon feature or a cluster distance output created by a General Spatial Weights Matrix tool (Esri, 2016a). A 2 km square grid shapefile was developed at the extent of Maricopa County through the Create Fishnet tool in the ArcGIS Toolbox. Initially a 1Km grid was attempted to correspond with MODIS satellite data, but was aborted due to low neighborhood values. The geocoded mortalities and Census population data were spatially joined to both the U.S. Census block group and Fishnet polygon grid shapefiles. Cluster analysis was conducted with ArcGIS Spatial Analysis tools, including the Spatial Autocorrelation Report (Moran's I) and High-Low Clustering Report (Getis-Ord Global). These two methods were chosen to identify the presence of anomalous-local spikes and neighborhood trends, respectfully. Analysis was conducted on the shapefiles during four temporal datasets: 2009, 2010, 2011, and All Years. Both the Spatial Autocorrelation and High-Low Clustering Getis-Ord General G tools used inverse distance conceptualization of spatial relationships over Euclidean distance parameters. Although it was indicated these parameters can be less efficient for features with inconsistent shapes and sizes, such as is present in the census boundaries, this method was used across all studies for consistency purposes (Esri, 2016a). Results were considered significant if they passed a P value of 0.05 or a Z score of 1.96.

Results of the mortality clustering were also compared to population and poverty records from the U.S. Census, aggregated at the block group level. Population and socioeconomic data were collected using the 2010 Decadal Summary File 1, Summary File 3, and American Community Survey 5 year data (U.S.Census, 2015, 2016a, 2016b). Since this data provides uniform information within the boundary, rather than a continuous representation of the actual population, a raster was developed to approximate population dispersion for the 2 km grid assessment. Census population and poverty data were converted to feature centroids, remaining within the political boundary, and a Kernel Density Function created a raster to estimate the variables' distribution. Kernel Density output cell size is determined by the shorter of the data extent's width or height, divided by 250. This produces a standardized representation of the dataset across the features. The search radius for the Kernel was set at 590 km using a spatial variant of Silverman's rule of thumb, so more distant points are less influential on the output dataset (Esri, 2016a; Silverman, 1986). A larger search radius would not improve the analysis as the features are relatively close within Maricopa County. The 2 km x 2 km fishnet grid was used to extract the raster values to compare against the mortality data. The socioeconomic grid was also processed through the Moran's I and Getis-Ord tools, just as the mortality data, to identify if natural or anthropogenic clustering of these variables could be identified. Investigation of the combined influence of poverty and population outputs were processed by averaging the mapped Getis-Ord result files together. Locations which maintained high averaged hot spots indicate areas of population and poverty overlap. These procedures produced a choropleth map of consistent hot spots and allowed for another Getis-Ord assessment of contributory risk. The images were visually compared against mortality hot spots to identify whether there was an overlap in the results.

The Time-Space Cube tool in ArcMap 10.3 was included in this study as an exploratory comparison to the static year methods previously described; this was an exploratory method as limited literature was available which incorporated the method for research. It was intended to be used as a validation method for the previously mentioned cluster analysis methods. The Cube required a standardized grid input, so the 2 km square grid shapefile was used. The Space Time Cube produces a significance value (P value) and a Getis-Ord result to demonstrate the presence or absence of clustering. The cube requires a minimum of ten time bins due the use of a Mann-Kendal statistic. A limitation to this tool's temporal set up is the bin design. To reduce potential temporal bias, bin development is initiated on the most recent records and built backwards, rather than conclude on a calendar year. The use of a two month time bin for a three year study, ending in September 2011, due to the dataset's extent, culminated in the development of fifteen time periods over the three year study. A three month time interval, culminating in ten time bins, was similarly developed to investigate how the winter months would bias the analysis. The three month interval still accrued a winter bias, but the comparison of results were expected to indicate how the results would be impacted if additional data were available to enlarge the time bins and reduce/eliminate empty winter periods.

A measure of the cluster pattern over time can assist in understanding whether the clusters occur in patterns. The Emerging Hot Spot Analysis calculates temporal clustering using the Space Time Cube results. Similar to the Getis-Ord neighbor cluster methods discussed previously, this process investigates the geospatial neighbors as well as the temporal bin neighbors for a feature. The results identify whether a bin has a consistent, increasing, decreasing, sporadic, or oscillating value of either high or low (hot or cold) value through the data's temporal component (Esri, 2016b). The study of mortality would not allow for negative,

or cold, spots in this assessment as the lowest value for a mortality count is zero, or no mortality records.

A supplementary analysis using the Space Time Cube and Emerging Hot Spot analysis tools were conducted using weekly and biweekly temporal bins on the dataset. The assessments would use the years independently to look for within season trends, and use a composite dataset transcribing all mortality records to within a single season. Weekly temporal bins over the three year period would not produce results due to the large winter season bias, and were therefore not reported. These tests were investigated to see if additional information could be extracted from the data and further identify any clustering of mortality within the study period.

3.4 Results

Multiple heat waves were recorded during the three year study period, using the Robinson's derived criteria of a two standard deviation threshold (110.4°F) of the 2008-2012 hourly temperature data (Robinson, 2001). In particular, the more severe annual heat events occurred July 2009 and 2010, and August 2011. Representations of the Robinson and NWS heat wave criteria and associated mortality trends can be viewed graphically in Figure 3-1, which also demonstrates the temporal similarities between increases in daily average temperature and heat mortality. A corresponding increase in mortalities was found to follow each heat wave. It should be noted that the graph demonstrates a heat spike before mortality increases, supporting the theory that consecutive oppressive days increase a heat wave's threat level (Robinson, 2001). A simple regression between daily average temperatures and mortality, using Microsoft Excel, indicates an R-squared value of 0.19 (p value < 0.05), further demonstrating a positive relationship between heat and mortalities. The low R-squared value was not considered to minimize the relationship, as winter season days were also included in the comparison. It is therefore anticipated the low R-squared value is due to the cumulative impact of temporal lag,

winter's lack of mortalities, and the use of average daily temperature reducing the temperature range.

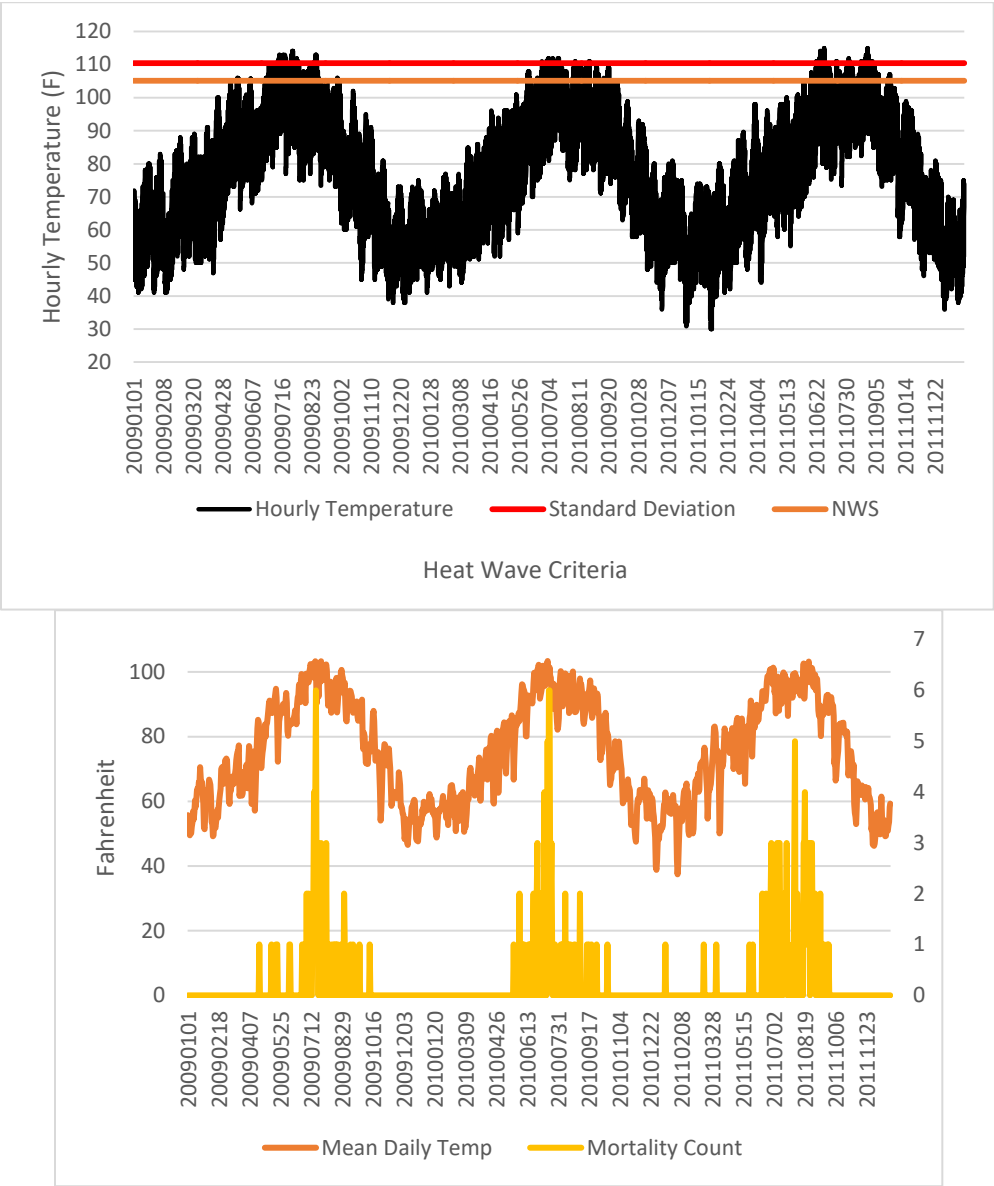

Figure 3-1: (Top) Temperature and heat wave thresholds. (Bottom) Mortality data compared to average daily temperatures recorded at the Phoenix Sky Harbor International Airport

Geocoding the mortality data was successfully conducted at 88% when using the residential or injury postal address. Post this process, the inclusion of coordinate locations resulted in a 100% successful location of the approximately 230 mortality records used in the

study. This information was able to be successfully used to conduct spatial and temporal analysis for this project.

3.4.1 Census Boundary Assessment

The Census boundary assessment was not able to provide results during any of the temporal assessments. Due to the varying size and border lengths, the eight required neighbors to calculate the spatial statistic could not be reached (Esri, 2016a). This was potentially due to the large features in the rural areas. As previously mentioned, the Inverse Distance conceptualization of neighbors, used in this study for consistency purposes, is complicated by the use of larger features. It may suggest vulnerability monitoring, particularly for heat, should be conducted at finer resolutions, as has been suggested by previous projects (Daniel P Johnson et al., 2012; Colleen E Reid et al., 2012; Stanforth, 2011).

3.4.2 2009 Grid Assessment

The mortality data from 2009 identified clustering of mortalities using both the Moran's I and the Getis-Ord results. Both assessments demonstrated the significance of the clustered results due to larger Z-score results of 16.2 for both statistics, and significant P values, displayed in Table 3-1. The mapped results in Figure 3-2 demonstrate the majority of the clustering was focused within the downtown Phoenix area, but rural clustering is also visible in the study area. One instance of rural area clusters is in the southwest of the study area, near the nature preserves. It is important to consider the more rural risk may be focused by recreational activities or proceedings on the nearby Air Force base, but cannot be confirmed with this data. Rural areas would require different mitigation practices than those used in populated areas.

Table 3-1: Cluster Analysis Results of the Fishnet Grid 2 km squared Shapefiles

		2009	2010	2011	All Years	2010 Census MHI	2010 Census Population
High-Low Clustering Report (Getis Ord)	Observed General G	0.000006	0.000004	0.000005	0.000058	0.000001	0.000002
	Z-Value	16.202028	10.508853	21.429159	4.870221	150.445525	240.284227
	P-Value	0	0	0	0.000001	0	0
	Results	Clustered	Clustered	Clustered	Clustered	Clustered	Clustered
	Random Chance?	< 1%	< 1%	< 1%	< 1%	< 1%	< 1%
Spatial Autocorrelation (Moran's I)	Moran's Index	0.150128	0.097354	0.200347	0.08031	0.554212	0.0885624
	Z-Value	16.1906	10.483176	21.408205	9.378838	150.356326	240.183852
	P-Value	0	0	0	0	0	0
	Results	Clustered	Clustered	Clustered	Clustered	Clustered	Clustered
	Random Chance?	< 1%	< 1%	< 1%	< 1%	<1%	< 1%

3.4.3 2010 Grid

Mortality data from 2010 indicated clustering from both the Moran's I and Getis-Ord results, containing Z-scores of 10.5 ($p < 0.05$) for both statistics. The results were slightly lower than the 2009 data. This is potentially due to the fewer documented mortalities in 2010, which contained the least number of reported heat mortalities of the study. The mapped results in Figure 3-2 indicate the mortalities occurred predominantly in the downtown Phoenix area and surrounding urban populated areas. There were limited cluster results in the peripheral rural landscapes, potentially an indication of less recreational mortalities during the year.

3.4.4 2011 Grid

Mortality cluster results from 2011 demonstrated significant P values, and Z scores of 21. The Moran's I result was the strongest between study iterations, at 0.20. The Getis-Ord value, however, was slightly weaker than the 2009 study. The increased quantity of mortalities reported in 2011 may have impact these results due to better reporting, or a bias in the dataset. Heat mortalities were also reported during the winter season of 2011, which could have exasperated the results. It may also indicate the difference between weather conditions through the years, as temperature records indicate 2011 may have been a more dangerous temperature season. The average May – September temperature was slightly higher during 2009 at 92.12°F (2010 = 90.5°F and 2011 = 91.04°F), but 2011 held 8/10 of the highest summer temperatures recorded during the study period. The average, or more consistently, increased temperature in

2009 might have caused more acclimation, while the more extreme records in 2011 may have contributed to more health complications and mortality.

3.4.5 All Years Grid

For the three year combined assessment, the results have significant P values for both the Moran's I and Getis-Ord tools. The Moran's I value of 0.08 suggests the data contained lower individual spikes within the dataset; this value was lower than any independent year. Whereas, the Getis-Ord value is larger than all other assessments at 0.000058. The Moran's I indicates small local spikes of mortalities decreased, but the Getis-Ord demonstrates a stronger neighborhood cluster of mortalities occurred, predominantly within urban areas.

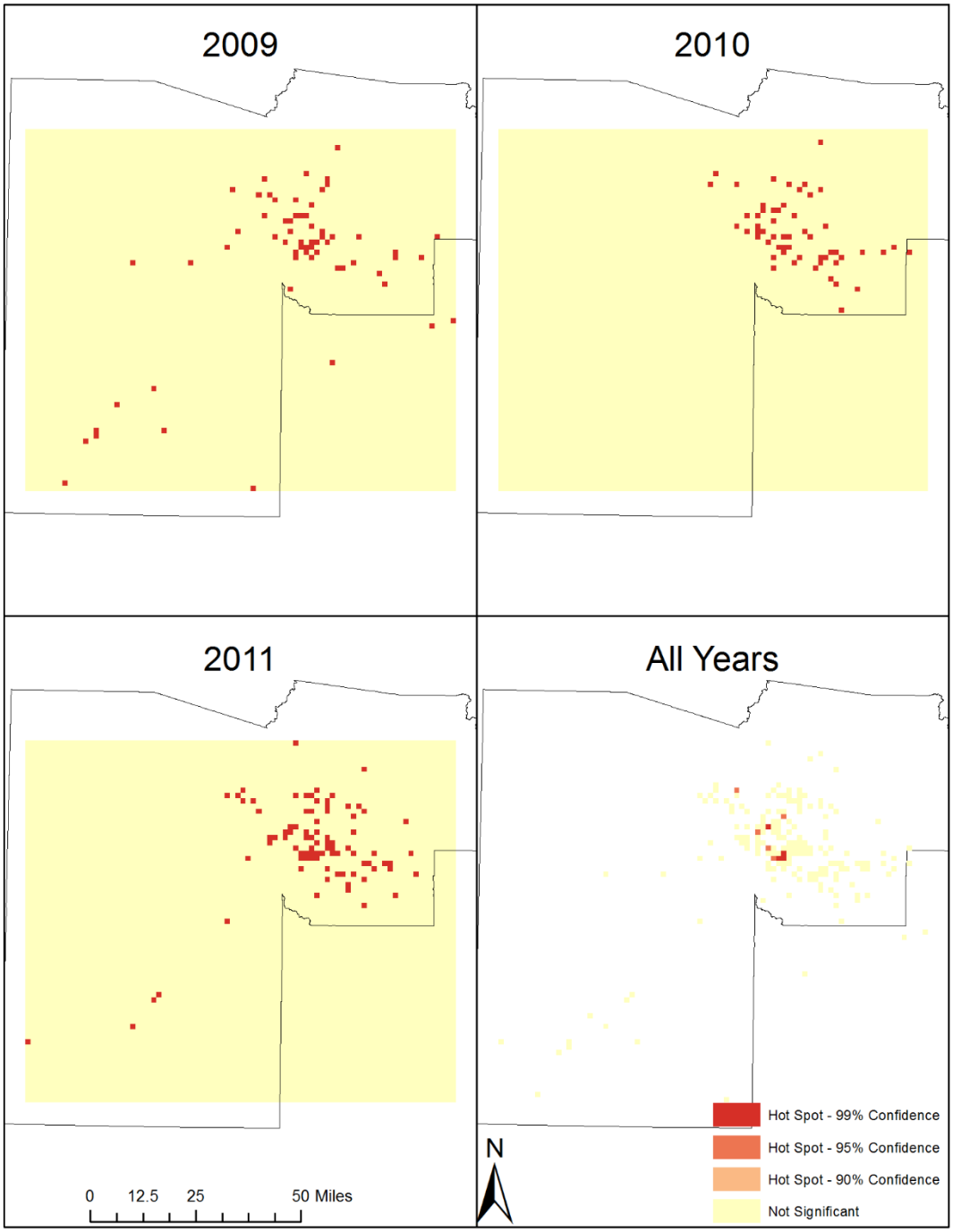

Figure 3-2: Yearly Fishnet grid outputs of mortality from the ArcGIS Hot Spot Analysis (Getis-Ord GI*) tool. Red spots identify derived hot spots for the study period

3.4.6 Space Time Cube results

The results from the Space Time Cube included the mortality data across all three study years. Due to the requirement of a minimum of ten temporal intervals, the results contained bias resulting from the short temporal bin size and the summer seasonality of heat mortalities.

The initial two month time bin contained a total of fifteen time intervals, identifying sporadic hot spot in the urban area. Similar to the Getis-Ord results in the All Years results previously discussed, a wider distribution (rather than individual spikes) was observed. The results also suggest non-consisted features, which may be related to the several vacant time bins occurring over the winter season. Since heat mortalities occur primarily between May-September, there were multiple two-month time intervals over the winter when mortalities were not frequently documented, creating a sporadic rhythm of events using this time resolution. The output of Figure 3-3 indicates the sporadic or oscillating hot spots in the Phoenix downtown populated areas. The overall results indicate a positive cluster statistic value of 0.3009, but the P value of 0.7635 indicates the results are not statistically significant.

The second assessment utilized a temporal bin size of three months, with ten intervals, with similar results. The larger time interval was anticipated to reduce the number of winter intervals and improve the temporal relationship. The cluster trend statistic increased to 0.3607, but the p value of 0.7183 remained not significant. The results may suggest the longer interval improved the identification of hot spot classification, as both the trend static and p value improved. However, since the p value is not significant no accurate interpretation of the Space Time Cube results can be formally investigated at this time. A future analysis with ten or more years of data would be better suited to implement this tool.

The Emerging Hot Spot analysis did not identify any cold spots in the results. This is supported by the mortality count dataset used, which did not contain any negative values, and the large number of zero values for grids which did not report mortalities during the study period. The results identified three types of results classifications: New Hot Spot, Sporadic Hot Spot, and Oscillating Hot Spot. A New Hot Spot is identified as a statistically significant hot spot within the final time step, which has not been statistically significant before. The Sporadic Hot

Spot is a location with intermitted identification of a statistically significant hot spot, which has occurred less than ninety percent of the time but never identified as a cold spot. An Oscillating Hot Spot is a similar to a Sporadic Hot Spot in which the last time period was a significant hot spot, but has the potential to be a cold spot if the data allowed for it (Esri, 2016b). All other empty grids displayed in Figure 3-3 represent areas where mortalities were present, but no trend or pattern was discernable. Due to the seasonality of the data, and the winter moth bias, it is inconceivable the data would have been able to produce a consistent hot spot. The oscillating and sporadic designations should indicate a location which repeatedly identified as a hot spot during the summer months. Future analysis should further test this theory with a longer dataset.

Table 3-2: Space Time Cube results

SPACE TIME CUBE DEVELOPMENT	**2 month bin**	**3 month bin**
Time steps intervals	15	10
Trend Statistic	0.3009	0.3607
Trend P value	0.7635	0.7183
Trend Direction	Not Significant	Not Significant
EMERGING HOT SPOT		
Consecutive Hot Spots	12	20
Sporadic Hot Spots	3	20
Oscillating Hot Spots	15	0

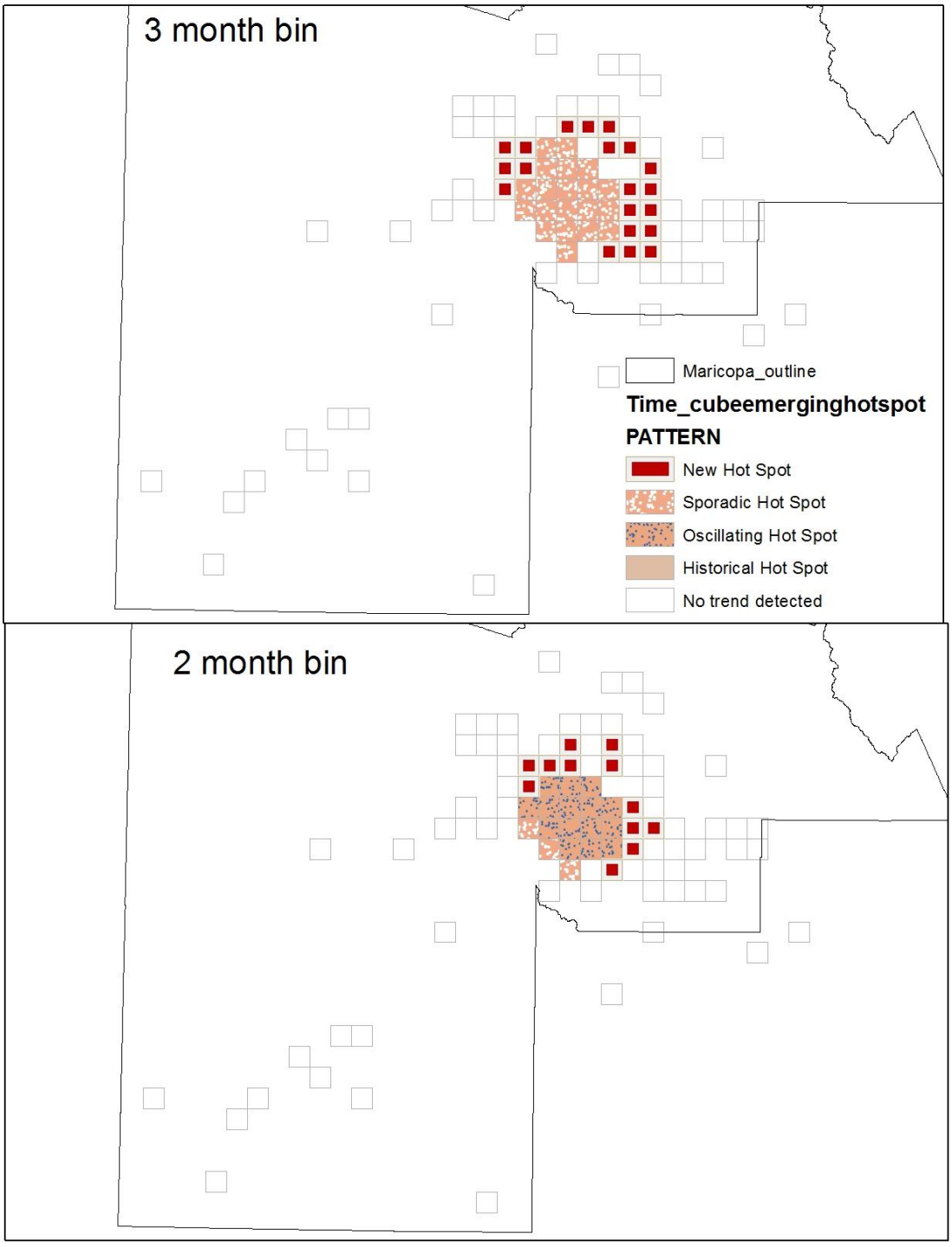

Figure 3-3: Time-Space Cube output

3.4.7 Confounding Variable Assessment

Assessment of potential confounding variables indicated population and poverty values did overlap mortality clusters in the downtown Phoenix area, as can be seen in Figure 3-4 . Both

socioeconomic variables contained significant p values and cluster results for both the Moran's I and Getis-Ord outputs, demonstrated in Table 3-1, demonstrating strong local and neighborhood statistical clustering. Population contained a Getis-Ord value of .000002 and Moran's I of 0.885624, indicating stronger local clustering. The poverty outputs were very similar at 0.000004 for Getis-Ord and Moran's I at 0.858154. Likewise, the averaged population/poverty Hot Spot identified much of the urbanized areas overlapped with both high population and poverty rates. The population and poverty hot spots encompass a much larger amount of space than the mortality clusters, which suggests population and poverty were not the only contributing factors. If they were, the mortality clusters would have conformed to more similar shapes/extensions as the population and poverty overlay cluster map. This supports previous research indicating heat vulnerability is the culmination of several contributing factors (such as age, education) during extreme heat events (S. L. Cutter et al., 2008; Susan L. Cutter et al., 2003; Ebi et al., 2003; S. L. Harlan et al., 2006; Daniel P Johnson et al., 2012).

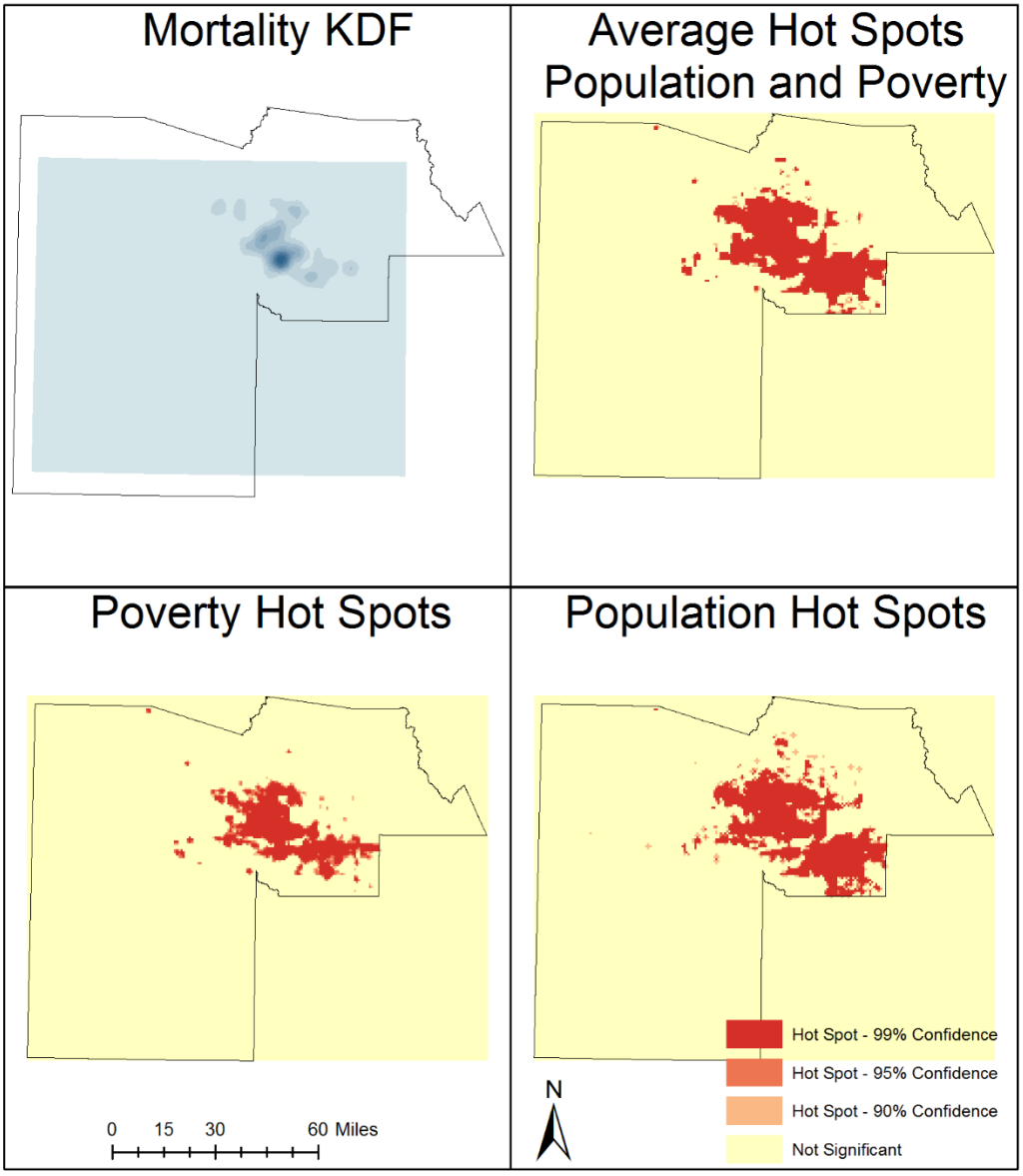

Figure 3-4: Confounding Variable Assessment on Mortality Clustering

Descriptive statistics of the mortality identified a mean age at death of 58.4 years, with a median age of 57.5 years. The background literature suggests the 65 and older demographic is one of the highest risk categories for heat mortality, and would therefore be expected to dominate the statistics of those who expired during the study period (Changnon et al., 1996; Dolney & Sheridan, 2006; Ebi et al., 2003; Whitman et al., 1997). Although the range of 65 and older does include some of the largest overall quantity, due to the large range, the descriptive

statistics do not suggest they are overwhelmingly represented. With a total range of 1-102 years of age, the dataset does not specifically identify older populations as the only at risk age group. In fact the mode, or most numerous value, was 49 years of age. This suggests that although elderly populations tend to report a higher risk, due to numerous chronic health conditions, their presence in this study may not disproportionately increase risk in this dataset (Susan L. Cutter et al., 2003; Naughton et al., 2002). It should be noted Phoenix, AZ is known to be a common retirement or seasonal home for older populations. Any seasonal visitors may not identify Phoenix as their residence during the Census, which could impact their classification or location of residence, or demonstrate their financial security is sufficient to allow for more medical or preventative assistance. The median age in the study area is higher than the 2010 Census national median age of 37.2, but the Census value is related to living persons rather than mortality records (Bureau, 2011).

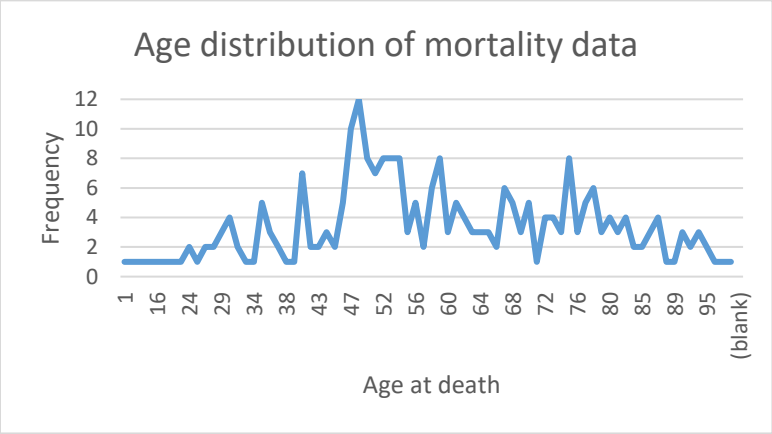

Figure 3-5: Distribution of Age at time of death

3.4.8 Supplemental Weekly assessment

The supplemental analysis of mortality data using weekly and biweekly intervals for the Space Time Cube did not improve the analysis over the previously discussed not significant results of the two and three month bins. Analysis on the individual years (2009, 2010, and 2011) provided no significant results for the biweekly assessment. The weekly designation contained

one significant result using the 2011 data with a trend value of 3.27 ($P < 0.05$), indicating an increasing trend in mortalities toward the end of the summer. It should be noted that 2010 did not contain enough data points to even run the model. Whereas the only significant result, using the 2011 data, contained the most data points.

Similar results resulted from consolidating the data into a single year, which extended between March 14 and October 17, representing the earliest and latest reported heat mortalities of any year. The biweekly results were not significant, as with the previous single year studies. The weekly assessment did have a significant result, $p = 0.03$, and a trend statistic of 2.17 to indicate an increase in clustering towards the conclusion of the study period. However, the Emerging Hot Spot analysis did not identify any discernable pattern in the dataset. This is potentially due to the consolidated test compiling the actual yearly or seasonal patterns into single event.

3.5 Discussion

The methodology was able to study the spatial and temporal clustering of mortality in Maricopa County, AZ. This analysis improves upon previous work by investigating multiple years of mortality data, and combining the spatial and temporal components, rather than study them independently (Christakos et al., 2012; Cleckner & Allen, 2014; Curriero et al., 2002; Daniel Johnson et al., 2009; Daniel P Johnson et al., 2014; D. P. Johnson & Wilson, 2009; Miron et al., 2010; Ruddell et al., 2009; Semenza et al., 1996). This study also investigated the use of newer space time models, such as the Space Time Cube and Emerging Hot Spot. However these tools were unable to support the current results as they required a longer time period than was available in the current dataset.

One of the simplest, and most important results to note is the benefit of using the Robinson heat wave threshold of 105°F for this study location. The Robinson threshold was able

to identify several distinct heat waves. The NWS heat wave criteria (daytime high temperature of 105°F and nighttime temperature lower limit of 80°F) identified nearly the entire summer season as heat wave conditions in this warm, arid climate. These results alone demonstrate the need for location specific heat wave thresholds (Curriero et al., 2002). These results also indicate that the implementation of the NWS criteria could incur increased message fatigue, due to the over estimation of dangerous conditions, in this and other climate regions.

The results of the Moran's I and Getis-Ord fishnet grids did identify several areas which contain hot spots, in close proximity, through all of the study years. It should be reminded that the two methods indicate different types of clustering. The Moran's I has the ability to detect a specific neighborhood (grid) which has increased mortalities compared to its neighbors, while the Getis-Ord is able to identify a group of neighborhoods (section of town) which all have higher risk. Therefore, a consistent neighborhood with a high Moran's I throughout the years indicates a very localized, but consistent, problem. An area with a Higher Getis-Ord and intermittent Moran's I suggest the larger area needs to be considered vulnerable, but specific neighborhoods experience occasional increased relative risk disproportionately, possibly depending on the time or event. These would require different mitigation practices. A high, consistent Moran's I may suggest a stationary cooling center would be useful in an area, while a high Getis-Ord may better utilize a mobile system which could relocate to distribute aid as vulnerability migrates through the area.

The results of the individual year assessments, using the Fishnet grid, consistently identified clustering particularly in or near the downtown Phoenix urban environment. The All Years assessment similarly indicated the presence of clustered mortalities, but the decrease in the local Moran's I value from those of the independent years indicated the specific location of the clustering may slightly fluctuate between years. Since a high Moran's I suggests a single

neighborhood contained a hot spot, in relation to its 'colder' neighbors, but the All Years data does not show these oasis hot spots to be consistent through time. Therefore, mortality spikes may have occurred in adjacent neighborhoods (grid cells) between adjacent years. The transfer of a hot spot between neighborhoods would elicit a high Moran's I during the independent year studies, but reduce the identification of continuous hot spots (low Moran's I value) in the All Years study. This supports the result of a higher Getis-Ord value in the All Years study, which aggregates these individual spikes into a larger feature. This could be explained by the theory of mortality displacement, which generally refers to a temporal displacement of mortalities, but could also suggest a spatial displacement. Previous research on mortality displacement have used lower resolution features (such as County or States) for their assessment (Hales, Edwards, & Kovats, 2003; Laurence S. Kalkstein, 1993; Laurence S. Kalkstein & Greene, 1997).

The Getis-Ord results from the All Years grid supports the theory of displaced mortality. Its increased value indicates the presence of neighboring grids with mutually higher rates of mortality. These results indicate only a few of the 2km grids had consistently elevated mortalities through the entire three year dataset, but several areas oscillated mortality spikes between years. This can suggest prevention and mitigation strategies need to incorporate ongoing or updatable designs to compensate for seasonal shifts in vulnerability. The current use of larger scale, standardized mitigation plans may provide aid in the same location every year; these results suggest this may not appropriately meet a population's need. Future studies could also investigate why the mortalities fluctuate between neighboring spaces. Do population demographics fluctuate significantly between years? Or is it a representative of mortality displacement, where the vulnerable population in an area is reduced temporarily by a previous event, and therefore the higher rate of mortality 'displaces' to another proximal area? Additional research could also investigate the few neighborhoods with consistently increased

mortalities to identify social or environmental variables which contribute to the consistently higher rates of heat mortalities.

The Space Time Cube did not produce many significant results during this study, presumably due to the length of time available in the dataset. It was unable to support the results of the yearly assessments during the 2-3 month temporal bin resolution. If data was available for ten or more years, allowing for a temporal bin size of one year, it may theoretically reduce the winter season interval bias and provide improved results. Additionally, the yearly independent assessments, using weekly and biweekly temporal bins, similarly were only able to produce significant results for the 2011 year, which had more data points than any other year. The 2011 data indicated an increase in clustering points, suggesting the end of summer season was hotter, which is supported by the weather data. Similarly, when the data was consolidated into a single year the analysis produced significant results with the weekly bin resolution demonstrating an increase trend towards the end of summer.

The Emerging Hot Spot Analysis demonstrated sporadic and oscillating hot spots using the 2009-2011 dataset, though the Space Time Cube did not contain significant trend results. These locations are proximal to the location identified during the yearly Getis-Ord mapped results in Figure 3-2, and may suggest similar hot spots. The oscillating or sporadic connotation may simply refer to the seasonality of the data, as mortalities were rarely recorded in the winter season. It should be noted that when mortalities were consolidated into a single yearly dataset no pattern was discernable using the Emerging Hot Spot Analysis. This tool would benefit from a longer dataset to help identify mortality clusters using an annual bin size. Future studies are not encouraged to use the consolidated dataset to identify temporal pattern with such a tool, as they would miss the seasonality of heat risk, which this study attempted to explore.

A future study should incorporate a larger temporal dataset to see whether significant Spatial Time Cube results would support the hot spots identified using the Moran's I and Getis-Ord results from independent years in Table 3-1 and Figure 3-2. Although not significant, the comparison between the 2 and 3 month temporal bin sizes in the Space Time Cube suggested the use of larger time bins, such as a year, could improve the use of this type of analysis. Such future studies could benefit mitigation specialists by helping to identify the seasonality of risk. It would be useful for public health officials to visualize whether the location of specific hot spots were consistent between all years - requiring repetitive aid, increasing- requiring additional assistance, or decreasing- due to previously implemented aid.

The fishnet grid analysis demonstrated the presence of several hot spots in downtown Phoenix, AZ through both visual and statistical results, even without the confirmation of the Space Time Cube results. The application of repetitive study through time identified a few locations to be visually consistent hot spots in the downtown area. The proximity of clustering within the urban areas may support previous vulnerability studies which focus predominantly on urban features, as the surrounding rural areas were less prone to mortality clustering. However, due to the short amount of time covered in this study, all of these results should be considered preliminary. A longer dataset would improve future investigates into the dynamic nature of the mortality hot spots identified in this study.

3.6 Conclusions

Overall, this study demonstrated the presence of heat mortality hot spots within Maricopa County. The results support previous research indicating heat vulnerability is not a continuous feature across urban landscapes, but rather can vary across both space and time. These methods identified discontinuous spatial clusters of heat mortalities which overlapped or oscillated between years. Therefore, mitigation practices and aid (preventative and responsive)

can be planned or applied when specific areas are identified to be higher risk areas. This high resolution mitigation practice could improve overall community health and resilience during inclement conditions, as well as help plan for future events. Understanding the temporal trend of hot spots, and considering additional thermal or vulnerability study results, could further improve the ability to document and aid specific neighborhoods which maintain consistently elevated vulnerability levels. Not only could mitigation practices be tailored to areas with increased risk, the identification of season risk could help mitigation managers distribute limited resources to areas and populations during the time when they are necessary.

Future studies should incorporate larger datasets to further document and compare local and neighboring increases in mortalities by the Moran's I and Getis-Ord statistics, respectfully. The addition of extra data can further improve the analysis and allow for the use of additional tools, such as the Space Time Cube, which could support and improve the results for researchers and public health managers. This study successfully demonstrated higher rates of population and poverty were present in areas with increased mortalities, but it suggests additional variables also contribute.

The preliminary methodologies of this study explored the ability to investigate heat mortalities through both space and time. It furthers previous research by expanding upon mono-temporal studies to demonstrate clusters of mortalities can occur consistently or fluctuate over time. It also suggests the theory of mortality displacement may need to be expanded to include spatial displacement. This methodology was designed to advance future studies and mitigation practices, by pushing the process and development of local scale mitigation strategies. The limitations of the data do not allow for the extrapolation of the results to other locations or study periods, and additional assessment would be required to confirm these results in other climate areas or time periods.

Dissertation Conclusions

The projects described in the above chapters appropriately demonstrate some of the potential applications of GIS and remote sensing techniques, and how they can be used in Public Health applications for medical and vulnerability studies. All of the studies were able to appropriately collect, store, process, and retrieve spatial data to advance the understanding of socioeconomic and environmental influences on health.

The study of Chicago, IL heat wave mortality was able to further demonstrate the application of the EHVI model's improvement over more commonly implemented low resolution methods. The results support previous studies which identified specific socioeconomic variables, including age and economic attainment, which are attributed to increasing vulnerability within populations. The use of smaller spatial units to identify vulnerable populations was also supported due to the ability to better identify where the highest risk populations were located. The results do indicate the presence of aggregation bias at multiple resolutions, which can also be referred to as the MAUP. Lower resolution data, such as that provided by satellite data, reduces the inclusion of environmental variables within the analysis. Future research and products will need to further study these limitations.

Remote sensing and spatial analysis tools were also used for vector disease mapping in Colombia. A PCA analysis was used across multiple models, and demonstrated consistent environmental and population variables were influential for the identification of vector habitats and human transmission. Results for temperature and elevation support the literature, by demonstrating their potential as limiting factors for the dispersion of the *Aedes aegypti* mosquito vector and the viruses which it can transmit. The graphical mapped results show reduced virus transmission risk corresponds to higher elevations in the Andes Mountain range, above the mosquito's threshold. The population models did suggest the potential for higher

than expected mosquito – human disease transmission rates in urban/populated areas, even if they existed above the anticipated habitat elevation boundary. Due to the lower resolution of the available data, and the use of a non-predictive statistical model, future studies should further investigate the relationships identified in this model. The use of different statistical methods could provide predictive or forecasting models, by using a longitudinal data set not included in this study. Additional studies should also consider whether the UHI and impervious surfaces in populated areas could provide oasis breeding grounds for the vector, as is suggested by the population based models.

Finally, the assessment in Phoenix, AZ demonstrated heat mortalities tend to cluster. Moran's I and Getis-Ord statistical results indicated significant cluster results when compared within individual years or seasons. The composite data covering three years of heat mortality data also demonstrated a significant population clustering result, but did not demonstrate as many location spikes through the Moran's I as the individual years. The results may have indicated a potential expansion in the definition of mortality displacement. The majority of the mortality data resided within the urban center of Phoenix or its surrounding cities. The repeated clustering of mortalities in the same, or proximal, neighborhoods supports previously conducted EHVI results encouraging the use of local or small boundary vulnerability modeling. It also demonstrates potential uses of focused mitigation planning procedures. This study was limited by the number of years and mortalities in the dataset. The Space Time Cube and Emerging Hot Spot results were not able to support spatial or temporal clustering results due to an inefficient quantity of data. Future studies could build off these results by utilizing a longer study period, to further document the temporal cluster of mortality. Additional vulnerability studies could also be assessed within the sites which consistently identified as clustered, or hot spot, locations to improve future understanding of vulnerability variables to improve mitigation projects.

Overall these projects were successfully carried out and support the application of GIS and remote sensing operations for medical and environmental research. The projects were all capable of providing statistical analysis to support and advance the field's literature. Although continued research and investigation is encouraged, the results are sufficient to provide guidance and advice for future researchers, public health officials, or city managers to continue to improve current mitigation practices.

References

Andrew, M. K., Mitnitski, A. B., & Rockwood, K. (2008). Social Vulnerability, Frailty and Mortality in Elderly People. *Plos One, 3*(5), -. doi: Artn E2232

Doi 10.1371/Journal.Pone.0002232

Beebe, N. W., Cooper, R. D., Mottram, P., & Sweeney, A. W. (2009). Australia's Dengue Risk Driven by Human Adaptation to Climate Change. *PLoS Negl Trop Dis, 3*(5), e429. doi: 10.1371/journal.pntd.0000429

Bolstad, P. (2015). *GIS Fundamentals: A first text on Geographic Information Systems* (fourth ed.): XanEdu Publishing Inc.

Brady, O. J., Johansson, M. A., Guerra, C. A., Bhatt, S., Golding, N., Pigott, D. M., . . . Maciel-de-Freitas, R. (2013). Modelling adult Aedes aegypti and Aedes albopictus survival at different temperatures in laboratory and field settings. *Parasites & vectors, 6*(1), 1-12.

Brandel, J. (2004). *Empirical Bayes methods for missing data analysis.* Uppsala University, June 2004. Retrieved from http://uu.diva-portal.org/smash/get/diva2:305249/FULLTEXT01.pdf

Brody, H., Rip, M. R., Vinten-Johansen, P., Paneth, N., & Rachman, S. (2000). Map-making and myth-making in Broad Street: the London cholera epidemic, 1854. *The Lancet, 356*(9223), 64-68.

Brunkard, J. M., Cifuentes, E., & Rothenberg, S. J. (2008). Assessing the Roles of Temperature, Precipitation, and ENSO in Dengue Re-Emergence on the Texas-Mexico Border Region. *Salud pública de México, 50*(3), 227-234.

Bureau, U. C. (2011). 2010 Census Shows Nation's Population is Aging [Press release]. Retrieved from https://www.census.gov/newsroom/releases/archives/2010_census/cb11-cn147.html

Cameron, D., & Jones, I. G. (1983). John Snow, the Broad Street pump and modern epidemiology. *International Journal of Epidemiology, 12*(4), 393-396.

CDC. (2002). Heat-Related Deaths—Four States, July-August 2001, and United States, 1979-1999. *Mortality and Morbidity Weekly Report, 51*(26), 567-570. http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5126a2.htm

CDC. (2005). Heat-Related Mortality --- Arizona, 1993--2002, and United States, 1979--2002. *Morbidity & Mortality Weekly Report, 54*(25), 628-630.

CDC. (2006). Heat-Related Deaths - United States, 1999 - 2003. *Morbidity and Mortality Weekly Report, 55*(29), 796-798. http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5529a2.htm

CDC. (2015). Extreme Heat Prevention Guide - Part 1. (June 22, 2016). http://emergency.cdc.gov/disasters/extremeheat/heat_guide.asp

CDC. (2016, January 19, 2016). Dengue. Retrieved February 2016, 2016, from http://www.cdc.gov/dengue/

Ceccato, P., Connor, S., Jeanne, I., & Thomson, M. (2005). Application of Geographical Information Systems and Remote Sensing Technologies for Assessing and Monitoring Malaria Risk. *Parassitologia, 47*(1), 81-96.

Chander, G., Markham, B. L., & Helder, D. L. (2009). Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment, 113*, 893-903.

Changnon, S. A., Kunkel, K. E., & Reinke, B. C. (1996). Impacts and Responses to the 1995 Heat Wave: A Call to Action. *Bulletin of the American Meteorological Society, 77*, 1497-1506.

Chen, X. L., Zhao, H. M., Li, P. X., & Yin, Z. Y. (2006). Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes. *Remote Sensing of Environment, 104*(2), 133-146. doi: DOI 10.1016/j.rse.2005.11.016

Christakos, G., Bogaert, P., & Serre, M. (2012). *Temporal GIS: advanced functions for field-based applications*: Springer Science & Business Media.

Cleckner, H., & Allen, T. R. (2014). Dasymetric Mapping and Spatial Modeling of Mosquito Vector Exposure, Chesapeake, Virginia, USA. *ISPRS international journal of geo-information, 3*(3), 891-913. doi: 10.3390/ijgi3030891

Costello, A. B., & Osborne, J. W. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Gettign the Most From Your Analysis. *Practical Assessment, Research & Evaluation, 10*(7), 9.

Curriero, F. C., Heiner, K. S., Samet, J. M., Zeger, S. L., Strug, L., & Patz, J. A. (2002). Temperature and mortality in 11 cities of the eastern United States. *American Journal of Epidemiology, 155*(1), 80-87.

Curtis, A., Ye, X., Heob, E., Targhetta, J., Salvato, V., Reyna, M., . . . Holmes, L. (2014). A comparison of three approaches to identify West Nile Virus mosquito space-time hotspots in the Houston Vicinity for the period 2002–2011. *Applied Geography, 51*, 58-64.

Cutter, S. L., Barnes, L., Berry, M., Burton, C., Evans, E., Tate, E., & Webb, J. (2008). A place-based model for understanding community resilience to natural disasters. *Global Environmental Change-Human and Policy Dimensions, 18*(4), 598-606. doi: DOI 10.1016/j.gloenvcha.2008.07.013

Cutter, S. L., Boruff, B. J., & Shirley, W. (2003). Social Vulnerability to Environmental Hazards. *Social Science Quarterly (Blackwell Publishing Limited), 84*(2), 242-261.

Cutter, S. L., Burton, C. G., & Emrich, C. T. (2010). Disaster Resilience Indicators for Benchmarking Baseline Conditions. *Journal of Homeland Security and Emergency Management, 7*(1). doi: Artn 51

Cutter, S. L., & Finch, C. (2008). Temporal and spatial changes in social vulnerability to natural hazards. *Proceedings of the National Academy of Sciences, 105*(7), 2301-2306.

DANE. (2010). *Colombia. Proyecciones De Población Por Área 2005 - 2020*.

Davis, M. (1997). The Radical Politics of Shade. *Capitalism, Nature, Socialism, 8*, 35-39.

DeCoster, J. (2004). Data Analysis in SPSS. http://www.stat-help.com/notes.html

Delmelle, E., Dony, C., Casas, I., Jia, M., & Tang, W. (2014). Visualizing the Impact of Space-Time Uncertainties on Dengue Fever Patterns. *International Journal of Geographical Information Science, 28*(5), 1107-1127.

Dolney, T. J., & Sheridan, S. C. (2006). The relationship between extreme heat and ambulance response calls for the city of Toronto, Ontario, Canada. *Environmental Research, 101*, 94-103.

Donoghue, E. R., Graham, M. A., Jentzen, J. M., Lifschultz, B. D., Luke, J. L., & Mirchandani, H. G. (1997). Criteria for the diagnosis of heat-related deaths: National Association of Medical Examiners: position paper. *The American journal of forensic medicine and pathology, 18*(1), 11-14.

Duneier, M. (2006). Ethnography, the ecological fallacy, and the 1995 Chicago heat wave. *American Sociological Review, 71*(4), 679-688.

Ebi, K. L., Teisberg, T. J., Kalkstein, L. S., Robinson, L., & Weiher, R. F. (2003). Heat Watch/Warning Systems Save Lives: Estimated Costs and Benefits for Philadelphia 1995-98. *Bulletin of the American Meteorological Society, 85*, 1067-1073.

Esri. (2016a). *ArcGIS for Desktop* Tools, Retrieved from http://desktop.arcgis.com/en/arcmap/10.3/main/tools

Esri. (2016b). How Emerging Hot Spot Analysis Works. *ArcGIS PRO - Tool Reference*. http://pro.arcgis.com/en/pro-app/tool-reference/space-time-pattern-mining/emerginghotspots.htm

Fuller, D., Troyo, A., & Beier, J. (2009). El Nino Southern Oscillation and Vegetation Dynamics as Predictors of Dengue Fever Cases in Costa Rica. *Environmental Research Letters, 4*(1), 014011.

Greiling, A. D., Jacquez, M. G., Kaufmann, M. A., & Rommel, G. R. Space-time visualization and analysis in the Cancer Atlas Viewer. *Journal of Geographical Systems, 7*(1), 67-84. doi: 10.1007/s10109-005-0150-y

Gubler, D. J. (1998). Dengue and Dengue Hemorrhagic Fever. *Clinical microbiology reviews, 11*(3), 480-496.

Hajian-Tilaki, K. (2013). Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian Journal of Internal Medicine, 4*(2), 627-635.

Hales, S., Edwards, S. J., & Kovats, R. S. (2003). Impacts on Health of Climate Extremes. In A. J. McMichael, D. H. Campbell-Lendrum, C. F. Corvalan, K. L. Ebi, A. K. Githeko, J. D. Scheraga & A. Woodward (Eds.), *Climate Change and Human Health: Risk and Responses* (pp. 79 - 102). Geneva: World Health Organizatoin.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*(1), 29-36. doi: doi:10.1148/radiology.143.1.7063747

Harlan, S. L., Brazel, A. J., Prashad, L., Stefanov, W. L., & Larsen, L. (2006). Neighborhood microclimates and vulnerability to heat stress. *Social Science & Medicine, 63*(11), 2847-2863.

Harlan, S. L., Declet-Barreto, J. H., Stefanov, W. L., & Petitti, D. B. (2013). Neighborhood Effects on Heat Deaths: Social and Environmental Predictors of Vulnerability in Maricopa County, Arizona. *Environmental Health Perspectives (Online), 121*(2), 197.

Hattis, D., Ogneva-Himmelberger, Y., & Ratick, S. (2012). The spatial variability of heat-related mortality in Massachusetts. *Applied Geography, 33*(1), 45-52. doi: Doi 10.1016/J.Apgeog.2011.07.008

Health, M. C. P. (2015). *Community Assessment for Public Health Emergency Response (CASPER) Heat Vulnerability and Emergency Preparedness Needs Assessment Maricopa County, Arizona*. Retrieved from http://www.maricopa.gov/publichealth/Services/EPI/pdf/heat/Special/2015-CASPER-Report.pdf.

Hondula, D. M., Davis, R. E., Leisten, M. J., Saha, M. V., Veazey, L. M., & Wegner, C. R. (2012). Fine-scale spatial variability of heat-related mortality in Philadelphia County, USA, from 1983–2008: a case-series analysis. *Environ Health, 11*(16), 1-11.

idre. (2016). Annotated SPSS Output - Principal Components Analysis. Retrieved 4/12/2016, 2016, from http://statistics.ats.ucla.edu/stat/spss/output/principal_components.htm

Jackson, R., & Shields, K. N. (2008). Preparing the US health community for climate change. *Annual Review of Public Health, 29*, 57-+.

Jensen, J. (2005). *Introductory Remote Sensing - a remote sensing perspective* (3rd ed.). Upper Saddle River: Pearson - Prentice Hall.

Jerrett, M., Gale, S., & Kontgis, C. (2010). Spatial modeling in environmental and public health research. *International journal of environmental research and public health, 7*(4), 1302-1329.

Johnson, D., Lulla, V., Stanforth, A., & Webber, J. (2011). Remote Sensing of Heat-Related Health Risks: The Trend Toward Coupling Socioeconomic and Remotely Sensed Data. *Geography Compass, 5*(10), 767-780.

Johnson, D., Wilson, J., & Luber, G. (2009). Socioeconomic indicators of heat-related health risk supplemented with remotely sensed data. *International Journal of Health Geographics, 8*(1), 57.

Johnson, D. P., Stanforth, A., Lulla, V., & Luber, G. (2012). Developing an applied extreme heat vulnerability index utilizing socioeconomic and environmental data. *Applied Geography, 35*(1), 23-31.

Johnson, D. P., Webber, J. J., Urs Beerval Ravichandra, K., Lulla, V., & Stanforth, A. C. (2014). Spatiotemporal variations in heat-related health risk in three Midwestern US cities between 1990 and 2010. *Geocarto International, 29*(1), 65-84.

Johnson, D. P., & Wilson, J. S. (2009). The socio-spatial dynamics of extreme urban heat events: The case of heat-related deaths in Philadelphia. *Applied Geography, 29*(3), 419-434.

Kaiser, H. F. (1960). The Application of electronic computers to factor analysis. *Education and Psychological Measurements, 20*, 141-150.

Kalkstein, L. S. (1991). A New Approach to Evaluate the Impact of Climate on Human Mortality. *Environmental Health Perspectives, 96*, 145-150.

Kalkstein, L. S. (1993). Direct impacts in cities. *Lancet, 342*(8884), 1397.

Kalkstein, L. S., & Greene, J. S. (1997). An Evaluation of Climate/Mortality Relationships in Large U.S. Cities and the Possible Impacts of a Climate Change. *Environmental Health Perspectives, 105*(1), 84-93.

Kalkstein, L. S., Greene, S., Mills, D. M., & Samenow, J. (2011). An evaluation of the progress in reducing heat-related human mortality in major U.S. cities. *Natural Hazards, 56*(1), 113-129. doi: Doi 10.1007/S11069-010-9552-3

Kalkstein, L. S., Jamason, P. F., Greene, J. S., Libby, J., & Robinson, L. (1996). The Philadelphia Hot Weather-Health Watch/Warning System: Development and Application, Summer 1995. *Bulletin of the American Meteorological Society, 77*(7), 1520-1528.

Kanaroglou, P., & Delmelle, E. (2015). *Spatial Analysis in Health Geography*: Routledge.

Kraemer, M. U., Sinka, M. E., Duda, K. A., Mylne, A. Q., Shearer, F. M., Barker, C. M., . . . Van Bortel, W. (2015). The Global Distribution of the Arbovirus Vectors *Aedes aegypti* and *Ae. albopictus*. *Elife, 4*, e08347.

Kulldorff, M., Heffernan, R., Hartman, J., Assunçao, R., & Mostashari, F. (2005). A space–time permutation scan statistic for disease outbreak detection. *PLoS Med, 2*(3), e59.

Kurland, K. S., & Gorr, W. L. (2012). *GIS Tutorial for Health* (fourth ed.). Redlands, CA: Esri Press.

Kwan, M.-P. (2012a). The uncertain geographic context problem. *Annals of the Association of American Geographers, 102*(5), 958-968.

Kwan, M.-P. (2012b). How GIS can help address the uncertain geographic context problem in social science research. *Annals of GIS, 18*(4), 245-255.

Li, K., & Yu, Z. (2008). Comparative and Combinative Study of Urban Heat island in Wuhan City with Remote Sensing and CFD Simulation. *Sensors, 8*(10), 6692-6703. doi: Doi 10.3390/S8106692

Li, Z., Roux, E., Dessay, N., Girod, R., Stefani, A., Nacher, M., . . . Seyler, F. (2016). Mapping a Knowledge-Based Malaria Hazard Index Related to Landscape Using Remote Sensing: Application to the Cross-Border Area between French Guiana and Brazil. *Remote Sensing, 8*(4), 319.

Lin, C.-H., & Wen, T.-H. (2011). Using geographically weighted regression (GWR) to explore spatial varying relationships of immature mosquitoes and human densities with the

incidence of dengue. *International journal of environmental research and public health, 8*(7), 2798-2815.
Liu, H., Weng, Q., & Gaines, D. (2008). Spatio-Temporal Analysis of the Relationship between WNV Dissemination and Environmental Variables in Indianapolis, USA. *International Journal of Health Geographics, 7*(1), 1.
Lo, C. P., & Quattrochi, D. A. (2003). Land-use and land-cover change, urban heat island phenomenon, and health implications: A remote sensing approach. *Photogrammetric Engineering and Remote Sensing, 69*(9), 1053-1063.
Lozano-Fuentes, S., Hayden, M. H., Welsh-Rodriguez, C., Ochoa-Martinez, C., Tapia-Santos, B., Kobylinski, K. C., . . . Monaghan, A. J. (2012). The Dengue Virus Mosquito Vector *Aedes aegypti* at High Elevation in Mexico. *The American journal of tropical medicine and hygiene, 87*(5), 902-909.
Luber, G., & McGeehin, M. (2008). Climate Change and Extreme Heat Events. *American Journal of Preventive Medicine, 35*(5), 429-435. doi: DOI 10.1016/j.amepre.2008.08.021
Mathian, H., & Sanders, L. (2014). *Spatio-temporal Approaches: Geographic Objects and Change Process*: John Wiley & Sons.
McCabe, M. F., Wang, L., & Lu, X. (2016). Elevated CO2 as a driver of global dryland greening.
McMichael, A. J., Wilkinson, P., Kovats, R. S., Pattenden, S., Hajat, S., Armstrong, B., . . . Nikoforov, B. (2008). International Study of Temperature, heat and urban mortality: The 'ISOTHURM' project. *International Journal of Epidemiology, 1*(1-11).
Miron, I. J., Montero, J. C., Criado-Alvarez, J. J., Diaz, J., & Linares, C. (2010). Effects of temperature extremes on daily mortality in Castile-La Mancha (Spain): trends from 1975 to 2003. *Gaceta Sanitaria, 24*(2), 117-122. doi: Doi 10.1016/J.Gaceta.2009.10.016
Moreno-Madriñán, M. J., Crosson, W. L., Eisen, L., Estes, S. M., Estes Jr, M. G., Hayden, M., . . . Zielinski-Gutierrez, E. (2014). Correlating Remote Sensing Data with the Abundance of Pupae of the Dengue Virus Mosquito Vector, Aedes aegypti, in Central Mexico. *International Journal of Geo-Information, 3*(2), 8. doi: 10.3390/ijgi3020732
Morin, C. W., Comrie, A. C., & Ernst, K. (2013). Climate and dengue transmission: evidence and implications. *Environmental Health Perspectives (Online), 121*(11-12), 1264.
Mylne, A. Q., Pigott, D. M., Longbottom, J., Shearer, F., Duda, K. A., Messina, J. P., . . . Hay, S. I. (2015). Mapping the zoonotic niche of Lassa fever in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene, 109*(8), 483-492.
NASA. (2015, December 17, 2015). TRMM 3B42 daily Product Summary. *Mirador - Data Access Made Simple.* Retrieved July 20, 2016, from http://mirador.gsfc.nasa.gov/collections/TRMM_3B42_daily__007.shtml
NASA. (2016a). MODIS. Retrieved 4/9/2016, 2016, from http://modis.gsfc.nasa.gov/
NASA. (2016b, May 18, 2016). TRMM - Tropical Rainfall Measuring Mission. Retrieved May 23, 2016, 2016, from http://trmm.gsfc.nasa.gov/
Naughton, M. P., Henderson, A., Mirabelli, M. C., Kaiser, R., Wilhelm, J. L., Kieszak, S. M., . . . McGeehin, M. (2002). Heat-Related Mortality During a 1999 Heat wave in Chicago. *American Journal of Preventive Medicine, 22*, 221-227.
O'Neill, M. S., Zanobetti, A., & Schwartz, J. (2005). Disparities by Race in Heat-Related Mortality in Four US Cities: The Role of Air Conditioning Prevalence. *Journal of Urban Health, 82*(2), 191-197.
Pantavou, K., Theoharatos, G., Mavrakis, A., & Santamouris, M. (2011). Evaluating thermal comfort conditions and health responses during an extremely hot summer in Athens. *Building and Environment, 46*(2), 339-344. doi: Doi 10.1016/J.Buildenv.2010.07.026

Papanastasiou, D., Melas, D., Bartzanas, T., & Kittas, C. (2010). Temperature, comfort and pollution levels during heat waves and the role of sea breeze. *International Journal of Biometeorology, 54*(3), 307-317. doi: Doi 10.1007/S00484-009-0281-9

Patz, J. A., McGeehin, M. A., Bernard, S. M., Ebi, K. L., Epstein, P. R., Grambsch, A., . . . Trtanj, J. (2000). The potential health impacts of climate variability and change for the United States: Executive summary of the report of the health sector of the US National Assessment. *Environmental Health Perspectives, 108*(4), 367-376.

Pinto, E., Coelho, M., Oliver, L., & Massad, E. (2011). The Influence of Climate Variables on Dengue in Singapore. *International journal of environmental health research, 21*(6), 415-426.

Reid, C. E., Mann, J. K., Alfasso, R., English, P. B., King, G. C., Lincoln, R. A., . . . West, N. L. (2012). Evaluation of a heat vulnerability index on abnormally hot days: an environmental public health tracking study. *Environmental Health Perspectives, 120*(5), 715.

Reid, C. E., O'Neill, M. S., Gronlund, C. J., Brines, S. J., Brown, D. G., Diez-Roux, A. V., & Schwartz, J. (2009). Mapping Community Determinants of Heat Vulnerability. *Environmental Health Perspectives, 117*(11), 1730-1736. doi: Doi 10.1289/Ehp.0900683

Rinner, C., Patychuk, D., Bassil, K., Nasr, S., Gower, S., & Campbell, M. (2010). The Role of Maps in Neighborhood-level Heat Vulnerability Assessment for the City of Toronto. *Cartography and Geographic Information Science, 37*(1), 31-44.

Robinson, P. J. (2001). On the Definition of a Heat Wave. *Journal of Applied Meteorology, 40*, 762-775.

Rossen, L. M., Khan, D., & Warner, M. (2014). Hot spots in mortality from drug poisoning in the United States, 2007–2009. *Health & Place, 26*, 14-20.

Ruddell, D. M., Harlan, S. L., Grossman-Clarke, S., & Buyantuyev, A. (2009). Risk and exposure to extreme heat in microclimates of Phoenix, AZ *Geospatial techniques in urban hazard and disaster analysis* (pp. 179-202): Springer.

Rueda, L., Patel, K., Axtell, R., & Stinner, R. (1990). Temperature-dependent development and survival rates of Culex quinquefasciatus and Aedes aegypti (Diptera: Culicidae). *Journal of medical entomology, 27*(5), 892-898.

Ruiz-López, F., González-Mazo, A., Vélez-Mira, A., Gómez, G. F., Zuleta, L., Uribe, S., & Vélez-Bernal, I. D. (2016). Presence of Aedes (Stegomyia) aegypti (Linnaeus, 1762) and its natural infection with dengue virus at unrecorded heights in Colombia. *Biomédica, 36*(2), 303-308.

Salano, R., Didan, K., Jacobson, A., & Huete, A. (2010). *MODIS Vegetation Index User's Guide (MOD13 Series)* (pp. 42). Retrieved from https://vip.arizona.edu/documents/MODIS/MODIS_VI_UsersGuide_01_2012.pdf

Salud, I. N. D. (2016). *Vigilancia Rutinaria*. Retrieved from: http://www.ins.gov.co/lineas-de-accion/Subdireccion-Vigilancia/sivigila/Paginas/vigilancia-rutinaria.aspx

Schwartz, J. (2005). Who is Sensitive to Extremes of Temperature. *Epidemiology, 16*, 67-72.

Semenza, J. C., Rubin, C. H., Falter, K. H., Selanikio, J. D., Flanders, H. L., Howe, H. L., & Wilhelm, J. L. (1996). Heat-related deaths during the July 1995 Heat Wave in Chicago. *The New England Journal of Medicine, 335*, 84-90.

Shen, T. F., Howe, H. L., Alo, C., & Moolenaar, R. L. (1998). Toward a broader definition of heat-related death: Comparison of mortality estimates from medical examiners' classification with those from total death differentials during the July 1995 heat wave in Chicago, Illinois. *American Journal of Forensic Medicine and Pathology, 19*(2), 113-118.

Shimazaki, H., & Shinomoto, S. (2007). A method for selecting the bin size of a time histogram. *Neural computation, 19*(6), 1503-1527.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis* (Vol. 26): CRC press.
Smoyer, K. E. (1998). A comparative analysis of heat waves and associated mortality in St. Louis, Missouri - 1980 and 1995. *International Journal of Biometeorology, 42*(1), 44-50.
Stanforth, A. C. (2011). *Identifying Variations of Socio-spatial Vulnerability to Heat-related Mortality During the 1995 Extreme Heat Event in Chicago, IL, USA.* faculty of the University Graduate School in partial fulfillment of the requirements for the degree Master of Science in the Department of Geography, Indiana University.
StatSoft, I. (2015). Electronic Statistics Textbook. Retrieved December 2015, from StatSoft http://www.statsoft.com/textbook/.
Stone, B., Hess, J. J., & Frumkin, H. (2010). Urban Form and Extreme Heat Events: Are Sprawling Cities More Vulnerable to Climate Change Than Compact Cities? *Environmental Health Perspectives, 118*(10), 1425-1428. doi: Doi 10.1289/Ehp.0901879
Tan, J. G., Zheng, Y. F., Tang, X., Guo, C. Y., Li, L. P., Song, G. X., . . . Chen, H. (2010). The urban heat island and its impact on heat waves and human health in Shanghai. *International Journal of Biometeorology, 54*(1), 75-84. doi: Doi 10.1007/S00484-009-0256-X
Tape, T. G. (*Interpreting Diagnostic Tests* Retrieved from http://gim.unmc.edu/dxtests/roc3.htm
U.S.Census. (2000). *Census 2000, Summary File 3 ; generated by Austin Stanforth; using American FactFinder; http://factfinder2.census.gov. (2 December 2014).*
U.S.Census. (2011). *TIGER/Line® Shapefiles and TIGER/Line® Files*. 2011 TIGER/Line Street Centerline shapefile. Retrieved from: https://www.census.gov/geo/maps-data/data/tiger-line.html
U.S.Census. (2015, June 10, 2015). American Community Survey (ACS): Comparing 2010 American Community Survey Data. from https://www.census.gov/programs-surveys/acs/guidance/comparing-acs-data/2010.html
U.S.Census. (2016a). *Census 2010, ACS 5 YR; generated by Austin Stanforth; using American FactFinder; http://factfinder2.census.gov. (22 March 2016).*
U.S.Census. (2016b). *Census 2010, Decadal Summary File 1 and Summary File 3; generated by Austin Stanforth; using American FactFinder; http://factfinder2.census.gov. (22 March 2016).*
UCLA. (*Introduction to SAS* Retrieved from http://www.ats.ucla.edu/stat/SPSS/output/principal_components.htm
Uejio, C. K., Wilhelmi, O. V., Golden, J. S., Mills, D. M., Gulino, S. P., & Samenow, J. P. (2011). Intra-urban societal vulnerability to extreme heat: The role of heat exposure and the built environment, socioeconomics, and neighborhood stability. *Health & Place, 17*(2), 498-507. doi: Doi 10.1016/J.Healthplace.2010.12.005
USGS. (2016). *Remote Sensing Data*. Retrieved from: http://earthexplorer.usgs.gov/
Venkatesan, M., & Rasgon, J. L. (2010). Population genetic data suggest a role for mosquito-mediated dispersal of West Nile virus across the western United States. *Molecular Ecology, 19*(8), 1573-1584.
Vogelmann, J. E., Howard, S. M., Yang, L., Larson, C. R., Wylie, B. K., & Driel, J. N. V. (2001). Completion of the 1990's National Land Cover Data Set for the conterminous United States. *Photogrammetric Engineering and Remote Sensing*(67), 650-662.
Voogt, J. A., & Oke, T. R. (2003). Thermal remote sensing of urban climates. *Remote Sensing of Environment, 86*, 370-384.
Wang, W. W., Zhu, L. Z., & Wang, R. C. (2004). An analysis on spatial variation of urban human thermal comfort in Hangzhou, China. *Journal of Environmental Sciences-China, 16*(2), 332-338.

Weng, Q., & Lu, D. (2008). A sub-pixel analysis of urbanization effect on land surface temperature and its interplay with impervious surface and vegetation coverage in Indianapolis, United States. *International Journal of Applied Earth Observation and Geoinformation, 10*(1), 68-83. doi: 10.1016/j.jag.2007.05.002

Weng, Q. H., Lu, D. S., & Schubring, J. (2004). Estimation of land surface temperature-vegetation abundance relationship for urban heat island studies. *Remote Sensing of Environment, 89*(4), 467-483. doi: DOI 10.1016/j.rse.2003.11.005

Weng, Q. H., & Quattrochi, D. A. (2006). Thermal remote sensing of urban areas: An introduction to the special issue. *Remote Sensing of Environment, 104*, 119-122.

Whitman, S., Good, G., Donoghue, E. R., Benbow, N., Shou, W., & Mou, S. (1997). Mortality in Chicago attributed to the July 1995 heat wave. *American Journal of Public Health, 87*(9), 1515-1518.

WHO. (2016, May 2015). Dengue and Severe Dengue. *Fact Sheets.* Retrieved February 2016, 2016, from http://www.who.int/mediacentre/factsheets/fs117/en/

Wilhelmi, O. V., & Hayden, M. H. (2010). Connecting people and place: a new framework for reducing urban vulnerability to extreme heat. *Environmental Research Letters, 5*(1), -.

Wisner, B. (2004). *At risk : natural hazards, people's vulnerability, and disasters* (2nd ed.). London ; New York: Routledge.

Woodward, M. (2013). *Epidemiology: study design and data analysis*: CRC press.

Xiao, R. B., Ouyang, Z. Y., Zheng, H., Li, W. F., Schienke, E. W., & Wang, X. K. (2007). Spatial pattern of impervious surfaces and their impacts on land surface temperature in Beijing, China. *Journal of Environmental Sciences-China, 19*(2), 250-256.

YourWeatherService. (2016). Climate Phoenix - Arizona. Retrieved March 22, 2016, 2016, from http://www.usclimatedata.com/climate/phoenix/arizona/united-states/usaz0166/2010/1

Yuan, F., & Bauer, M. E. (2007). Comparison of impervious surface area and normalized difference vegetation index as indicators of surface urban heat island effects in Landsat imagery. *Remote Sensing of Environment, 106*(3), 375-386. doi: DOI 10.1016/j.rse.2006.09.003

Zhang, S., & Zhao, J. (2015). Spatio-Temporal Epidemiology of Hand, Foot and Mouth Disease in Liaocheng City, North China. *Experimental and therapeutic medicine, 9*(3), 811-816.

Zhang, X., Zhong, T., Wang, K., & Cheng, Z. (2009). Scaling of Impervious Surface Area and Vegetation as Indicators to Urban Land Surface Temperature Using Satellite Data. *International Journal of Remote Sensing, 30*(4), 841-859. doi: Doi 10.1080/01431160802395219

Zhang, Y. S., Odeh, I. O. A., & Han, C. F. (2009). Bi-Temporal Characterization of Land Surface Temperature in Relation to Impervious Surface Area, NDVI and NDBI, using a Sub-Pixel Image Analysis. *International Journal of Applied Earth Observation and Geoinformation, 11*(4), 256-264. doi: DOI 10.1016/j.jag.2009.03.001

Zhou, Y., & Shepherd, J. M. (2010). Atlanta's urban heat island under extreme heat conditions and potential mitigation strategies. *Natural Hazards, 52*(3), 639-668. doi: Doi 10.1007/S11069-009-9406-Z

CURRICULUM VITAE

Austin Curran Stanforth

EDUCATION

Indiana University, Indianapolis, IN — *September 2016*
Doctor of Philosophy: Applied Earth Sciences — *Dissertation Defended: May 2, 2016*
Minor: Epidemiology — *Candidate: September 2014*

Indiana University, Indianapolis, IN — *June 2011*
Master of Science: Geographic Information Science

Butler University, Indianapolis, IN — *May 2007*
Bachelor of Arts: Anthropology
Minor: Biology and Psychology

HONORS AND AWARDS

- IUPUI's Elite 50 – Premier 10, and Best Student in the School of Science awards 2015
- American Meteorological Society Annual Conference Student Presentation Award 2015
  - Tenth Symposium on Societal Application
- Graduate and Professional Student Government – Vice President elected 2014
- Golden Key International Honour Society *2009*
- Lesley A. Sharp Award for Excellence in Original Field Research *2007*
- Sigma Nu Fraternity
- Kappa Kappa Psi Service Fraternity
- Eagle Scout

TEACHING EXPERIENCE

Sci 100 Earth Science with Laboratory – Ivy Tech Community College Fall 2015
Geog 100 Geography of World Regions – FSU Fall 2016
Geog 108 Physical Systems of the Environment: Laboratory – IUPUI Spring 2014
Geog 112 Cultural Geography – FSU Fall 2016
Geog 130 World Geography – IUPUI Fall 2015
Geog 303 Weather and Climate – IUPUI Fall 2015
Geog 311 Social Aspects of GIS – FSU Fall 2016
Geog 338/539 Geographic Information Science – Online TA – IUPUI Spring 2016
Geog 536 Advanced Remote Sensing – TA – IUPUI Spring 2016
Geog 602 Topical Seminar in Climate, Land, & Environmental Change –IUPUI Fall 2015

PROFESSIONAL EXPERIENCE

Ferris State University, Big Rapids, MI
Adjunct Lecturer *August 2016 –*

- Performed duties as the instructor of courses for the department including preparing lecture and presentation material, prepare labs, grade assignments and exams, and perform office hours

Indiana University – Purdue University Indianapolis, Indianapolis IN
Fairbanks School of Public Health –
Research Assistant *Oct 2015 – July 2016*

- Geospatial and Spatial Statistic Analysis
- Tropical Disease modeling in Colombia – Focus on Dengue Fever and Zika
- Identification of *Aedis aegypti* vector infection sites

Indiana University – Purdue University Indianapolis, Indianapolis IN
Department of Geography –
Graduate Assistant / Teaching Assistant *Oct 2015 – July 2016*

- Department Teaching Assistant for courses in GIS, Spatial Modeling, and Remote Sensing of Human Geography topics

Indiana University – Purdue University Indianapolis, Indianapolis IN
Lecturer *Jan 2014 – Dec 2015*

- Performed duties as the instructor of courses for the department including preparing lecture and presentation material, prepare labs, grade assignments and exams, and perform office hours

Ivy Tech Community College, Indianapolis IN
Lecturer *Aug 2015 – Dec 2015*

- Performed duties as the instructor of courses for the department including preparing lecture and presentation material, prepare labs, grade assignments and exams, and perform office hours

NASA Goddard Space Flight Center, Greenbelt, MD
Research Intern *- Universities Space Research Association June – August 2015*

- GIS – Geospatial Analysis, Landscape Pattern Metrics, Spatial Analysis
- Remote Sensing – Earth Observations and cross sensor analysis (Commercial, MODIS, Landsat)
- Oral and Poster presentations

Institute for Research on Social Issues, Indianapolis, IN
Research Assistant *July 2012 – December 2015*

- GIS – Geospatial Analysis
- Research in Human Geography, Disaster, Vulnerability, Environmental Hazards, Public Health
- Work with aerial and satellite imagery to identify, digitize and map different land cover/land uses
- Conduct multispectral image interpretation and analysis
- Spatial Statistics and predictive modeling of environmental variables
- Create vulnerability models of social and spatial vulnerability indicators for research areas
- Create buffers, statistics, and maps in ESRI ArcMap to show analysis of data and project status
- Coordinate with end users
- Design and create presentations and publications of original work
- Laboratory Supervisor

The Polis Center

Graduate Research Intern *July - August 2014*

- Assist Domestic Violence Network initiative – reporting and characteristic analysis
- Spatial Statistics and predictive modeling of neighborhood socioeconomic variables
- Conducted project specific application of software and editing of GIS/map layers
- Prepare and present presentation materials

DEVELOP – NASA Internship

Contract Research Intern *June - August 2013*

- Design and implement research project to address ability of satellite data to identify impact of the invasive Hemlock Woolly Adelgid on the Great Smoky Mountains National Park's native Hemlock trees
- Produced end-user driven deliverables and GIS tools
- Created multiple media products with results, including: presentation poster and slides, technical report, fact sheets, electronic media articles, tutorials, and site maps.
- Worked with Satellite imagery and GIS shapefiles and geodatabases

Department of Geography at IUPUI, Indianapolis, IN

Graduate Research Assistant *February 2008 – June 2011*

- Worked with aerial and satellite imagery to identify, digitize and map different land cover/land uses
- Conducted image interpretation and analysis
- Utilized ERDAS Imagine for multispectral image investigation
- Spatial Statistics and predictive modeling of environmental variables
- Created network files
- Analyzed Origin/Destination Cost Matrixes
- Conducted field research to support ongoing projects
- Created buffers, statistics and maps in ArcMap to show analysis of data and project status
- Used Garmin GPS tracking methods and converted waypoints into Shapefiles
- Design and create presentations and publications of original research

Center for Health Geographics, The Polis Center, IUPUI, Indianapolis, IN

Graduate Research Assistant *June 2008 – January 2009*

- Geocoded Areas of Interest for research projects
- Network relational analysis
- Integrated satellite imagery to produce land use/land cover maps
- Prepared and interpreted satellite imagery indexes for research
- Created intensity Raster images for analysis
- Database maintenance

Indy Parks and Recreation, Office of Land Stewardship, Indianapolis, IN

GIS Intern *June -December 2008*

- Designed and updated the Trimble GPS data collection system
- Built catalog of plants and parks used by the office within ArcCatalog
- Designed 'Quickforms' to improve field data acquisition on Trimble GPS devices
- Collected GPS field data
- Implemented GPS data into maps and statistics to analyze project efficiency and outsource work contracts
- Created presentations in Power Point for management pertaining to project audits and future project/fund planning
- Created GPS 'How to' documentation and manuals to assist new GPS users
- Mapped re-naturalization projects, exported maps to various programs for visualization
- Performed spatial analysis on re-naturalization projects and vegetation survivability
- Identified statistically superior trees to plant within flood zones for future planting projects

Office of the State Archaeologist, University of Iowa, Iowa City, IA

Field Technician *October – December 2007*

- Perform surface analysis for artifacts
- Excavate for Pre-Historical occupation materials
- Document and map artifacts

CONFERENCE PRESENTATIONS

Stanforth, A. C., Johnson, J. P. "Space and Time—Impacts on Modeling Extreme Heat Warning Systems" AMS Annual Meeting. New Orleans, LA. 13 January 2016. Paper Presentation.

Stanforth, A. C., Van Den Hoek, J. "Comparison of Landscape Pattern across Multiple Scales, Sensors, and Features." Goddard Space Flight Center, Greenbelt, MD. August 2015, Poster Presentation.

Stanforth, A.C., Johnson, D.P. "Comparing the Impact of Spatial Resolution on Modeling Extreme Heat Vulnerability." AMS Annual Meeting. Phoenix, AZ. 6 January 2015. Paper Presentation.

Stanforth, A. C. "Identifying Socio-Spatial Indications of Extreme Heat Risk for Mitigation Practices." AAG Annual Meeting. Tampa, FL. 8 April 2014. Paper Presentation.

Stanforth, A.C. "Advancing mitigation plans through the use of Socio-Spatial modeling of extreme heat events." CUH Annual Meeting – New Dimensions in Urban Health and Action. Indianapolis, IN. 1 April 2014. Paper Presentation.

Stanforth, A. C. "Developing Future Mitigation Practices: The Application of Research Identifying Socio-Spatial Indications of Extreme Heat Risk." AMS 94th Annual Meeting. Atlanta, GA. 5 February 2014. Paper Presentation.

Stanforth, A. C., J. He, L. Craft, P. Narenpitak, X. Li. "Great Smoky Mountains National Park Ecological Forecasting: Utilizing NASA Earth Observations to Monitor Loss of Hemlock Forest and Advance Mitigation Practices Against the Invasive Hemlock Woolly Adelgid." DEVELOP Internship Review. Huntsville, AL. 6 August 2013. Oral Presentation.

Stanforth, A. C., D. P. Johnson. "Utilizing geospatial technologies to assess heat-related health vulnerability." AGU Annual Conference – NASA Session. San Francisco, CA. 3 December 2012. Paper presentation.

Stanforth, A.C., D.P. Johnson. "Using NASA data and models to improve heat watch/warning systems for decision support." NASA Public Health Review. Rhode Island, NJ. 19 September 2012. Review presentation.

D. P. Johnson, Stanforth, A. C., V. Lulla "Identifying heat vulnerable populations across an urban environment, a case study of the 1995 Chicago, IL extreme heat event." Centers for Disease Control and Prevention, Atlanta, GA. May 2011. Research Update presentation

Stanforth, A. C. "Identifying Variations of social-spatial vulnerability to heat-related mortalities during the 1995 extreme heat event in Chicago, IL." The Association of American Geographers Annual Conference. Seattle, WA. 15 April 2011. Paper presentation.

Stanforth, A. C. "Relating socio-spatial variables to heat induced mortalities during the 1995 extreme heat event in Chicago, IL." The Joseph Taylor Symposium. Indianapolis, IN. February 2011. Poster presentation.

Stanforth, A. C., D. P. Johnson, V. Lulla, and J. Webber III. "Relating socio-spatial variables to heat induced mortalities in Chicago, IL during the 1995 extreme heat event." ICEPHI Symposium. Indianapolis, IN. September 2010. Poster presentation.

Stanforth, A. C. "Relating socioeconomic variability to heat-related mortality during the 1995 extreme heat event in Chicago." The Association of American Geographers Annual Conference. Washington, D.C. April 2010. Poster presentation.

Stanforth, A. C., J. Green. "Remote archaeology- An experiment on the implications of remote sensing for archaeological research." IUPUI Scholar Research Conference. Indianapolis, IN. May 2009. Poster presentation.

Stanforth, A. C., S. L. Gidley, J. Green. "Edge detection of historical African American homesteads in Indiana National Forest preserves." IUPUI Scholar Research Conference. Indianapolis, IN. May 2008. Poster presentation.

PUBLICATIONS

Stanforth, A. C., M. J. Moreno-Madriñán, J. Ashby (2016). Exploratory Analysis of Dengue Fever Niche Variables within the Río Magdalena Watershed. *Remote Sensing*, Accepted.

Stanforth, A. C., D. P. Johnson. (2016). Sociospatial Modeling for climate-based emergencies: Extreme Heat Vulnerability Index. S. Steinberg, W. Sprigg (Eds.), *Extreme Weather, Health, and Communities: Interdisciplinary Engagement Strategies*. Springer. ISBN 978-3-319-30626-1.

Lulla, V., A. C. Stanforth, N. Prudent, G. Luber, D. P. Johnson. (2015). Modeling Vulnerable Urban Populations in the Global Context of a Changing Climate. R. Ahn (Ed.), *Innovations, Urbanizations and Global Health*. Springer. ISBN 978-1-4899-7597-3

Johnson, D. P., Stanforth, A. C. (2015). Extreme and Changing Meteorological Conditions on the Human Health Condition. G. Luber (Ed.), *Global Climate Change and Human Health*. Jossey-Bass. ISBN 978-1118505571

Johnson, D. P., Webber, J. J., Urs Beerval Ravichandra, K., Lulla, V., & Stanforth, A. C. (2014). Spatiotemporal variations in heat-related health risk in three Midwestern US cities between 1990 and 2010. *Geocarto International*, *29*(1), 65-84.

Johnson, D. P., V. Lulla, A. C. Stanforth. (2013). Intra-urban variations in vulnerability associated with extreme heat events in relationship to a changing climate. S. C. Pryor (Ed.), *Climate Change in the Midwest: Impacts, Risks, Vulnerability, and Adaptation* (134-145). Bloomington, IN: Indiana University Press.

Johnson, D. P., A. C. Stanforth, V. Lulla, and G. Luber. (2012). Developing an applied extreme heat vulnerability index utilizing socioeconomic and environmental data. *Applied Geography, vol 35, issues 1-2*, 23-31. ISSN 0143-6228, 10.1016/j.apgeog.2012.04.006.

Johnson, D. P., V. Lulla, A. C. Stanforth, and J. Webber. (2011). Remote sensing of heat-related risks: the trend towards coupling socioeconomic and remotely sensed data. *Geography Compass, vol 5, Issue 10*, 767-780. DOI: 10.1111/j.1749-8198.201.00442.x

Stanforth, A. (2011). *Identifying variations of socio-spatial vulnerability to heat-related mortality during the 1995 extreme heat event in Chicago, IL, USA.* MS Thesis. IUPUI: USA. http://hdl.handle.net/1805/2643

PROFESSIONAL MEMBERSHIP

American Meteorological Society	2013-Present
-Board on Environment and Health: Student Representative	2013-Present
American Association of Geographers	2009-Present

SOFTWARE EXPERIENCE

ESRI ArcGIS (Arcmap, ArcCatalog, ArcToolbox, ArcPad)
QGIS
ERDAS Imagine
ENVI
Microsoft Office Suite (Word, Excel, Power Point, Outlook, Access)
SPSS
Fragstats
Experience with a variety of Windows and Microsoft operating systems
Experience with Python, C++, and R