IMPROVING NCEI'S CLIMATE EXTREMES INDEX AND REVISING THE CDC'S SOCIAL VULNERABILITY INDEX TO ANALYZE CLIMATE EXTREMES VULNERABILITY IN THE UNITED STATES

by

EMILY LOUISE PAULINE

(Under the Direction of John Knox)

ABSTRACT

The occurrence of extreme weather and climate events has increased in recent decades. This increasing frequency has adversely impacted economic and health outcomes, leading to an increasingly urgent need to study climate extremes. The National Centers for Environmental Information (NCEI) created the Climate Extremes Index (CEI) in 1996 to quantify climate extremes. In this thesis, the CEI is improved by recalculating it using Z-scores instead of the prior approach of using 10th and 90th percentiles. This provided a more accurate method of quantifying climate extremes while calculating the CEI on a climate division basis. CEI values were then combined with recalculated Social Vulnerability Index (SVI) values to create a new Extremes Vulnerability Index (EVI) that calculates climate extremes vulnerability in the United States. The information contained in the EVI can be used by policymakers to implement policies and changes in infrastructure that mitigate risk in vulnerable climate divisions.

INDEX WORDS: Climate, Extremes, Vulnerability, Indices

IMPROVING NCEI'S CLIMATE EXTREMES INDEX AND REVISING THE CDC'S SOCIAL VULNERABILITY INDEX TO ANALYZE CLIMATE EXTREMES VULNERABILITY IN THE UNITED STATES

By

EMILY LOUISE PAULINE

B.S., Valparaiso University, 2017

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2019

IMPROVING NCEI'S CLIMATE EXTREMES INDEX AND REVISING THE CDC'S SOCIAL VULNERABILITY INDEX TO ANALYZE CLIMATE EXTREMES VULNERABILITY IN THE UNITED STATES

by

EMILY LOUISE PAULINE

Major Professor: John Knox

Committee: Andrew Grundstein
Lynne Seymour

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2019

DEDICATION

For my family, who have supported me throughout every step of my Meteorology journey, including my decision to move to Athens. Nothing I’ve done in my life would be possible without you.

Also, to all those I’ve met during my time in Athens. This thesis has been shaped by all of you.

ACKNOWLEDGEMENTS

There are several individuals that were instrumental in the completion of this project and require mentioning. First and foremost, Karin Gleason for all of her help, guidance, advice, and wonderful conversations about all things Valpo. Your importance in the completion of this project cannot be overstated, and you will always have my heartfelt thanks. Deke Arndt for all of the string pulling to bring me to Asheville, and for sparking the idea for this project. I wouldn't be here at all if you hadn't pointed me to the CEI. Matt Menne, Shannon Fields, Simone, and Jude for welcoming me into your home and your lives while I was in Asheville. Y'all made this whole experience possible for me, and your support and friendship throughout this process have been such a gift. I cannot thank you enough for your kindness.

Many thanks to the Department of Geography at the University of Georgia for providing me with funding to complete this work. I would also like to thank Alison Smith from the College of Environment and Design for her invaluable assistance when making the maps for this project. These maps look amazing, and I have you to thank for that. Finally, I am so grateful to my committee for their never-ending support and assistance throughout the completion of this project. Your guidance helped make this project what it is; I am so thankful for everything you've done for me.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Since the Industrial Revolution of the 19th Century, human activities have caused the phenomenon known as anthropogenic climate change that has disrupted normal weather patterns. This disruption has manifested itself in the increased frequency of temperatures and precipitation that are significantly different from normal values (IPCC Fifth Assessment Summary 2014). These statistically significant departures from normal, whether they are above or below normal, can be used to form the basis for climate extremes. Studying climate extremes is increasingly urgent because the rising number of extreme events has adversely impacted economic and health outcomes (IPCC Fifth Assessment Summary 2014). A key organization working within this field is the National Centers for Environmental Information (NCEI), the U.S. government agency responsible for the collection, interpretation, and dissemination of climate data. In order to educate the public about climate extremes, a Climate Extremes Index (CEI) was developed in the 1990s to quantify the extremes that the country as a whole was experiencing (Karl et al. 1996). However, as it has only been minimally updated since its inception, the use of new statistical techniques to identify values that are extreme could further improve the CEI.

This project applied the Z-score statistic to calculate the CEI on a numerical scale and explored the accuracy of calculating the CEI with this statistic. Connecting climate and human health is an important potential application of this index. This project also created a climate extremes vulnerability index using the results from the updated CEI that describes vulnerability to various health concerns and climatic conditions due to the combination of exposure to climate extremes and demographic characteristics. Chapter 2 provides background information on the CEI, similar indices, issues facing stakeholders when interpreting climate indices, statistical techniques that can be used to quantify climate extremes, and an introduction to the connection between climate and human vulnerability. Chapter 3 introduces the research questions. Chapter 4 details the methodology to improve the CEI and quantify vulnerability to climate extremes. Chapter 5 explains the results of this project, and Chapter 6 concludes the thesis document.

CHAPTER 2

BACKGROUND INFORMATION

A. The CEI and its Strengths and Weaknesses

In 1996, the National Centers for Environmental Information (NCEI) introduced their Climate Extremes Index (CEI) as a way to quantify observed changes in climate within the continental United States (Karl et al. 1996). This index was released in an environment where the idea of climate change was still relatively new to the public, so being able to quantify the extremes helped serve as a way to educate the public about what these changes would mean for them. A minimally updated index, released in 2008, was meant to be used as a tool to help aid policy makers in legislative decisions and educate the general public about changes that would result from climate change (Gleason et al. 2008).

The original index is defined as being the arithmetic average of five indicators of the percentage of the continental United States experiencing extremes of each indicator. As a note, NCEI uses the term “Step” to refer to each indicator in the CEI. The term “Component” is used in this thesis. The five components selected were:

1. Maximum temperatures much below normal and maximum temperatures much above normal

2. Minimum temperatures much below normal and minimum temperatures much above normal

3. Severe drought (representing a severe lack of moisture) and severe moisture surplus

4. Greater than normal amounts of extreme 1-day precipitation totals, measured by comparing the proportion of daily rainfall to yearly total across all grid points. In the original CEI, a grid receiving 2 inches or 50.8 millimeters of precipitation on any given day was considered the extreme threshold. This criterion was later amended, and a detailed explanation is available in Gleason et al. 2008.

5. Greater than normal number of days with precipitation and greater than normal number of days without precipitation, used as a way to gauge total number of days with rainfall per analysis period

Data to calculate the current CEI comes from three different datasets. Temperature data (used in the first two components) comes from NCEI's cClimGrid data set. It formerly came from the U.S. Historical Climatology Network (HCN), utilizing only the stations with monthly data that is at least 90% complete (Karl et al. 1990). Precipitation data comes from the Cooperative Summary of the Day (TD3200) dataset and the TD 3206 dataset for all years prior to 1948 (see Gleason et al. 2008 for a more detailed explanation). Data for moisture availability comes in the form of monthly values of the Palmer Drought Severity Index (PDSI) (Palmer 1965).

The first step in calculating the CEI was determining the threshold for "extreme" patterns within each component. It was decided that the extreme portion of each dataset would be the portion falling in the highest-ten percent of the dataset (high extremes), or the lowest-ten percent of the dataset (low extremes). To calculate the final product, the percent of the country that was experiencing extremes in each of the five components was first determined. Once that percentage for each component was established, the final index value was calculated by averaging all percentages together to come up with a total percentage of the United States experiencing extreme conditions:

$$\mathbf{CEI = [(T_{max}\%) + (T_{min}\%) + (Moisture\%) + 2*(1\text{-}day\%) + (Precip\_days\%)] / 5}.$$

Since the calculation is based on the highest and lowest 10% of each indicator, the assumption is that about 20% of the country should be experiencing extreme conditions. Values not equal to this amount indicate a change in the area of the United States experiencing extreme weather patterns (Karl et al. 1996; Gleason et al. 2008).

Although this index serves as a good first step to quantifying climate extremes, there are several problems with this index in its current form. These problems include:

1) The math used to calculate the extreme portion of each dataset is at times questionable and often difficult to comprehend. The assumption that 20% of the US is experiencing extremes in climate is a little simplistic when there are five different indicators included in this index. There may be trends within each indicator that get masked in the calculation of the entire index. In addition, stating that a certain percent of the country is affected by extremes doesn't tell users how extreme a variable is, or what may happen in their area in the future. The index merely indicates that extreme conditions are occurring. Analysis of the top and tail end of a distribution is a very elementary method and not a very robust technique. Assuming that the highest and lowest percentiles represent extreme values does not compare individual values to a mean value, meaning that certain values in the percentiles deemed as extreme may not even be extreme. They just may happen to fall in one of the tail ends of the distribution, but that value itself may be "normal" (Seymour,

personal communication, March 2018). Additionally, the calculation of the value for component #4 involves doubling the top end of the dataset for component #4 to equal the 20% that the CEI assumes is the percent of the country that is experiencing extremes. This is also a very unusual approach to establish a 20% range within a distribution.

2) The original CEI produces a value for the entire country, and a recent update produces a breakdown of the CEI for nine regions within the United States. However, the CEI would greatly benefit from a state or climate division-level breakdown because it would give Americans the chance to see what the trend is for climate extremes in their area and how they will be impacted in the future.

3) On the webpage for the CEI, the issue of ineffective communication also surfaces. The maps and charts (visual tools) used to display the CEI are very hard to interpret (Figure 1), and the section detailing the CEI is hard to find and rather technical for someone without a background in meteorology. Since a very minimal description is given, it is very difficult to interpret the values of the CEI; this can be improved by writing a short description of how to interpret the CEI in practical language so that users can interpret it on their own, regardless of whether or not they have a meteorology or climatology background.

4) The previous three problems culminate in the most pressing problem of the CEI from an applications perspective: the CEI in its present form is unable to be applied to other fields. A tool that quantifies climate extremes is very powerful and can be used to help the general public understand how their lives may be changing and how different factors in their lives will be affected, but this index currently is not doing that. An important way the CEI can be applied is to the field of health. The

connection between climate and health is becoming stronger (Bell 2018, Reid et al. 2009, IPCC 2007 Impacts, Adaptation, and Vulnerability Chapter 8); being able to tell people where certain extremes are occurring and identifying any connections to potential health risks is extremely important. The impact is now extending to the insurance industry. Increasing climate variability and occurrence of extremes can cause deviations from the estimation of future health insurance costs, and this will have "considerable financial consequences" (Owen 2019). Applying the CEI on a regional basis to identify potential regional health risks is the main motivation for this project.

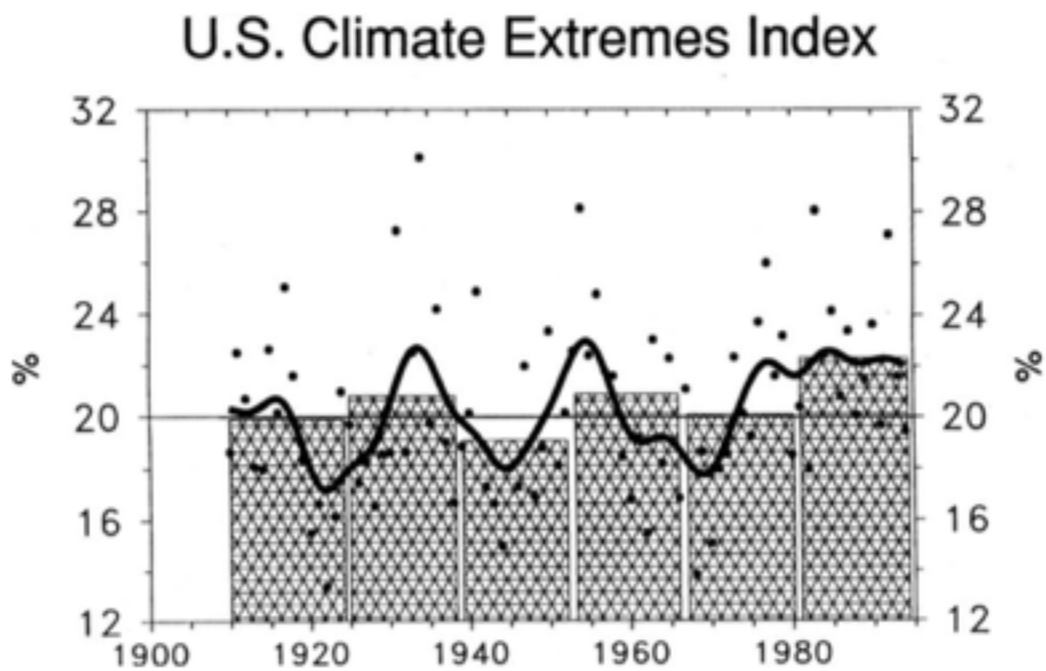

Figure 1: The original CEI from Karl et al. (1996), Figure 13

B. Other Climate Extremes Indices

Since the inception of the CEI, several other indices with the goal of calculating and communicating climate extremes have been developed, and there are many concepts that could be taken from these indices and implemented in the calculation of the CEI to improve it.

1. The ACI

A recent index released was the Actuaries Climate Index (ACI), released by the Society of Actuaries in 2016. This index was designed as a "tool to help inform actuaries, public policymakers, and the general public about climate trends and some of the potential impacts of a changing climate on the United States and Canada" (http://actuariesclimateindex.org). This index was created to be an objective measure of climate trends that actuaries could make use of when assessing risk of a particular location to changes in climate. Thus, the development of this index has placed a high priority on interpretability by non-scientists. The ACI is calculated for all of North America (US and Canada) and also twelve different subregions within North America, and is derived from six components, some similar to those used in the calculation of the CEI:

1. Frequency of temperatures above the 90th percentile
2. Frequency of temperatures below the 10th percentile
3. Maximum rainfall per month in five consecutive days (used to look at multiple day, heavy-precipitation events)
4. Annual maximum consecutive dry days
5. Frequency of wind speed above the 90th percentile
6. Sea level changes

All data come from components of the GHCNDEX dataset, which collects daily values of each variable; the only exception is the data for component 6, which come from tide gauges. To find the values that feed into the final ACI, the mean of each of the six components for a given year is calculated first. These are then turned into standardized anomalies of the reference period, 1961-1990. Using standardized anomalies allows for all of the components to be combined into one index. The final ACI, calculated on a monthly and seasonal (months ending in February, May, August, and November) basis, is calculated by finding the average of the standardized anomalies of the six components (actuariesclimateindex.org):

ACI = mean (T90std - T10std + Pstd + Dstd + Wstd + Sstd).

Rather than presenting the results as a percent, the ACI is expressed as a numerical index of standard deviations from the mean of the reference period of 1961-1990 (Figure 2). A positive number means that the region being examined is trending toward more extreme conditions, as compared to the reference period (actuariesclimateindex.org). To determine the type of extremes (e.g. higher maximum temperatures, more rain, etc.) contributing to the overall index value for a particular location, the user need only check the graphs made of each individual component. The result is an index that can be interpreted by almost any member of the public.

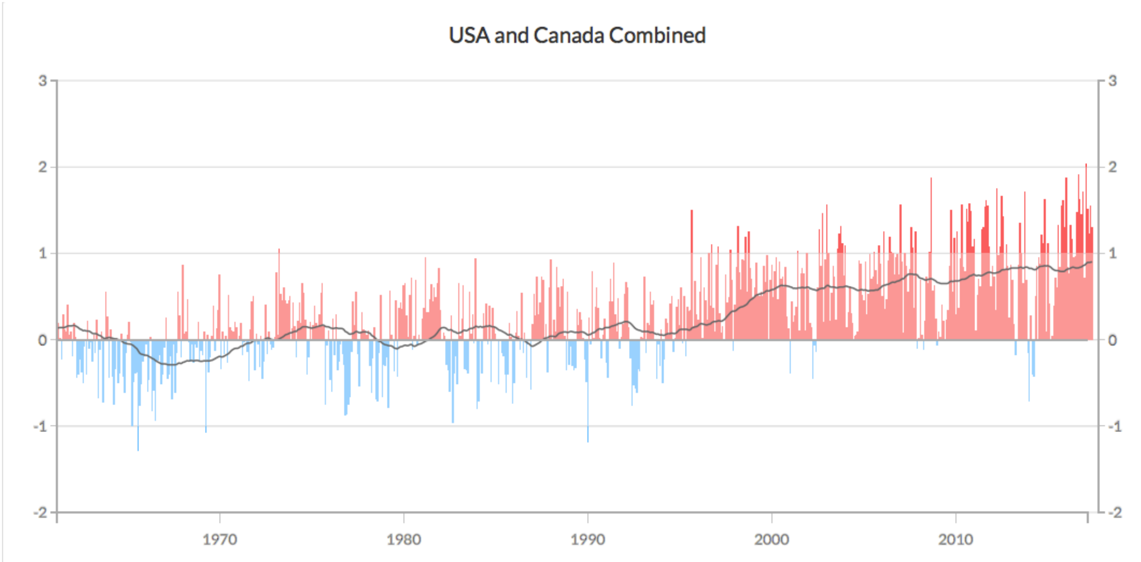

Figure 2: The ACI for the United States and Canada, from actuariesclimateindex.org

2. The AACI

The most recently released index that measures climate extremes is the Australian Actuaries Climate Index. Released in 2018 by the Actuaries Institute (of Australia), the AACI was created "to measure whether the frequency of extreme weather conditions is changing over time" in Australia (https://actuaries.asn.au/microsites/climate-index). The objective of this index is similar to the ACI, in that the goal is to provide primarily actuaries, but also decision-makers and the general public, with an easy-to-interpret climate monitoring tool. Despite a similar objective, nearly every other aspect of the AACI is different from the ACI.

The AACI is derived from six components, some similar to those used in the calculation of the ACI:

1. High Temperature: the monthly frequency of maximum and minimum temperatures above the 99^{th} percentile.

2. Low Temperature: the monthly frequency of maximum and minimum temperatures above the 1st percentile.
3. Precipitation: the monthly frequency of 5-day consecutive rainfall above the 99th percentile.
4. Wind: the monthly frequency of daily wind gusts above the 99th percentile.
5. Consecutive Dry Days: annual maximum consecutive dry days.
6. Sea Level: Monthly maximum sea level.

All data are taken from datasets available through the Bureau of Meteorology (BoM) of Australia. High and Low temperature data come from the ACORN-SAT dataset, precipitation data comes from BoM maintained weather stations that report rain, wind data comes from BoM maintained weather stations, and sea level data comes from 16 tide gauges maintained by the BoM as part of its Baseline Sea Level Monitoring Project (https://actuaries.asn.au/microsites/climate-index).

The values used in the calculation of the AACI are based on the calculation of the frequency of events that exceed a given threshold. To be considered extreme, a component value must exceed the value representing the 99th percentile. To determine the value that represents this threshold, all values for a given day, as well as the five days preceding and following it, between the years 1981-2010 are considered and then the 4th highest value is selected (https://actuaries.asn.au/microsites/climate-index). This value represents the 99th percentile. The monthly frequency of events that exceed this threshold is then calculated for each component. Six individual component indices are produced for each of the components using the method described above, with one value calculated per season (ending with months February, May, August, and November). Each component index produces one value for the entire country of Australia, as well as one value for twelve different subregions within the country.

Additionally, one composite index (the AACI) is also calculated that combines the measurements of high temperature, rainfall, and sea level. To calculate the composite AACI, the values from each individual component are first standardized so that they can be combined. Using a reference period of 1981-2010, the seasonal mean and standard deviation for each component are calculated. These values are then used to calculate the standardized anomaly of each component. After standardizing each component, the values for the three components are then combined by taking the average of the three standardized anomalies to determine the AACI value:

$$\mathbf{AACI = mean\ (HighTemp_{std} + Precip_{std} + SeaLevel_{std})}$$

One value per season (the same seasons defined above) is calculated for the entire country and also twelve subregions. Like the ACI, this index is also presented as a numerical scale of standardized anomalies (Figure 3). A positive number indicates that the frequency of extreme conditions was higher than that of the value for the reference period (1981-2010), and a negative number indicates that the frequency of extreme conditions was lower than that of the reference period (https://actuaries.asn.au/microsites/climate-index). Like the ACI, the result is an index that many people can interpret and understand.

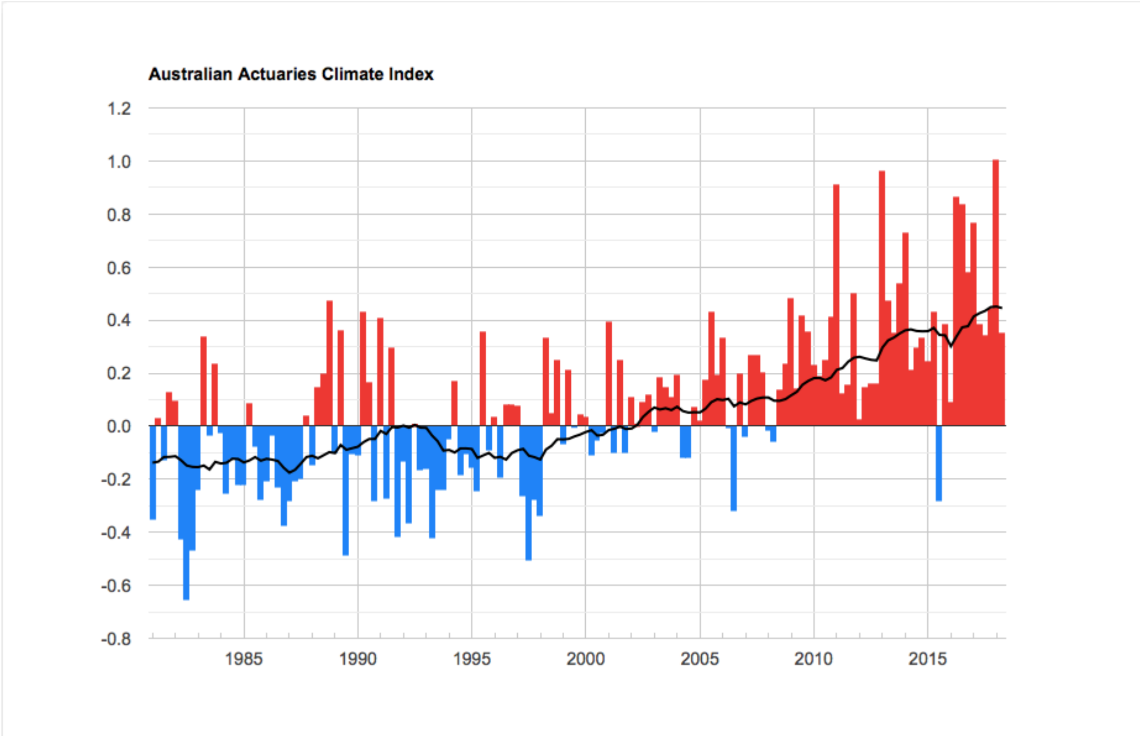

Figure 3: The AACI for Australia, from http://actuaries.asn.au/microsites/climate-index

3. mCEI and dmCEI

Another climate index that has been developed is actually a modification of the CEI, and it has been applied to regions in Australia. The Australian modified CEI (mCEI) and daily modified CEI (dmCEI) were developed to quantify the percent of Australia experiencing climate extremes (Gallant and Karoly 2010). The difference between the two is that the mCEI is based on monthly and annual values, but the dmCEI uses daily data to examine the effects of variables like extreme 1-day precipitation events. The mCEI and dmCEI address two major flaws in the CEI: the lack of a description of the types of extremes occurring, and the lack of a regional breakdown of the CEI to analyze specific components affecting certain regions. So, an index was developed for Australia that was based on the CEI, but that also sought to address these two flaws in the CEI. Both of these indices are derived from the same five components as the CEI:

1. Maximum temperature

2. Minimum temperature

3. Drought/moisture surplus

4. Heavy precipitation (intending to measure the area that experienced a high proportion of extreme precipitation days)

5. Wet/dry days

To calculate the temperature components of the mCEI, monthly averages were used; for the dmCEI, a comparison between the recorded values and hot and cold day thresholds were used. In the calculation of both indices, the lack of analyzing the direction of the extremes is addressed by employing what is referred to as "the subtraction method." In this method, the percentage of Australia below the 10th percentile is subtracted from the percentage of Australia above the 90th percentile. The result is an anomaly that reveals the direction of the extreme of a particular component. For example, a positive number for maximum temperature means that high temperatures are trending toward the upper end of the distribution and extreme temperatures are becoming warmer (Gallant and Karoly 2010). Thus, this directional aspect to the mCEI and dmCEI reveals what type of extremes a region is experiencing.

After calculating the mCEI and the dmCEI the developers of these indices realized that there were discrepancies between the index values for Australia as a whole versus specific regions within the country. In order to understand how four different regions of Australia are affected by different components, both indices were recalculated for each of these four regions. Calculating index values on a regional basis identifies components that impact the overall value of the mCEI and dmCEI, and this can highlight components that will pose more of a threat for individual regions. For example, the mCEI for Australia as a whole reveals a trend towards warmer and wetter patterns, but calculations of regional mCEI values indicates that the southeast and southwest regions are experiencing an increase in areas experiencing dry days (Gallant and Karoly 2010).

C. Applications of Other Indices to the CEI

There are several aspects of the previously discussed climate extremes indices that, if implemented, could greatly improve the calculation and functionality of the CEI. The ability of the ACI to be interpreted by any member of the general public is the most important feature that could be transferred back to the CEI. Changing the CEI to be represented as a number, rather than a percent, would help to increase the interpretability of the index substantially, and this would help make the index accessible to a broader audience. The strength of the AACI that can be used in the calculation of the CEI is the use of the 1981-2010 data to calculate the reference mean and standard deviation values (https://actuaries.asn.au/microsites/climate-index). Since this period is the most current and complete 30-year period available, it is a reliable period to use as a baseline for comparisons, even though the conditions during this period are not necessarily normal because it contains the climate change signal. The mCEI and dmCEI also introduce different methods that would help to improve the CEI. If the CEI continues to be calculated on a percentile basis, a "subtraction method" step similar to that used in the mCEI and dmCEI calculation could also be included within the calculation of the CEI to help users understand what kind of extremes they may experience in the future. Utilizing regional extremes can be applied to the CEI as well, because this will help identify specific patterns that will affect regions and will further help the general public prepare for weather that is to come.

D. Other Statistical Analysis Methods

The one commonality between all indices examined thus far is the statistical technique used to define and analyze extreme values within datasets. Even though the resulting indices all look different, they all stem from using the 90th percentile (top ten percent) of a dataset as high end extremes and the 10th percentile (bottom ten percent) as low-end extremes. The use of "top and tail analysis" is not the best method to use for statistical analysis of temperature or precipitation data. Reasons why can be seen in Figure 4 (courtesy of Dr. Lynne Seymour). This image is the result of ten temperature simulations, with 100 temperature values per simulation. For each simulation, the mean is 75 and the standard deviation is 5. Values in black are those which are less than the 90^{th} percentile and greater than the 10^{th} percentile. Values in red represent values exceeding the 90^{th} percentile, and values in blue represent values less than the 10^{th} percentile. The pink/light blue lines are drawn at +/- 2 standard deviations from the mean, and the red/dark blue lines is drawn at +/- 3 standard deviations. Using the NCEI definition of extreme, the same number of values will be considered to be extreme, regardless of distance from the mean. However, using the top and bottom 10th percentiles as the extremes of the dataset may not correctly identify values that actually are extreme. Because these raw values are not being compared to any value, e.g. a mean value, there can be situations in which a value that falls in the lowest 10^{th} percentile of a data distribution is not actually extreme in a meteorological or climatological sense.

Comparing raw data values to a mean value and selecting a certain distance from that mean to indicate extreme values restricts the number of values of and occurrence of extreme values, which should be rare anyway.

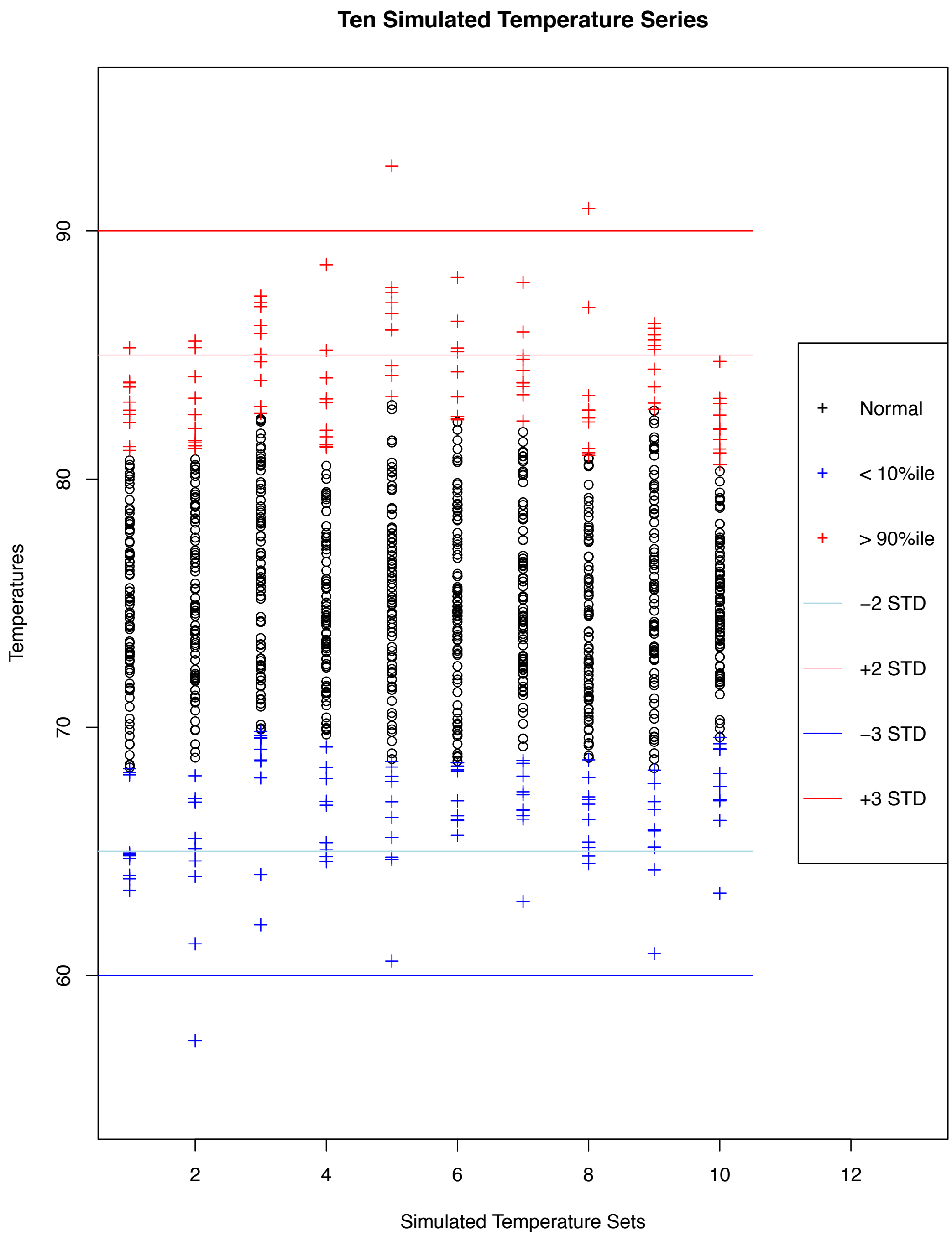

Figure 4: Temperature data simulation results (courtesy of Dr. Lynne Seymour)

Additionally, the distribution of the different datasets used in the calculation of this index are different, as precipitation is not normally distributed (Wilks 1989). In order to be effective and accurate indices, different statistical techniques should be applied to the datasets that derive the components of these indices to account for the irregularities that may be within each dataset.

In order to maintain a similar format when calculating the final index, a statistic is needed that can be used to identify extremes in both temperature and precipitation data. The Z-score is one such statistic (Larson and Farber 2006). A Z-score is mathematically defined as:

$$z = \frac{Value\ -\ Mean}{Standard\ Deviation} \ = \ \frac{x - \ \mu}{\sigma}$$

In this equation, x represents the data value, μ represents the mean of the dataset, and σ represents the standard deviation of the dataset. This statistic calculates the number of standard deviations a data value is from the mean of the dataset it comes from (Larson and Farber 2006). Calculating the number of standard deviations a data point lies from the mean of that dataset has foundations in Chebyshev's Theorem (Shafer and Zhang 2017). It applies to any data distribution, normal or otherwise.

Chebyshev's Theorem mathematically explains the relationship between mean and standard deviation by using the Z-score, which for any data point is just its distance from the mean in units of standard deviations. This theorem describes how all data points should behave with respect to the distance from the mean in terms of standard deviation. Mathematically, Chebyshev's Theorem states that the proportion of any dataset lying within a certain number of standard deviations of the mean is at least:

$$\left(1 - \frac{1}{k^2}\right) * 100\%.$$

In this statement, *k* represents the number of standard deviations. From this statement, this theorem gives a description of the number of data observations that are expected to lie within a certain number of standard deviations of the mean of a data set. When k=2, or two standard deviations, this statement equals $\frac{3}{4}$, which is also equal 75%. This means that at least 75% of the data must lie within two standard deviations of the mean. When k=3 (three standard deviations), the statement equals $\frac{8}{9}$, or 88.9%. This means that at least 88.9% of the data must lie within 3 standard deviations of the mean (Larson and Farber 2006). Since this metric also uses units of standard deviation, Z-scores are actual values of “k”.

Applying this math to the Z-scores of any dataset means that at least 89% of the data lie within 3 standard deviations of the mean; that is, at least 89% of the data will have a Z-score less than 3. This also serves as a statement about data that do not conform to this rule. In the context of this work, any data point with a Z-score larger than three is considered to be extreme. Again, Chebyshev’s Theorem can be applied to any data set, so it can be applied to data sets that are not bell shaped (Shafer and Zhang 2017). This is very relevant to weather and climate data, since oftentimes the frequency curves of various parameters are not bell shaped.

The concept of standard deviations is key to Z-scores and how extremes are determined. The Z-score of a certain data point is the number of standard deviations that a data point is from the mean, as defined by Chebyshev’s Theorem. Z-scores with an absolute value less than 3 mean that the data point falls within 3 standard deviations of the mean. Z-scores with absolute values greater than 3 indicate that the data point is outside of the 3 standard deviations. These are the values that are considered to be extreme because they do not conform to the rules set forth in Chebyshev's Theorem. The higher the Z-score, the more extreme the value.

The NCEI node of the NASA DEVELOP program used Z-scores in 2017 to rank and compare one year's annual temperatures with those of surrounding years. This technique utilized Z-scores to represent annual temperature anomalies to create this ranking system (Yang et al. 2017). The scale was created in three steps:

1. Identify the residuals for a given year's annual temperature values
2. Divide by the group's standard deviation to find the associated Z-scores
3. Transform the Z-scores to create a scale that ranks annual temperatures.

The scale of the transformed Z-scores ranged from 1 to 10, based on cut-off values of a Gaussian distribution, with a 1 representing a cold year and a 10 representing a warm year. Creating a scale and ranking system of temperature anomalies provides context to the rankings that organizations like NASA and NOAA release regarding how warm a year is, e.g. warmest on record. It also provides context to the natural fluctuations seen in temperature and allows for an analysis of natural vs. anthropogenic climate change (Yang et al. 2017). This is one technique that can be implemented to analyze and define extremes that go into the calculation of the CEI.

Two additional methods were considered in the event that using Z-scores did not help the calculation of the CEI. One technique that was considered to analyze temperature extremes was control charts, documented in Lund and Seymour (1999). A new method employed by Mattingly et al. (2017) takes a more "hyperlocal approach" to identifying precipitation extremes that may occur in one region over a certain period of time; it addresses the flaw in precipitation extreme calculations that most extreme precipitation calculations are point-based, meaning that the values at one location are used to make a generalization about a larger region. However, as the Z-score statistic worked the best to quantify climate extremes, neither of these approaches were used in the re-calculation of the CEI.

E. Stakeholder Use, Misuse, and Disuse of Climate Extreme Indices

The CEI has yet to fulfill its stated purpose of serving as a tool that the public can use to interpret changes in climate. From a basic Google search in January of 2018 that produced about 44,000 results, the only webpages that one can find regarding the CEI are the webpages from NCEI that present the index and a website about climate change created by climate scientists. The webpage from NCEI that gives a brief overview of the index is buried deep in the NCEI website and is not easily accessible (https://www.ncdc.noaa.gov/extremes/cei/). There is no framework present within the CEI that makes it accessible to the general public; citations of the CEI are numerous, but they are mainly academic in nature. Current literature acknowledges only that the CEI exists; no one has assessed its effectiveness and usability yet.

One study has attempted to apply the CEI to bank loan contracting (Huang et al. unpublished manuscript). The goal of this study was to analyze and quantify the impact that climate risk has and might continue to have on future bank loan contracts to firms. In addition to including economic variables like total firm assets and loan size in their analysis, these authors also included climate data to analyze how all of these factors combine to impact the types of conditions that a firm will face when trying to get a loan from a bank. The climate data that these authors utilized came in the form of the regional

CEI values between the years 1993-2010. However, there is a serious misinterpretation of this index by the authors that could potentially undermine their research.

After describing the CEI and its components, Huang et al. then state that "the regional CEI is indicative of the severity of climate risk a region faces" (Huang et al. unpublished manuscript). However, the original Karl et al. (1996) article describing the original CEI states that the index was developed "with the goal of summarizing and presenting … changes in climate" (Karl et al. 1996). The original CEI article does not make any mention about intending to make a statement about risk or a particular region's likelihood of experiencing extreme climate. The intent of the CEI is only to summarize and describe the *past and current* exposure of various regions to climate extremes. The CEI is not a tool to be used to make *future* projections. Since exposure is not necessarily indicative of risk, Huang et al. incorrectly assume that the CEI can be used as a measure of risk. As a result, the meaning of the CEI values was also misinterpreted in the construction of the regression model used to determine the significance of climate in the development of loan contracts.

One reason for the misuse (or disuse) of this index may be the way that the original index is presented (see Figure 1). The original index from 1996 is presented in such a way that interpretation is nearly impossible, due to the lack of labels on the graphs intended to model the CEI. The difficulty in interpretation has been acknowledged by NCEI employees, who have noted that "in its present form, the CEI is complex enough that most non-specialists do not understand how to interpret the results" (Gleason, personal communication, March 2018).

F. Climate-Health Interactions: The Measurement of Vulnerability

An important, yet understudied, application of climate data is to human health. Since there are many factors that have the potential to impact human health, an important concept used to encompass all aspects is climate vulnerability. Climate vulnerability is defined by the Intergovernmental Panel on Climate Change (IPCC) as the degree to which a system is exposed to, and unable to cope with, adverse climate conditions (IPCC 2007: Impacts, Adaptation, and Vulnerability 2007). This means that for the assessment of climate vulnerability to be accurate, it should be based not only on how often certain climate conditions occur, but also on social factors.

Two different approaches have been used to describe the incorporation of social factors: vulnerability as a starting point and vulnerability as an end point. End point vulnerability measures remaining impacts of climate variables after adaptation strategies have been determined. The starting point approach views vulnerability as a pre-existing state that is based on socio-economic processes. Using the starting point approach helps to incorporate the important concept of adaptive capacity, or the ability of a community to "bounce back" after a particular climate event (O'Brien et al. 2004). It is also important to note that climate variability depends on more than just climate events; especially for health, socio-economic factors, such as income and access to healthcare, must also be incorporated to best assess populations most at risk to a particular disease (Lim et al. 2004). Therefore, the best way to characterize climate variability is to represent it as a function of exposure, sensitivity, and adaptive capacity (KC et al. 2015), as in the equation below:

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

Exposure to climate variables (temperature, precipitation, and extreme events) forms the basis for exposure. Sensitivity is a measure of the social and demographic make-up of the system, and adaptive capacity is the measure of a system to adjust to climate related factors (KC et al. 2015, taken from IPCC 2007).

KC et al. used this framework to form the basis of a Vulnerability Index, which calculates climate change vulnerability based on climate exposure as well as social vulnerability (sensitivity and adaptive capacity combined).

Climate Change Vulnerability = exposure + social vulnerability

Both factors had equal weighting in the computation of this index, corresponding to the additive model of calculating indices. Using additive models means that the final vulnerability value calculated will be equally dependent on all components, as opposed to being more dependent on or weighted towards one or two elements (Allison et al. 2009, Cutter et al. 2003, Reid et al. 2009). However, one disadvantage of this method is that the calculated index values may be offset by counterbalancing components (Godber and Wall 2014). One method that can be used to address this is to account for as many variables as possible when calculating climate change vulnerability. By accounting for as many factors as possible, this framework gives a comprehensive and more accurate measure of communities that will be more susceptible to changes in climate in the future.

The Social Vulnerability Index (SVI) from the Centers for Disease Control and Prevention (CDC) is one index that measures social vulnerability by taking this approach. The SVI determines the overall social vulnerability of each census tract in the United States to fifteen different variables (including income, age, race, and housing conditions).

A complete list of variables used in the calculation of the SVI can be found in the SVI 2016 Documentation document (https://svi.cdc.gov/Documents/Data/2016_SVI_Data/SVI2016Documentation.pdf). The variables used come from census data; in order to calculate the index more frequently than every ten years, estimated values from all variables included from the American Community Survey (ACS) are used to calculate the 2014 and 2016 index values.

To begin calculating the SVI, the values for each census tract are ranked against each other to create percentile rankings. Once rankings have been created for all variables, individual variables are grouped into four different themes: Socioeconomic status, Household composition, Minority Status, and Housing/Transportation (Flanagan et al. 2011). The percentiles for all variables included in each theme are added together, and then the resulting values are re-ranked to get the overall value for each theme. To determine the final SVI value, the percentile values for each theme are added together, and then the resulting percentile values are re-ranked. The final index is a percentile value that describes the vulnerability of each census tract in relation to other census tracts across the United States (Flanagan et al. 2011). In addition to calculating values per census tract, this index is also calculated for each county in the United States using the same process described above. Incorporating many variables when calculating this index ensures that no one variable dominates the calculation of the final index.

While calculating a vulnerability index is an important step towards communicating potential health risks with the public, the more important step is representing the results in a way that is easy for users to understand. A study conducted by Oxfam sought to produce maps that illustrated vulnerable communities by using a similar framework as that of KC et al. They represented their results in the form of bivariate analysis maps that visualized the convergence of climate hazards and social hazards (Figure 5 below, from oxfamamerica.org). From the scale on these images, it is easy to see where the strongest convergence of climate and social factors is. The square in the top right corner of the scale, represented by the color purple, indicates where climate and social vulnerability combined is the strongest. Areas in Louisiana that are this color are the areas that are currently the most vulnerable to climate variability (oxfamamerica.org, accessed 24 March 2018).

These two methods of KC et al. and Oxfam can be combined to determine the impact of climate factors on human health. After calculating a vulnerability index, the results can be mapped to determine communities that are the most vulnerable. After identifying the most vulnerable communities, the climate specific factors, like temperature or precipitation, that contribute the most to the highest amount of vulnerability can be examined. Using these climate indicators can help determine specific health hazards associated with each indicator, and this can help communities to take preventative action to minimize the risk to their health.

Social Vulnerability in Louisiana

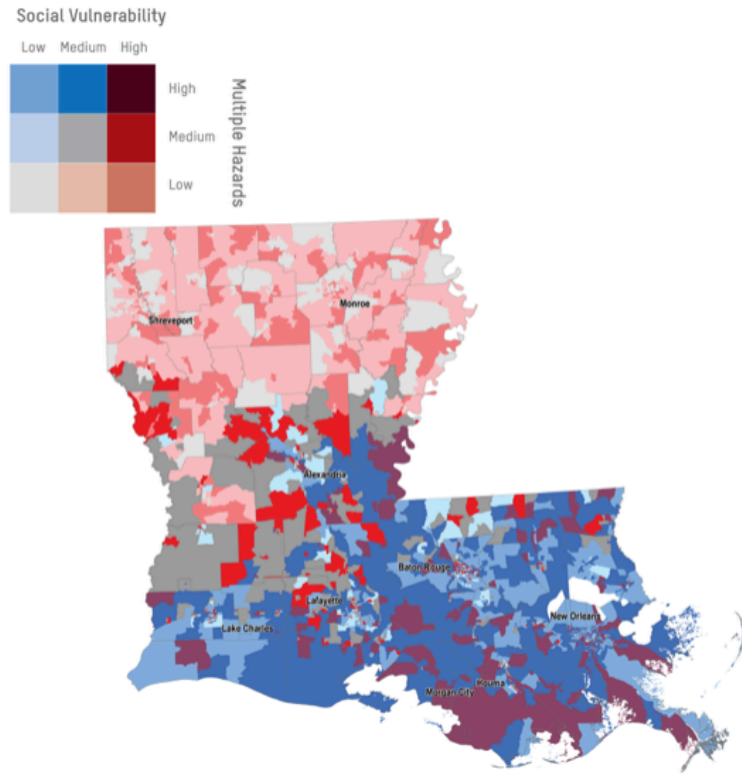

Figure 5: The resulting vulnerability map for the State of Louisiana, from oxfamamerica.org

CHAPTER 3

RESEARCH QUESTIONS

The original Climate Extremes Index (CEI) intended to quantify extreme weather patterns to educate policy makers and the general public about changes that they may see in the future. However, this index has been largely neglected by its target audience since its creation, and the primary users have become academics. Applying new statistics to analyze the climate data used to develop this index and implementing different displays of the information within the CEI will make it easier for users to interpret. The important link between climate and physical health warrants some attention as well. Investigating how elements of the CEI can be used to generate a new index that can be used to assess regional health risks will further help the public by identifying weather conditions that have the potential to negatively affect their health. This thesis seeks to answer the following questions:

1. Will employing different statistics to analyze temperature and precipitation data make the CEI more robust than in its current form?

    a. Will the use of Z-scores help to correctly identify extremes in temperature and precipitation?

    b. Will this approach correctly ignore non-extreme values?

2. How can the improved CEI be modified and applied on a regional level to assess health vulnerability to regional extremes?

CHAPTER 4

METHODOLOGY

A. Data and Sources

For the CEI update that forms the core of this thesis, the data for each component will come from datasets from NCEI:

Data for the calculation of Components 1 (maximum temperature) and 2 (minimum temperature) come from the nClimGrid dataset, which contains temperature data at a 5km resolution (Karl et al. 1990). This is the dataset used in the current calculation of the CEI.

For drought and moisture surplus (Component 3), the Palmer Drought Severity Index (PDSI) is used (Palmer 1965). This index incorporates temperature and precipitation data to assess drought or moisture conditions in various regions across the United States. The PDSI is calculated for nine climate regions (Northwest, West, Southwest, South, Central, West North Central, East North Central, Northeast, and Southeast), rather than for a specific station. These are the values used to calculate the current CEI. However, PDSI values are available at the climate division level, and these are the values used in this project to calculate the new CEI so that the resulting index is the more reflective of local patterns.

The data used to calculate Component 4 (maximum monthly precipitation) come from the Climate Divisional Database (nCLIMDIV), version 2, dataset (Vose et al. 2014). This dataset contains monthly values of various weather variables such as temperature and precipitation for each climate division in the United States.

The data used to determine vulnerability levels across the United States came from the 2016 county-level data used to calculate the Social Vulnerability Index (SVI) from the Centers for Disease Control and Prevention (CDC). Values for each variable, as well as each theme, are available for download on CDC's website (https://svi.cdc.gov/data-and-tools-download.html). These values were downloaded in the form of a CSV file, and the numerical values calculated for each theme (before percentiles rankings were calculated) were used in the calculation of the Extremes Vulnerability Index. These are the columns with the variable names SPL_THEME1, SPL_THEME2, SPL_THEME3, and SPL_THEME4. These variables are the sum of the values for all variables that are included in each theme.

B. Updating the CEI: An Overview

The code to calculate the new CEI has been adapted from current code that NCEI uses to calculate the existing CEI. Each component that is incorporated into the CEI is calculated in an individual script, and from there one final script calls the values calculated in the scripts for the individual components and calculates the final CEI. A similar process is used to re-calculate the CEI, but the scripts for each component have been edited to calculate the new CEI. The code and data for all components were made available for viewing and download on an NCEI ftp site. Changes to the code and calculation of the new CEI were done at NCEI as well as at the University of Georgia. The new CEI values are now stored on a personal external hard drive (and are available upon request) and were used for extremes vulnerability analysis.

Rather than basing each component on percentages of the country experiencing a certain extreme, Z-scores (see Chapter 2, Section D) were used to calculate the new CEI. To calculate each component, data discussed in Section A was used. The Z-score of the temperature or precipitation value at each station is determined by analyzing the value's location within the data distribution. Z-scores represent departures from a mean value, or standard deviations. The Z-score is used for each component to calculate the number of standard deviations of each component's monthly value from the 1981-2010 mean value. This mean value was selected to be consistent with current literature that also use the 1981-2010 normals as a base period for comparison (https://actuaries.asn.au/microsites/climate-index). Following Chebyshev's Rule, Z-scores with an absolute value greater than 2 are considered to be values that are "unusual", and Z-scores with absolute values greater than 3 are considered to be indicative of extreme values (see Table 1). The same Z-score threshold is considered to be extreme for all components; this leads to consistent analysis across all components and allows for the calculation of the final CEI to be on a numerical scale.

Table 1: Interpretation of Z-score Values

Absolute Value	Interpretation
0-2	Normal
2.01-3	Unusual
>3	Extreme

Once the Z-scores for each component were calculated, the number of extreme components per climate division were added together to calculate the final CEI value. Following this process leads to the calculation of the new CEI for every month from 1900-2017 for each climate division. Calculating the new CEI for each climate division in the continental United States (CONUS) obtains a more local CEI for optimal use by the public and other stakeholders. This statistic will identify extremes more accurately than the previous approach of Karl et al., because it will only identify values outside 3 standard deviations as extreme. Because the calculation of this statistic is based on the mathematical relationship between the mean and the rest of the data, there is no need for a secondary accuracy assessment of the components that make up the CEI.

C. Re-calculating the CEI

The original CEI incorporates five different components into its final calculation. However, it was mutually agreed upon with NCEI employees to omit the fifth component (extremes in the number of days with recordable precipitation) in the calculation of the new CEI for this thesis project. So, the new CEI includes only the first four components used in the original CEI.

Each of the four components of the CEI was recalculated so that extreme values were determined on a climate division basis (see Figure 6). The analysis in each script determines the Z-score of the value for each component in each climate division. Z-scores were then used to determine which climate divisions were extreme for the analysis period. Now, rather than measuring the percent of the US experiencing extremes as the existing version of the CEI does, this new CEI calculates the extremity of each climate division.

Figure 6: The climate divisions of the United States.

1. Components 1 and 2

Because Components 1 and 2 are calculated in the same way, the only difference being the datasets used in calculations and names of data files, the explanation of the recalculation of both is presented here. This explanation below refers to Component 1, but the same steps were also done to calculate values for Component 2. These components both represent temperature extremes. Component 1 is the extremity of maximum temperatures, and Component 2 is the extremity of minimum temperatures.

Again, data used to calculate these components came from the nClimGrid dataset, so the temperature data has a 5-km resolution. In order to calculate the CEI for a climate division, the first step was calculating a temperature value per climate division. This was done by calculating an average temperature value of all the data points in each climate division. After finding the average monthly value to represent each climate division, Z-scores (based on the 1981-2010 monthly average and standard deviation) were calculated for each climate divisional value to assess the extremity of each monthly climate division value.

There were several changes made to the original CEI code from NCEI in order to calculate the Z-score of each value based on climate division. The biggest adjustments were in the subroutine called "regread"; this subroutine reads in the polygon files of each climate division. Each of the latitude-longitude points in each file represents the boundary points of each climate division in the continuous United States. The subroutine in the original code reads in points for nine climate regions, rather than divisions. When reading in points for divisions, more changes needed to be made in addition to changing the dimension of a few arrays. There are a number of climate divisions (e.g., those along coastlines) that are made up of a number of different parts, or segments, meaning that the original polygon point files contain greater-than-sign symbols (>) dividing up the original divisions into segments. In order to account for each segment while also being able to calculate a value per climate division, a new one-dimensional data array called "iSegments" was created. This array contains the number of segments within each of the 344 climate divisions in the United States. Each of the original polygon files was also divided into one file per segment, if the climate division had multiple segments. Those that only had one segment were left alone. The integer array "iSeg" is then used later in the subroutine as the script loops through each state and division to determine which polygon file, or files, to read in. After looping through each state and division, the subroutine loops through segments as well to determine which polygon files to read in. If the value of iSegments for that particular climate division equals 1, the subroutine reads in the original polygon file. If, however, the value of iSegments for that climate division is greater than 1, then the subroutine loops through each of the segments and reads in the polygon file corresponding to that segment. In this way, the program is able to read in all polygon points for each climate division while preserving the boundaries between segments (see Appendix B).

In the subroutine and the main body of the code, a counter ("iKounter") is used to tell the program which value within "iSegments" represents the number of segments in the first climate division of each state. The program then uses this value when reading in climate division polygon files and also when assigning individual grid points to each climate division.

After each grid point is assigned to a climate division, the average Tmax value for each climate division is calculated by averaging together the values for the grid points that were assigned to that climate division. These average values are assigned to a new array, as well as written to a new file. The program then calculates the 1981-2010 monthly mean for each climate division using the newly calculated average values per division. The program also calculates the 1981-2010 monthly standard deviation values based on the average Tmax value for each month and the monthly mean. The final step, and final adjustment to the code for the temperature components, is to calculate the Z-score of each monthly average climate division Tmax value. The equation used to calculate each Z-score is the equation discussed previously. The code uses the Tmax value for each climate division and the 1981-2010 30-year mean and standard deviation values (per climate division) to calculate each Z-score. All Z-scores are stored in a new array and new text file. All Z-scores with an absolute value greater than or equal to 3 mean that the individual Tmax value is extreme for that month (see the previous description of Z-scores in Chapter 2).

Again, the process used to calculate Z-scores for Tmin (Component 2) and adjustments made to the original code are the same as Component 1. The only difference, from a computational standpoint, is the name of the data file that contains the monthly temperature values. The adjustments to the code were also the same. If one wants to understand how Z-scores for Tmin values were calculated, substitute "Tmin" wherever "Tmax" is printed in the previous section.

2. Component 3

Z-score values for Component 3, the moisture availability component, were based on monthly climate division data of the Palmer Drought Severity Index (PDSI). The current CEI uses monthly PDSI values for each climate division in the United States to determine which divisions had a severe moisture surplus (top 10%) and severe drought (bottom 10%). To determine whether a certain division was extreme, the original program loops through the PDSI values of the analysis period and then saves the values representing the 90th and 10th percentiles for the period of analysis. It then compares each monthly PDSI value to these values. If a PDSI value is greater than the 90th percentile value, it is determined to be a high end extreme (moisture surplus), and if a value is less than the 10th percentile value it is determined to be a low end extreme (drought). Once the program determines divisions that are extreme, it adds together the areas of all of these divisions to determine the area of the US that is experiencing extremes in moisture availability.

To calculate the new Component 3 of the CEI, the new script determines the Z-score of each monthly PDSI value. Because the Z-score calculation is based on means and standard deviations, those values per month were calculated first. The time period 1981-2010 was used as the period of comparison because this is the most current 30-year analysis period that is available. Since the original PDSI values were monthly, the script calculated the average monthly mean for 1981-2010; this process was repeated for standard deviation. Once these reference values were calculated, the script calculated the Z-score for every PDSI value within the analysis period of 1900-2017 using the respective 1981-2010 monthly mean and standard deviation values.

The result when using Z-scores to determine extreme values is values on a numerical scale. Because of Chebyshev's Theorem, described in any introductory Statistics textbook, 75% of the data will lie within 2 standard deviations of the mean, and about 89% need to lie within 3 (Larson and Farber 2006). Therefore, PDSI values with a Z-score value greater than 3 are considered to be high-end extremes and have a high amount of moisture available (severe moisture surplus), and PDSI values with a Z-score less than -3 are considered to be low end extremes and have very little available moisture (drought). Values with a Z-score of absolute value greater than or equal to 2 will be considered to be "unusual."

3. Component 4

The calculation of the fourth component, the extremity of monthly precipitation totals, differed the most from the methodology used by NCEI to calculate this component. This component of the CEI is currently based on daily precipitation data. To determine extreme values, the script determines the proportion of daily to total precipitation (per analysis period) for each 1° x 1° grid that would be the 90th percentile value. Once that threshold is in place, it determines which proportion is that threshold across all grids (to get a national threshold for extremity). For a given analysis period, each extreme grid-based precipitation proportion is compared to the national extreme proportion to determine the percent of the country with the greatest proportion of extreme daily precipitation.

Since the desired outcome was one monthly value per climate division, a different methodology and dataset were used for this project. Data for the calculation of this component came from the Climate Divisional Database, from version 2 of the nCLIMDIV dataset (Vose et al. 2014). This dataset calculates one monthly value of various climate variables per climate division, total monthly precipitation being one of them. Text files containing the data for each variable calculated are available for download from NCEI.

The dataset that is the source for this data is the Global Historical Climatology Network-Daily (GHCN-Daily) dataset. This dataset contains values from several major observing networks of station data, the primary networks being the Cooperative Observer (COOP) program and the Automated Surface Observing System (ASOS). To supplement data from regions where these two networks had sparse data, nCLIMDIV-version 2 also includes data from the National Interagency Fire Center (NIFC) Remote Automatic Weather Station (RAWS) network, the USDA Snow Telemetry (SNOTEL) network, the Environment Canada (EC) network, and part of Mexico's Servicio Meteorologico Nacional (SMN). To derive one value per climate division, a gridding method was used.

5 km gridpoint estimates were interpolated from station data, and a 5 km spacing was used to ensure that all divisions had fairly equal representation in the calculation of final values. Once one data point was assigned to each 5 km grid, final climate division values were determined by calculating the area-weighted average of all 5 km grids within each climate division. For full documentation of how this dataset is derived, see Vose et al. 2014.

Since the data file for monthly precipitation totals was in the same format as that of the file containing PDSI values, the steps used to calculate Z-scores of this component were the same as those used to calculate Z-scores of PDSI values. The 1981-2010 monthly means and standard deviation values were calculated for each climate division. Then the Z-score of the monthly precipitation total for each climate division was calculated. As with the three previous components, any Z-score with an absolute value greater than 3 is considered to be an extreme value for that climate division.

D. Calculating Social Vulnerability

After downloading county-level census data from the CDC, social vulnerability values were calculated. The original SVI is calculated using percentile rankings. However, the percentiles for the final SVI cannot be combined with the Z-score values of the CEI. Using Z-scores produces a numerical scale, so the CEI values cannot be used with the additive framework to calculate climate extremes vulnerability with the SVI in its current form.

To address this issue, SVI theme values were recalculated using Z-scores. The Z-score of the value of each theme for each county was calculated based on the national mean and standard deviation for the 2016 data. Calculations were performed using builtin functions in Excel. Once county-level Z-score values for each theme were calculated, the values were rescaled based on Z-score value so that they were on a numerical scale of only whole numbers, using a technique similar to Reid et al. (2009). Values used when re-scaling are in Table 2. All values less than 0 were rescaled to 0 because in terms of vulnerability, negative numbers indicate areas that are more resilient; since the goal of this project was to focus on identifying counties that are more vulnerable, re-scaling negative numbers to 0 ensured that only areas that are less resilient are included in the final SVI calculation. Aside from the negative numbers, all values were re-scaled based on their Z-score value.

Table 2: Re-scaled values for the re-calculated SVI

Re-scaled Value	Corresponding Z-score values
0	< 0
1	0-1
2	1.01-2
3	2.01-3

Once values were re-scaled, this CSV was imported into ArcMap from ArcGIS to perform analysis and create maps. The CSV file containing Z-score values was joined to the shapefile from the CDC so that re-scaled values could be mapped. This join needed to happen so that the tabular data (SVI theme Z-scores) was associated with spatial features. The "Feature to Point" tool was then used to identify the centroids of each county. This was done to ensure accuracy when performing analysis in GIS. Using the centroids of each county ensures that only counties completely contained within each climate division are considered to be located within a climate division. Otherwise, ArcMap will assume that any county that touches the boundary of a climate division is included in that division, and this could produce inaccurate results.

The centroids layer was then spatially joined with the climate divisions layer to summarize re-scaled Z-score values by maximum value per climate division. The maximum value statistic was used to assume a "worst case" scenario when calculating vulnerability to determine the impact of the highest amount of vulnerability and identify all climate divisions that had counties that are highly vulnerable. Using the maximum value per climate division ensures that these climate divisions can be identified; calculating average values per climate division could cause high values to be masked out, thus leading one to think that there are not vulnerable counties in that climate division. This resulting layer was then saved as a new file. The resulting calculation was the maximum Z-score value recorded for a county within each climate division, leading to one SVI theme value per climate division. Maps were then created in GIS to visualize social vulnerability values per theme to see how all climate division values compare to one another. To calculate the final SVI, the Z-scores for each theme were added together, resulting in an index ranging from 0-10. Higher values indicate a climate division that is more vulnerable from the perspective of social variables.

E. Calculating an Extremes Vulnerability Index

To calculate overall vulnerability for each climate division, values for both the December 2015 CEI and the 2016 SVI were broken into three groups: low, medium, and high values for each index. Each group was then given a numerical value. Low was given a value of 1, medium was given a value of 2, and high was given a value of 3. The values of each index for each climate division were then added together, following the additive framework, to produce a numerical scale of overall vulnerability, ranging from 2-6. A value of 6 indicates a division that is the most vulnerable.

To then calculate an index of extremes vulnerability, values from the CEI and SVI were combined to produce an overall value that indicates how vulnerable each climate division is and which index is contributing the most to overall vulnerability. To best communicate this, a bivariate scale was used in ArcMap when creating the map. The values of each index were broken down into three groups that correspond to low, medium, and high vulnerability. Values from each index included in each group are contained in Table 3. The different values included in each group of SVI vulnerability are based on the re-scored Z-score values of each theme. The maximum Z-score value recorded per theme was either 2 or 3. This means that the highest SVI value a climate division could be is 10. This number gives an indication of the number of themes that had the highest possible Z-score value (per climate division). An SVI value between 0-3 indicates that a maximum of 1 theme recorded the maximum possible Z-score value. An SVI value between 4-6 indicates a maximum of 2 themes per climate division recording the highest possible Z-score, and an SVI value between 7-10 indicates that 3 or 4 themes recorded the maximum Z-score value.

Table 3: Values included in each group of vulnerability

	CEI	SVI
Low Vulnerability	0	0-3
Medium Vulnerability	1-2	4-6
High Vulnerability	3-4	7-10

After creating different groups of vulnerability for each index, the values of both indices were compared to each other to create a bivariate scale ranging from 1-9 to determine overall vulnerability. This scale indicates the level to which each climate division is vulnerable to variables included in each index, and also which index is impacting the overall value the most. An explanation of each value is contained in Table 4. A value of 9 indicates a climate division that is vulnerable to both physical exposure and social variables. A value of 1 indicates a climate division that is not highly vulnerable, and values between 2-7 indicate climate divisions that are moderately vulnerable to at least one factor. Using a scale of this type can help to identify the factors that each climate division is vulnerable to so that effective mitigation and adaptation strategies and policies can be implemented to decrease overall vulnerability.

Table 4: Extremes Vulnerability Index interpretation

EVI Value	Interpretation
1	Low CEI value, Low SVI value
2	Medium CEI value, Low SVI value
3	High CEI value, Low SVI value
4	Low CEI value, Medium SVI value
5	Medium CEI value, Medium SVI value
6	High CEI value, Medium SVI value
7	Low CEI value, High SVI value
8	Medium CEI value, High SVI value
9	High CEI value, High SVI value

CHAPTER 5

RESULTS AND DISCUSSION

After recalculating the CEI and SVI and calculating the Climate Extremes Vulnerability Index, several results were found. All main findings are described below.

A. 1981-2010 CEI Component Statistical Distributions

After adjusting the original CEI code to calculate the Z-scores of each of the four components, additional analysis in R was performed to determine the percentage of each component's distribution that was within 2 and 3 standard deviations of their respective means. The mean values during the 30-year normals period (1981-2010) for the month of December were used for this analysis, since this was the period that all raw data values were being compared to during the calculation of Z-scores.

1. Component 1

The distribution of the values for Component 1 (maximum temperatures) for all Decembers from 1981-2010 is shown in Figure 7. From looking at this figure, it appears that the majority of the data are contained within 2 standard deviations of the mean. To confirm this, the percent of the data that are contained within 2 and 3 standard deviations was calculated in R (see Table 5). This analysis determined that 95.41% of the data lie within 2 standard deviations of the mean, and 99.29% of the data lie within 3 standard deviations. Thus, the December data for Component 1 (maximum temperatures) for the period 1981-2010 conform to Chebyshev's rule.

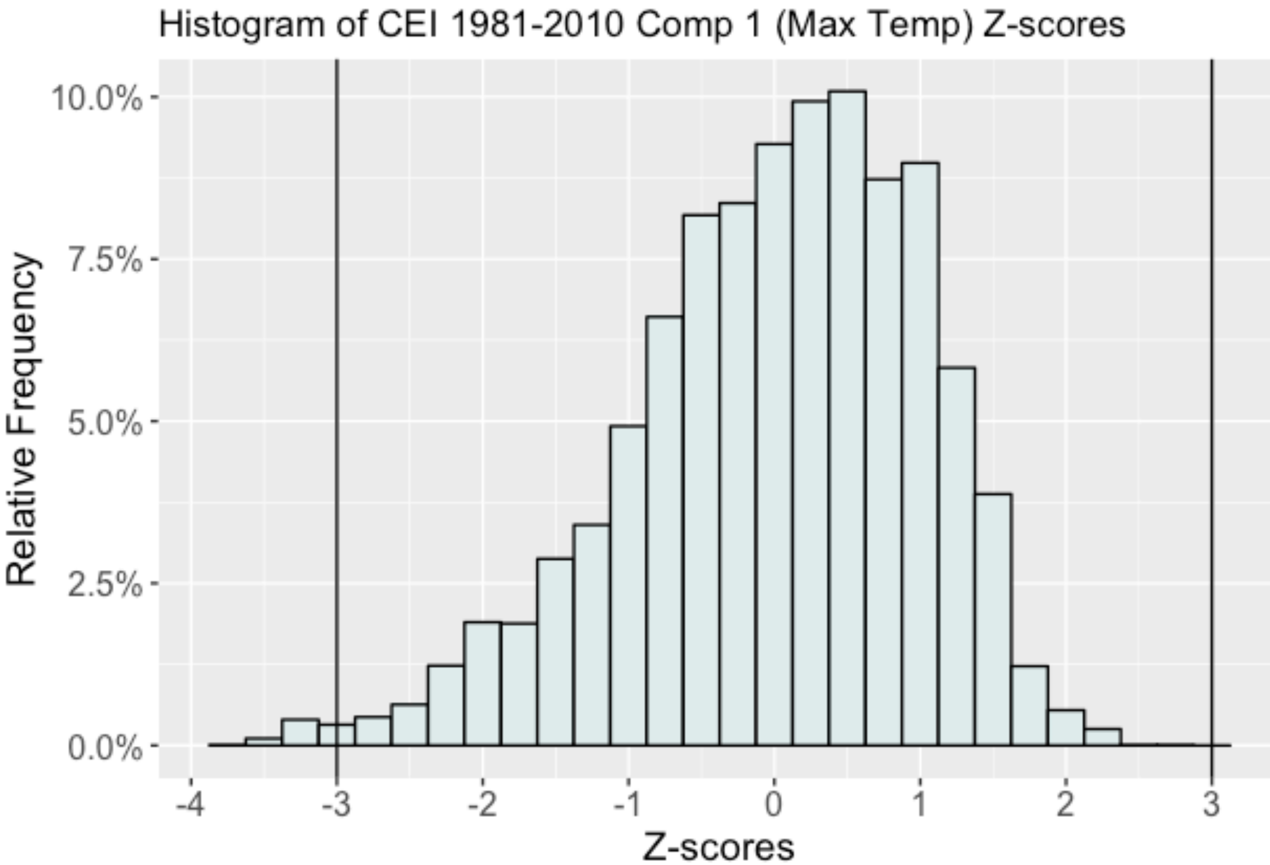

Figure 7: Distribution of 1981-2010 CEI Component 1 Z-scores

Table 5: Percent of each 1981-2010 CEI component distribution lying within a number of standard deviations of the 1981-2010 mean

CEI Component	Percent of Z-scores lying within 2 standard deviations	Percent of Z-scores lying within 3 standard deviations
1	95.41%	99.29%
2	95.03%	99.25%
3	96.77%	99.98%
4	95.40%	98.89%

2. Component 2

The distribution of values for Component 2 (minimum temperatures) during December from the period 1981-2010 is shown in Figure 8. As with Component 1, it appears that the majority of raw data values are within 2 standard deviations of the mean and that the data from this period conforms to Chebyshev's Rule. After completing additional analysis in R, this inference was confirmed. 95.03% of the data are within 2 standard deviations of the mean, and 99.25% are within 3 standard deviations (Table 5). Thus, the data from this component also conform to Chebyshev's Rule.

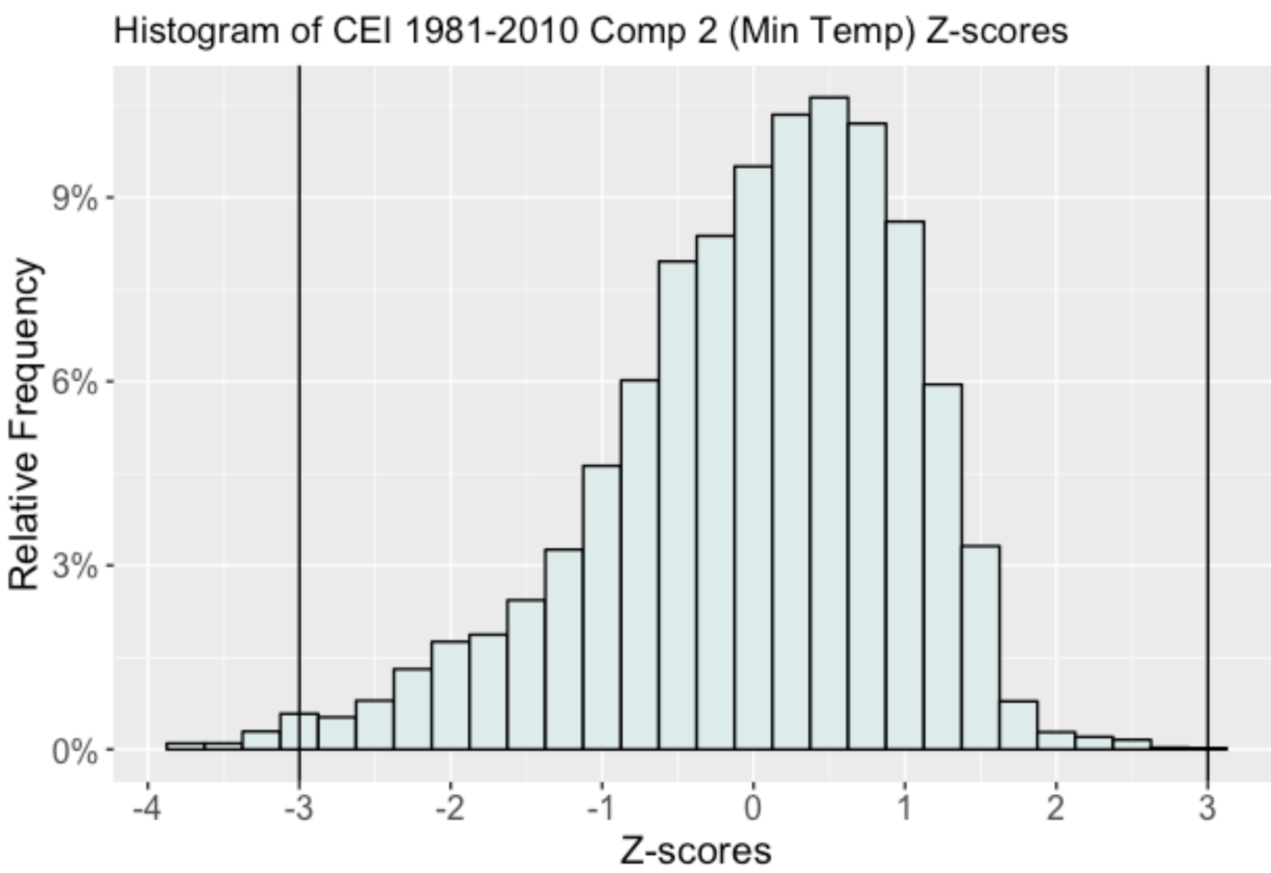

Figure 8: Distribution of 1981-2010 CEI Component 2 Z-scores

3. Component 3

The distribution of all Z-score values during the month of December between 1981-2010 is shown in Figure 9. This distribution bears a very strong resemblance to a normal distribution, and it appears that the majority of the values are within 2 standard deviations of the mean. After analysis in R, it was found that 96.77% of this data lie within 2 standard deviations of the mean, and 99.98% of the dataset lie within 3 standard deviations. Through this analysis, this subset of the Z-score values conforms to Chebyshev's Theorem, as more than 75% of the data lie within 2 standard deviations, and more than 89% lie within 3 standard deviations (Table 5).

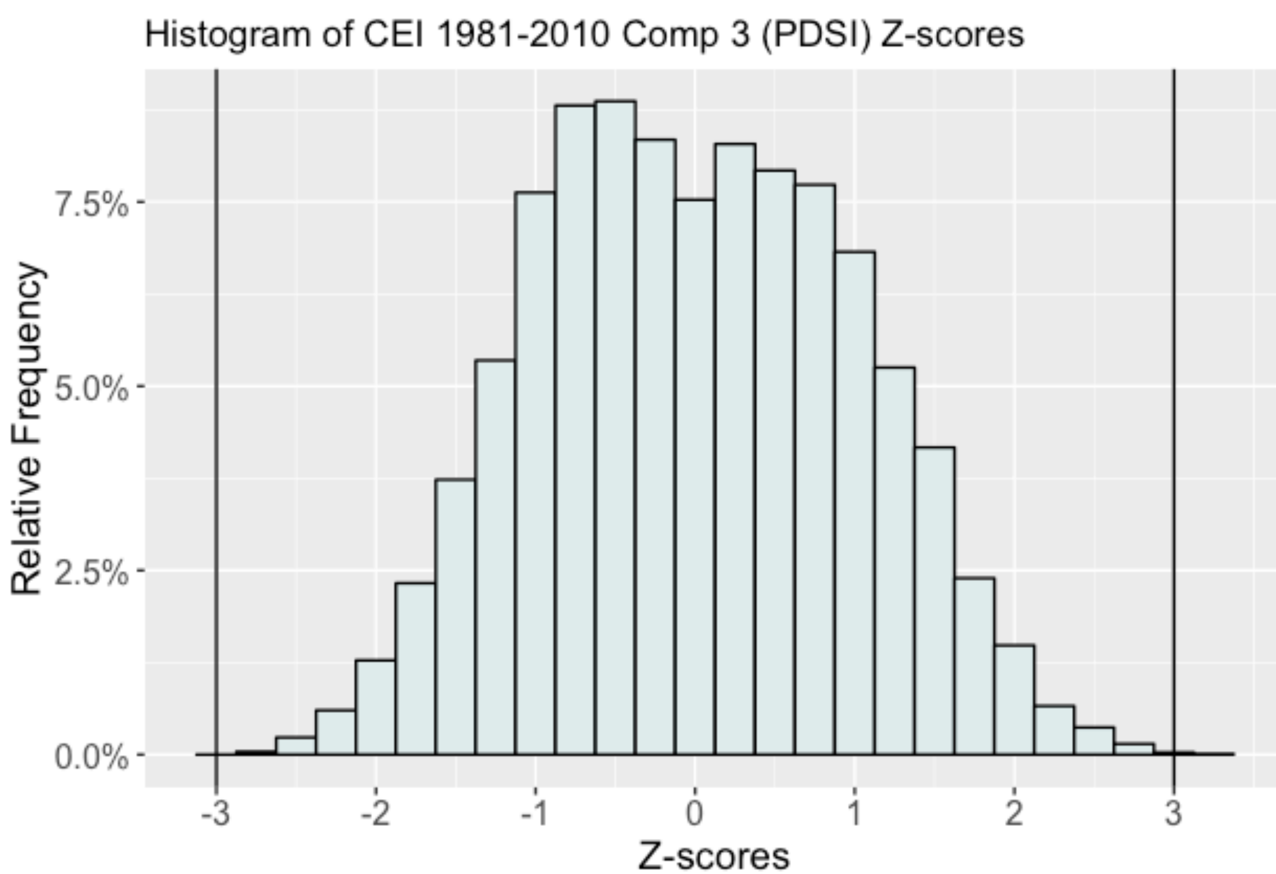

Figure 9: Distribution of 1981-2010 CEI Component 3 Z-scores

4. Component 4

The distribution of all Z-scores for Component 4 (monthly precipitation totals) for Decembers between 1981-2010 is displayed in Figure 10. This data distribution differs slightly from that of the other three components, in that it appears that the majority of data points had a Z-score less than 0. However, it still appears that the majority of values lie within 2 standard deviations of the mean. Analysis in R shows that about 95.40% of this data lie within 2 standard deviations, and 98.89% of the data lie within three standard deviations (see Table 5). This analysis shows that this subset of the data for this component conform to Chebyshev's Theorem.

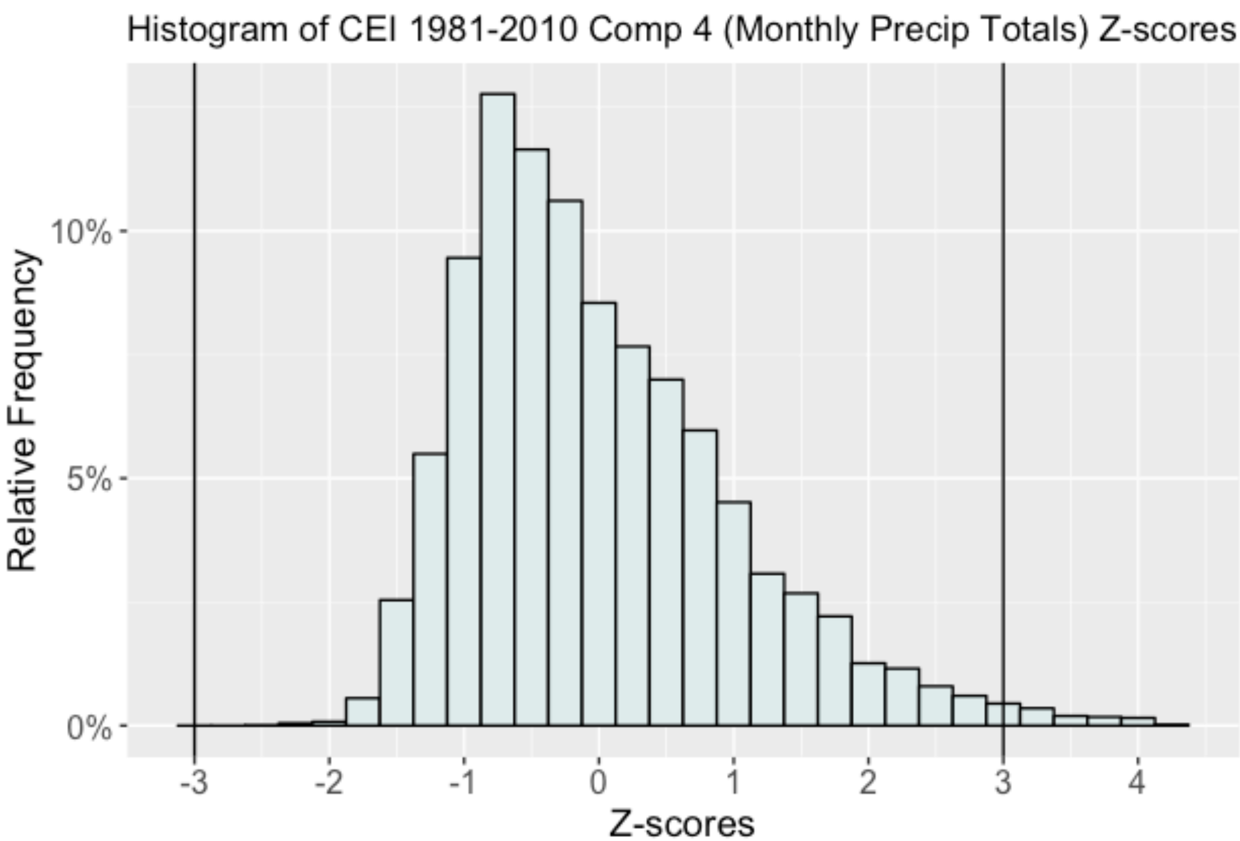

Figure 10: Distribution of 1981-2010 CEI Component 4 Z-scores

B. A Case Study comparing the current and new CEI for December 2015

Because the goal of this project was to focus on the identification of extreme values and communication of information contained within the CEI, one month was selected to use for comparison to the current CEI. The month of December 2015 was selected as the month for analysis because a very intense El Niño event occurred during this winter, and many states recorded both temperature and precipitation values that were described as "well above normal", making it both the warmest and wettest month on record for many reporting stations (https://www.ncdc.noaa.gov/sotc/national/201512). Normal wintertime responses to El Niño include above-average precipitation in the eastern United States, and below-average temperatures in the southeastern United States (Ropelewski and Halpert 1986).

Because the observed pattern in the eastern United States during December 2015 consisted of warmer temperatures *and* above-normal precipitation amounts, this month is highly abnormal for an El Niño event. The hypothesis when selecting this month was that, upon mapping out the new component values, there would be extreme values to map. This would help to test whether extreme values would be correctly identified and non-extreme values ignored. To display the new CEI, maps of the values for each of the four components, as well as the final CEI calculation, were created using ArcMap.

1. CEI Component 1

To determine the normality of this month's data, calculations were done in R to determine the percent of values of this component that fell within 2 and 3 standard deviations of the 1891-2010 mean. To reiterate from Chapter 2, Section D, Chebyshev's

Theorem states that for a distribution to be considered normal at least 75% of the data must lie within 2 standard deviations of the mean of a dataset, and at least 89% must lie within 3. For Component 1 (Maximum Temperature) values, calculations in R found that 54.65% of these values lie within 2 standard deviations, confirming that this month's data distribution was not normal (see Table 6). Even though 99.71% of the data lie within 3 standard deviations of the mean, less than 75% are within 2 standard deviations. Since such a high percentage of the data lie outside 2 standard deviations of the mean, this illustrates that the data for this month is not normal, as compared to the 1981-2010 mean and standard deviation.

Table 6: Percent of each December 2015 CEI component distribution lying within a number of standard deviations of the 1981-2010 mean

CEI Component	Percent of Z-scores lying within 2 standard deviations	Percent of Z-scores lying within 3 standard deviations
1	54.65%	99.71%
2	49.42%	88.37%
3	94.19%	99.71%
4	72.67%	86.05%

The histogram displaying the distribution of this data also illustrates that this month's distribution was not normal (Figure 11). This distribution is abnormal and uncommon in shape, in that the majority of the data are greater than the 1981-2010 mean, meaning that the data distribution for this month is very positively skewed (see Figure 11). The skewness of this distribution is the result of the departure of current temperatures from the 1981-2010 normals. Additionally, a large portion of the Z-scores for this month lie outside of 2 standard deviations from the 1981-2010 mean. Because the monthly values are compared to the 1981-2010 mean and standard deviation, the comparison between the two distributions reveals that the month of December 2015 was not normal. This confirms the abnormality of the temperature observations that were above normal for this month.

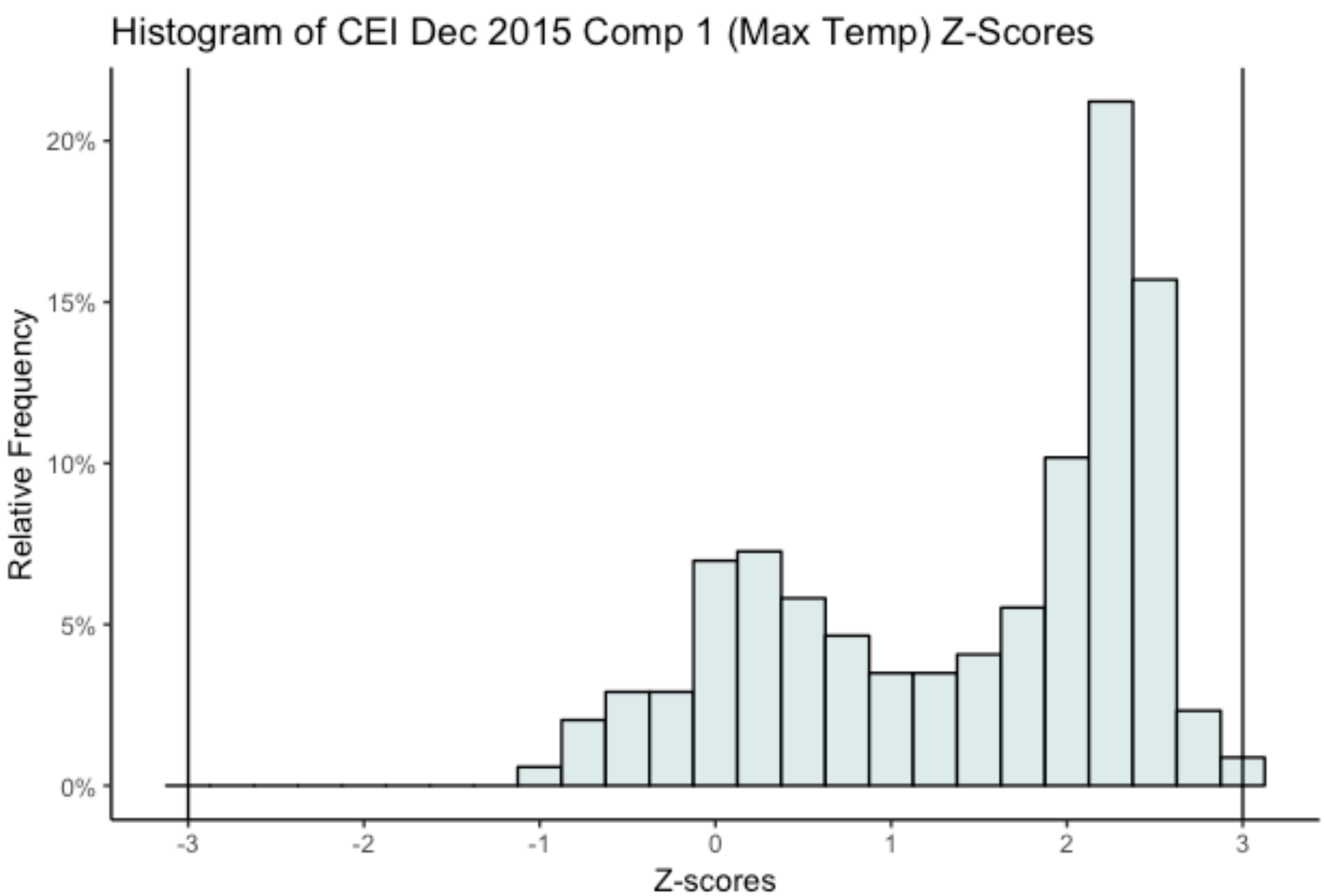

Figure 11: Histogram of the values for the CEI, Component 1 for December 2015

After updating the original code to calculate the CEI, maps were created to identify the locations that experienced unusual and extreme values in maximum temperatures for this month, as well as compare results to the original CEI. Upon comparing these two maps, several discrepancies between the new and current CEI emerged. First, the new calculation of the CEI does identify different values as extreme. The map from NCEI conveys the message that extremes in maximum temperature occurred in about half of the eastern United States, including Minnesota, Iowa, Kansas, Oklahoma, Texas, and all states to their east (Figure 12). However, the new CEI, mapped in Figure 13, shows that extreme values occurred in a much smaller portion of the country. Using the new scale of extremity, only one climate division experienced extremes in maximum temperature, meaning that the Z-score had an absolute value greater than 3. This was New York's climate division 4, which consists of Long Island, with a Z-score of 3.04. This means that during this month, this climate division's average maximum temperature value of 55.95 degrees Fahrenheit (°F), or 13.30 degrees Celsius (°C), was significantly above the mean (due to a positive Z-score). This maximum temperature, which was very above normal, means that this area experienced maximum temperatures that were much higher than values that they normally experience.

Several climate divisions during this month experienced temperatures that were unusual, meaning that the absolute value of their Z-score values was between 2 and 3 (see Table 1). These are the values shaded in dark pink; this area includes portions of Wisconsin, Illinois, Indiana, Missouri, Arkansas, and Louisiana, as well as all states to the east in the Midwest, Southeast, and Northeast. All of these values had positive Z-score values, meaning that the temperatures that these areas experienced were above the value that would be expected. So, these areas experienced daily high temperatures that were unusually higher than what they should normally experience in the month of December. The only state within the eastern United States that was not unusual or extreme during this month was Maine.

Visually, comparison of these two maps demonstrates the effectiveness of Z-scores at identifying values that, in relation to a mean value per climate division, actually are extreme and ignoring those that are not. Because each Z-score calculation is based on the mean for that climate division, Z-scores will correctly identify raw data values that are truly extreme. Top and tail analysis, as currently used by NCEI to calculate the CEI, identifies a much larger portion of the United States that is extreme; in reality the majority of the values highlighted as extreme really are not extreme.

Additionally, the map from NCEI also raises many questions. Values that NCEI considers to be extreme include the upper and lower 10^{th} percentiles of the dataset. This would imply that about 10% of the United States should be shaded, but close to half of the country is shaded in the map from NCEI (see Figure 12).

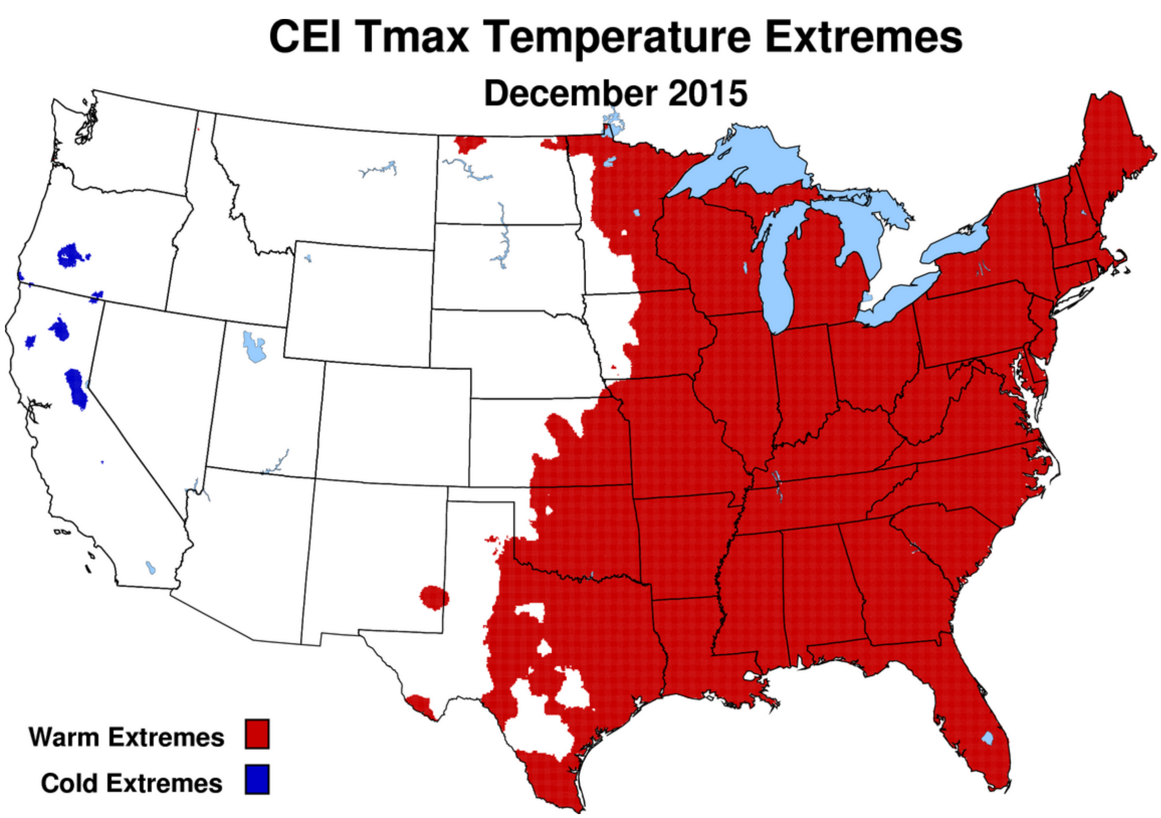

Figure 12: December 2015 Component 1 map (courtesy of Karin Gleason, NCEI)

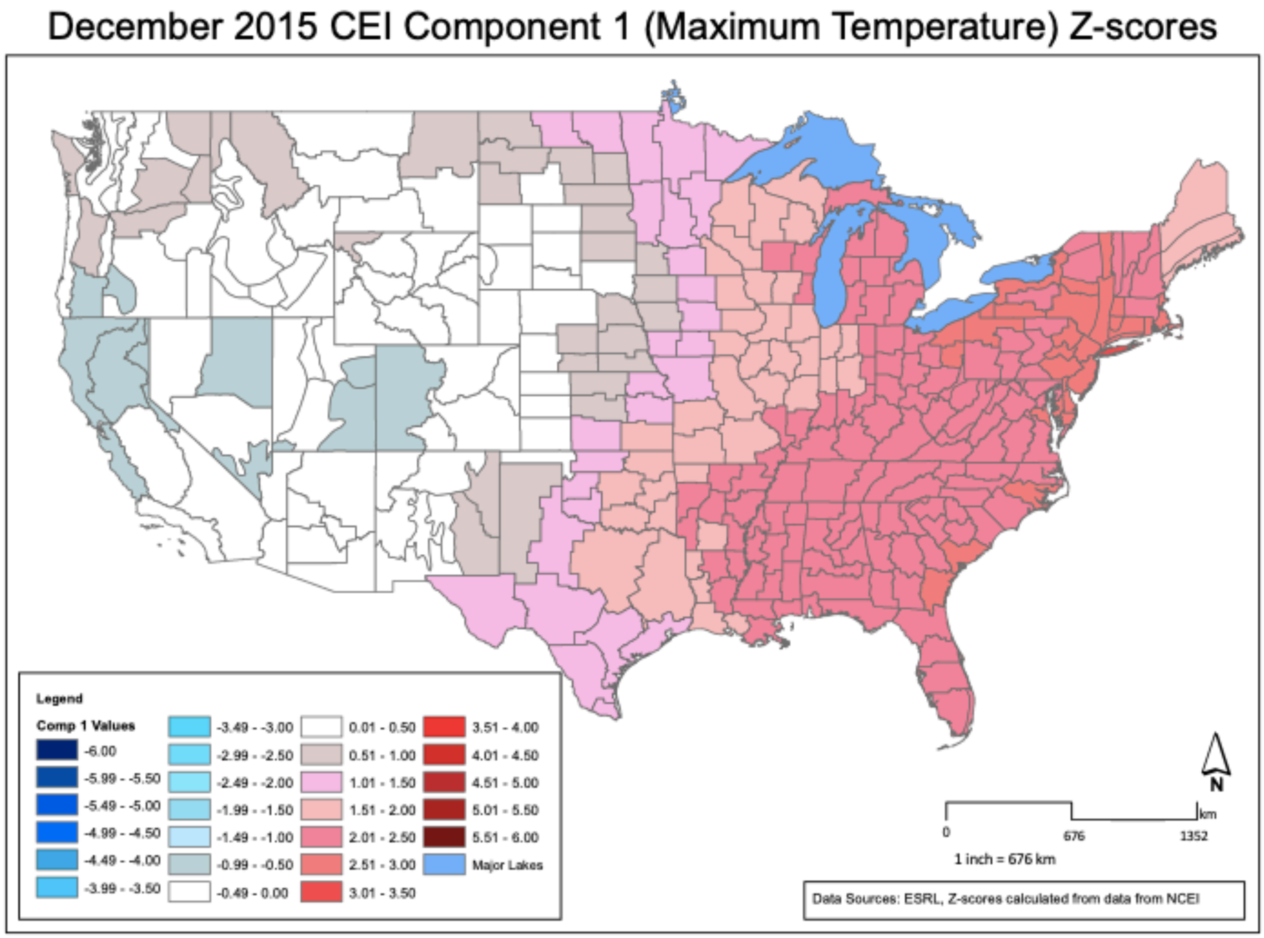

Figure 13: New December 2015 CEI Component 1 values for all 344 divisions of the United States

2. Component 2

As with the first component, additional analysis was done in R with the Component 2 (minimum temperature) Z-scores from December 2015. This was done to determine whether or not the distribution of the data for this month could be considered normal. This analysis found that only 49.41% of the data from this month lie within 2 standard deviations of the mean, and that only 88.37% lie within 3 standard deviations (see Table 6). This means that slightly more than 11% of the minimum temperature values are extreme. The distribution of the data for this component is even more abnormal than that of the first component, as just over half of the data are values that are either unusual or extreme. Again, the data for this month are being compared to 1981-2010 mean and standard deviation values, so these are the values and data distribution considered to be normal. Because such a large percentage of the data are outside of 2 standard deviations of the mean of the 30-year normals, the distribution of the values for Component 2 during December 2015 is not normal.

The histogram of the data for this month further illustrates how abnormal the distribution of minimum temperature values for this month were (Figure 14). This histogram shows that this data distribution, like that of Component 1, is heavily skewed towards the right end of the distribution. This means that the majority of the data for this component are greater than 0. Because this distribution is skewed, as compared to the distribution of the 1981-2010 December values, it is not normal. Additionally, this histogram also illustrates where the unusual and extreme values fall within this distribution. Looking at this histogram reveals that all of the unusual (absolute value greater than 2) and extreme (absolute value greater than 3) values for this month were positive, indicating that all unusual and extreme values for this component during December 2015 were above normal. The solid black line indicates where the Z-score value of 3 is located, and anything to the right of this line is considered to be extreme for this component. Because a higher percent of the data for Component 2, as compared to Component 1, lie outside of 3 standard deviations of the mean, this component recorded more extreme values during December 2015. This histogram confirms that the data distribution for minimum temperatures for this month is skewed, and that the majority of values are abnormally above normal.

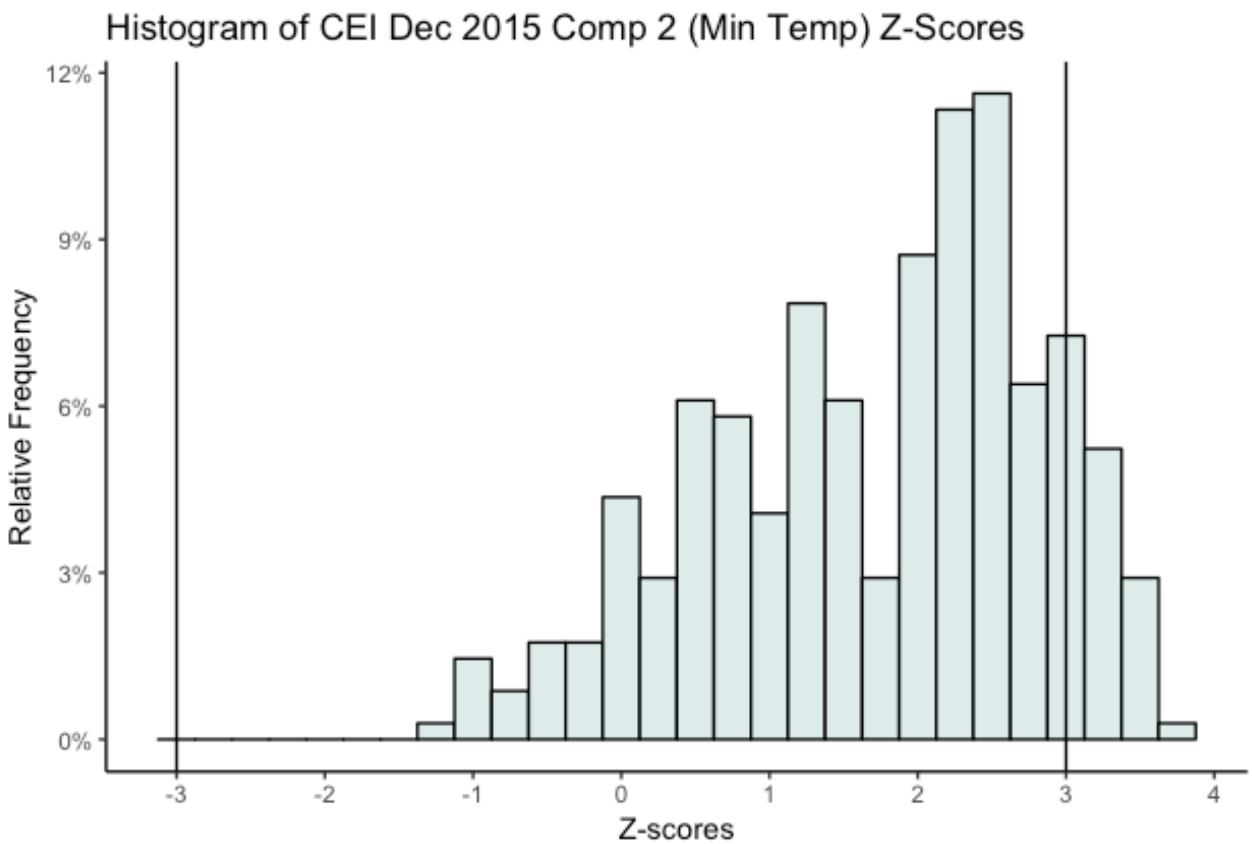

Figure 14: Histogram of the values for the CEI, Component 2 for December 2015

The map produced for this component helps to reveal where the extremes in minimum temperatures occurred during this month, and, like that of the map for extremes in maximum temperatures, tells a different story than the map produced by NCEI. The map from NCEI (Figure 15) conveys the message that extremes in minimum temperature for this month occurred in over half of the United States, extending from the Dakotas, Nebraska, Kansas, Oklahoma, and eastern Texas eastward to include the rest of the eastern United States. There are also portions of Colorado and New Mexico included. However, the recalculated CEI (Figure 16) conveys a very different message. Using the new scale of values, the values that are considered to be unusual (Z-scores with absolute values greater than 2) begin in very eastern Minnesota, Iowa, Arkansas, and Louisiana and extend eastward to include the rest of the eastern United States. Extreme values (Z-scores with absolute value greater than 3), colored in shades of red, are only located in states in the very eastern United States, such as the Carolinas, eastern Tennessee, Georgia, Virginia, Maryland, Delaware, New Jersey, Eastern Pennsylvania, Connecticut, and Rhode Island.

For example, the entire state of North Carolina was either unusual or extreme for this month. The one division in western North Carolina that was unusual had a Z-score of 2.9, corresponding to a temperature of 39.74°F (4.3°C). This means that, compared to the monthly mean of this climate division, this average minimum temperature was unusually higher than normal (because the Z-score is positive). The remainder of North Carolina experienced minimum temperatures that were extremely above normal, due to positive Z-scores greater than 3. These Z-scores ranged from 3.07-3.58, corresponding to temperatures ranging from 39.2°F-47.46°F (4.00°C-8.59°C). These average monthly minimum temperatures represent values that are extreme for these climate divisions.

Again, from a visual standpoint, the comparison of these two maps further demonstrates the effectiveness of Z-scores at identifying values that actually are extreme and ignoring those that are not, in relation to the mean value per climate division. Because each Z-score calculation is based on the mean for that climate division, Z-scores will correctly identify raw data values that are extreme. Top and tail analysis, as currently used by NCEI to calculate the CEI, identifies a much larger portion of the United States that is extreme; in reality the majority of the values highlighted as extreme really are not extreme. The portion of the United States experiencing extreme values is actually much smaller.

Like the map displaying maximum temperature extremes (Component 1), the map from NCEI for minimum temperatures (Component 2) also raises many questions. Values that NCEI considers to be extreme include the upper and lower 10^{th} percentiles of the dataset. This would imply that about 10% of the United States should be shaded, but more than half of the country is shaded in the map from NCEI (see Figure 15).

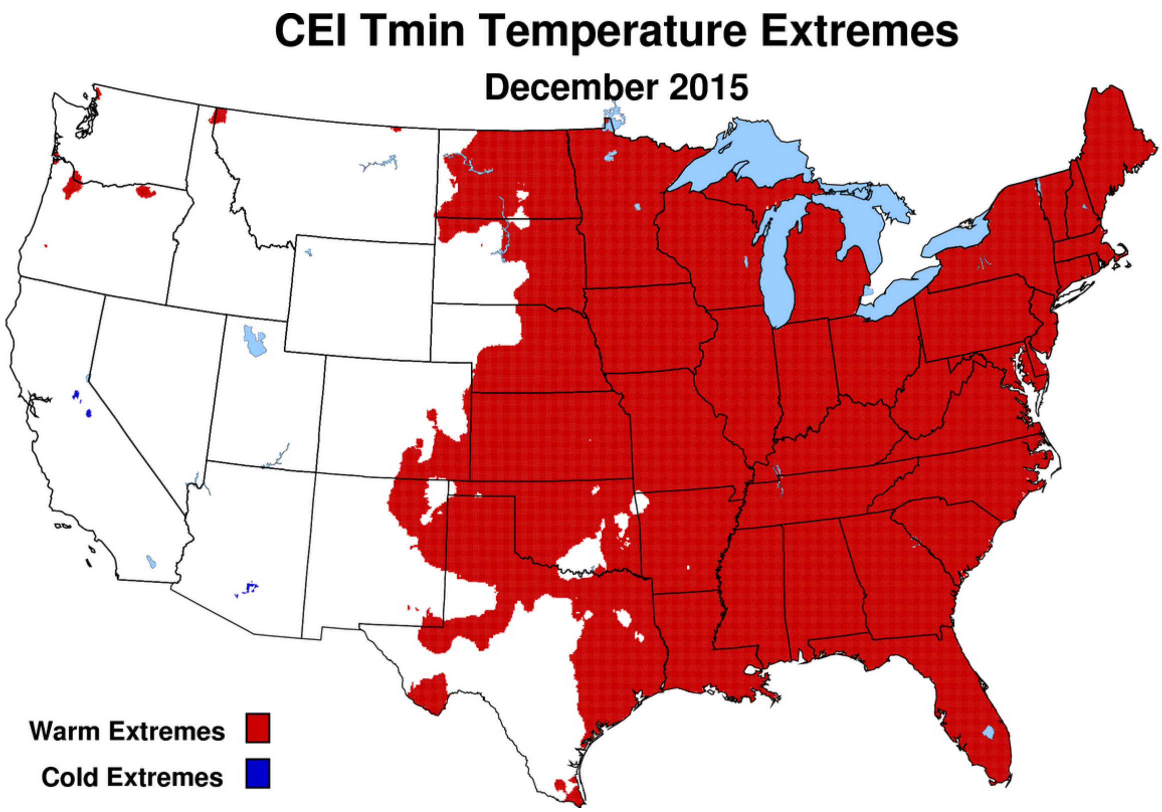

Figure 15: December 2015 Component 2 map (courtesy of Karin Gleason, NCEI)

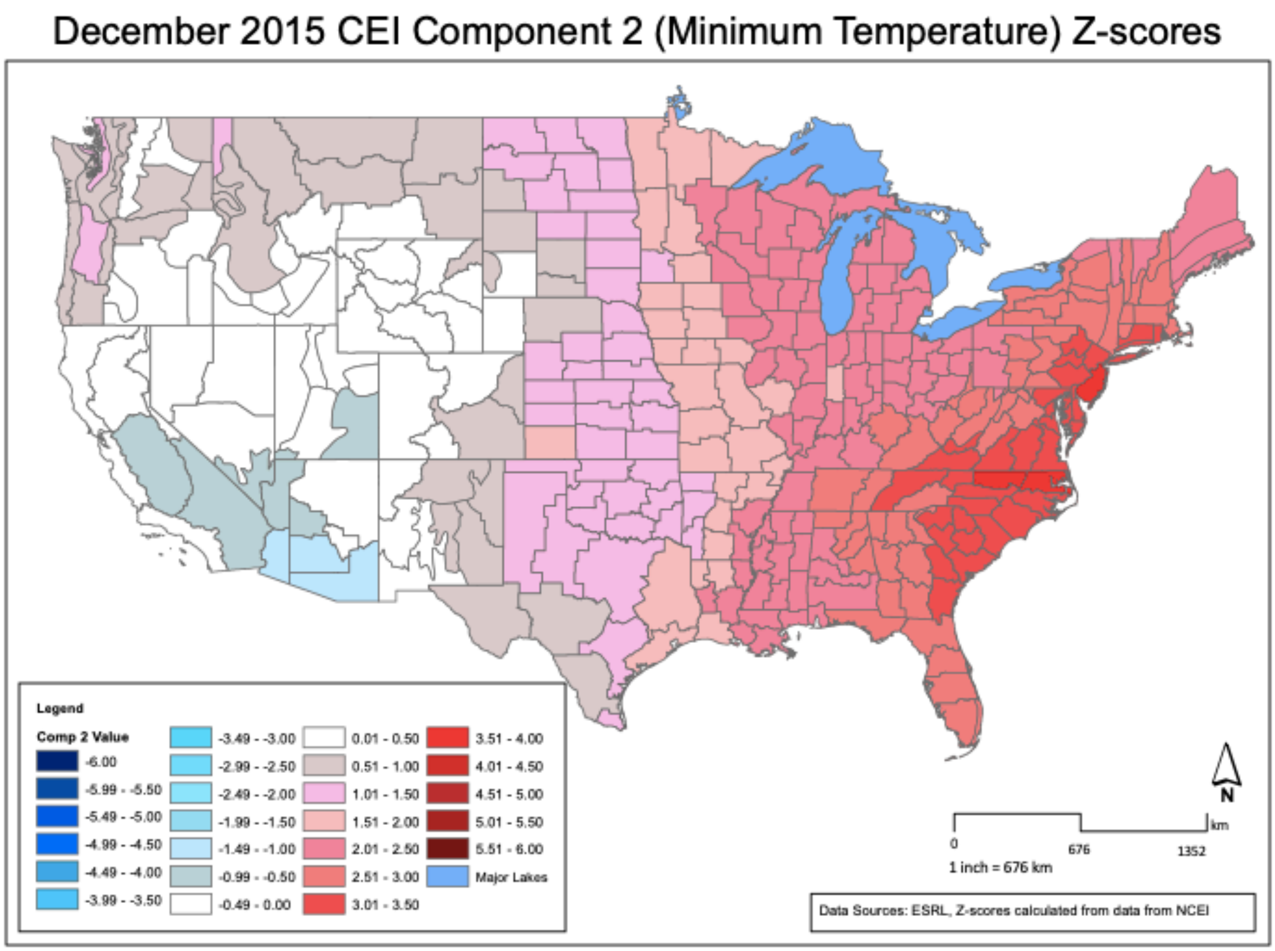

Figure 16: New December 2015 CEI Component 2 values for all 344 divisions of the United States.

3. *Component 3*

Continuing with the analysis of individual CEI components for December 2015, additional analysis in R tested the normality of the distribution of values from Component 3: PDSI values. As a reminder, the PDSI is an index that measures moisture availability. Positive values indicate moisture surplus, and negative values indicate drought conditions. The analysis in R found that 94.18% of the data lie within 2 standard deviations of the mean, and 99.71 % lie within 3 standard deviations (see Table 6). This means that there is only a very small percentage of the data that can be considered unusual or extreme for December 2015. Unlike the first two components, this month's data distribution is actually normal as compared to the mean and standard deviation values of the 30-year normals period of 1981-2010. This is also the only component to produce this result for the month of December 2015.

The histogram of this data distribution also demonstrates that this data distribution is fairly normal (Figure 17). The majority of values are concentrated around 0, mostly within 1 standard deviation. This data really isn't skewed at all, and only a small percentage of the values (a little less than 6%) are unusual or extreme for this month. This means that the distribution of the data for this month, as compared to that of the 30-year normals, is actually normal.

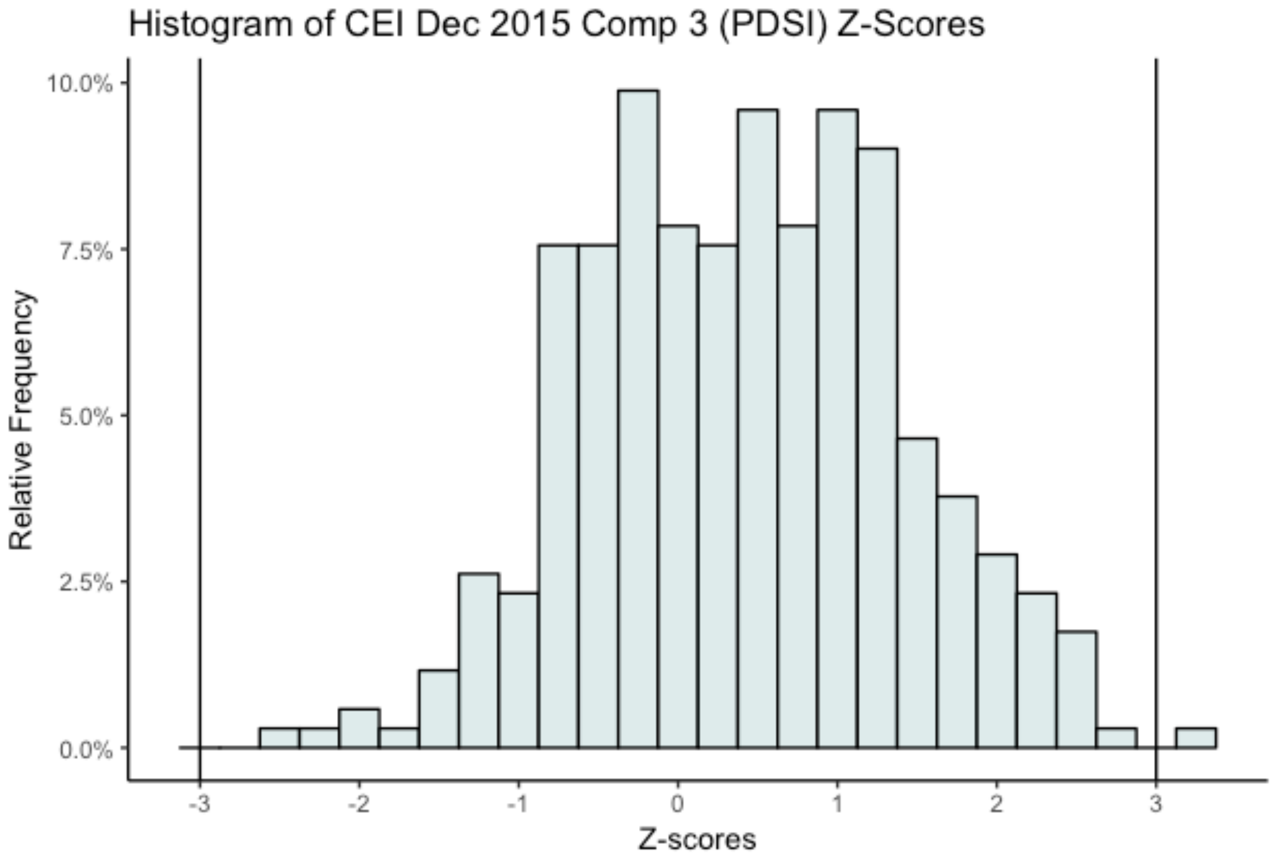

Figure 17: Histogram of the values for the CEI, Component 3 for December 2015

A map of this component's values was also created using ArcMap, and was compared to an image of the month's PDSI values downloaded from the NCEI website. For this component, NCEI considers the PDSI values in the upper and lower 10th percentiles to be the extreme values. Therefore, in the NCEI map (Figure 18), there is no indication of which climate divisions are extreme for this component. The highest values are located in the central United States, e.g., Illinois, Iowa, Oklahoma, and Texas, and the lowest values are located in southern California. These values are presumably the values that NCEI would consider to be extreme, but this image gives no indication of where the extreme values actually are. However, the map created based on values for the re-calculated CEI provides users with information about how extreme each value is, and where they are located (Figure 19). The only value that is extreme for this month is located in Oklahoma's climate division 6, with a Z-score value of 3.18. This means that the PDSI value for that climate division for December 2015, 2.91, is extremely above normal. This indicates that this climate division had a moisture surplus for this month.

Several states had values that were unusual, colored in pink, but there is no area where they are concentrated. States with unusual values include the Carolinas, Texas, and Oklahoma. These areas, with Z-score values with absolute value greater than 2, indicated unusually high levels of moisture, but were not high enough to be extreme. This is the component that had the highest amount of variability in the range of Z-score values, as this component was the only one to actually have negative values.

Even though this component did not have many extreme climate divisions, this shows how Z-scores can be used to identify climate divisions that are extreme, and ignore those than are not. Rather than just plotting values of the PDSI, this new method of calculating the CEI shows how the value per month compares to a mean value, giving an indication of normality of that value.

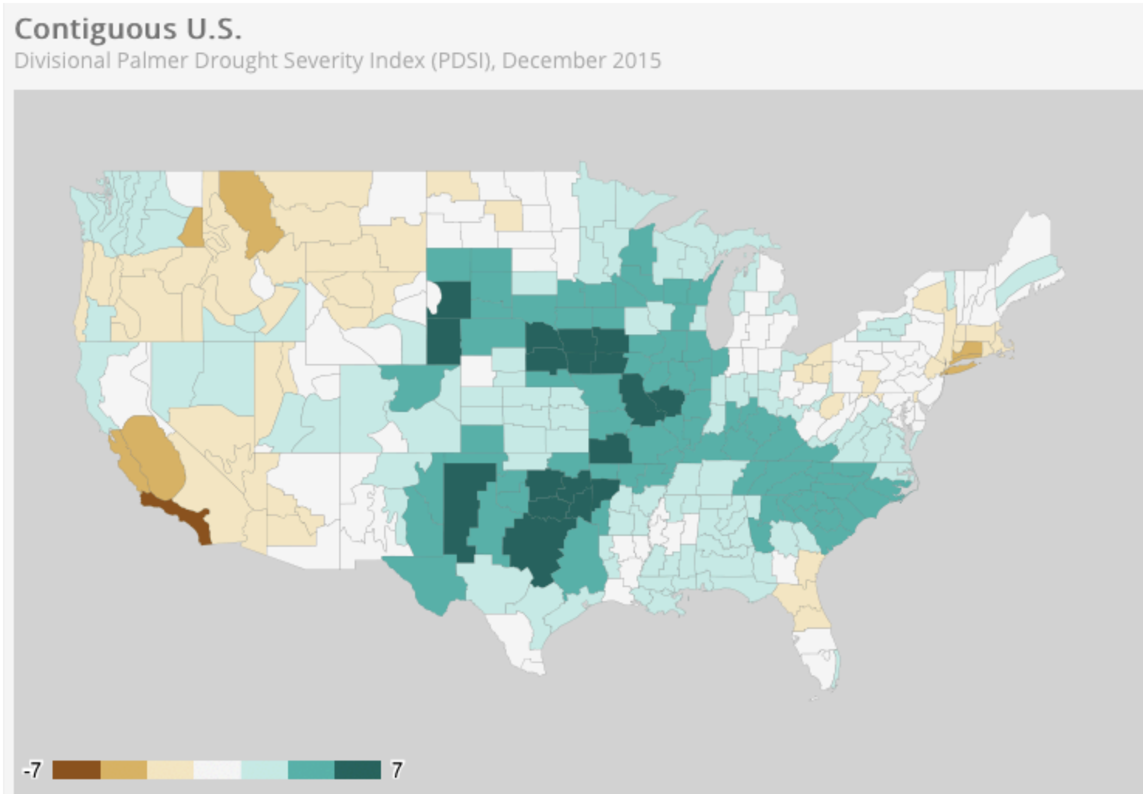

Figure 18: December 2015 PDSI values map from NCEI (http://www.ncdc.noaa.gov)

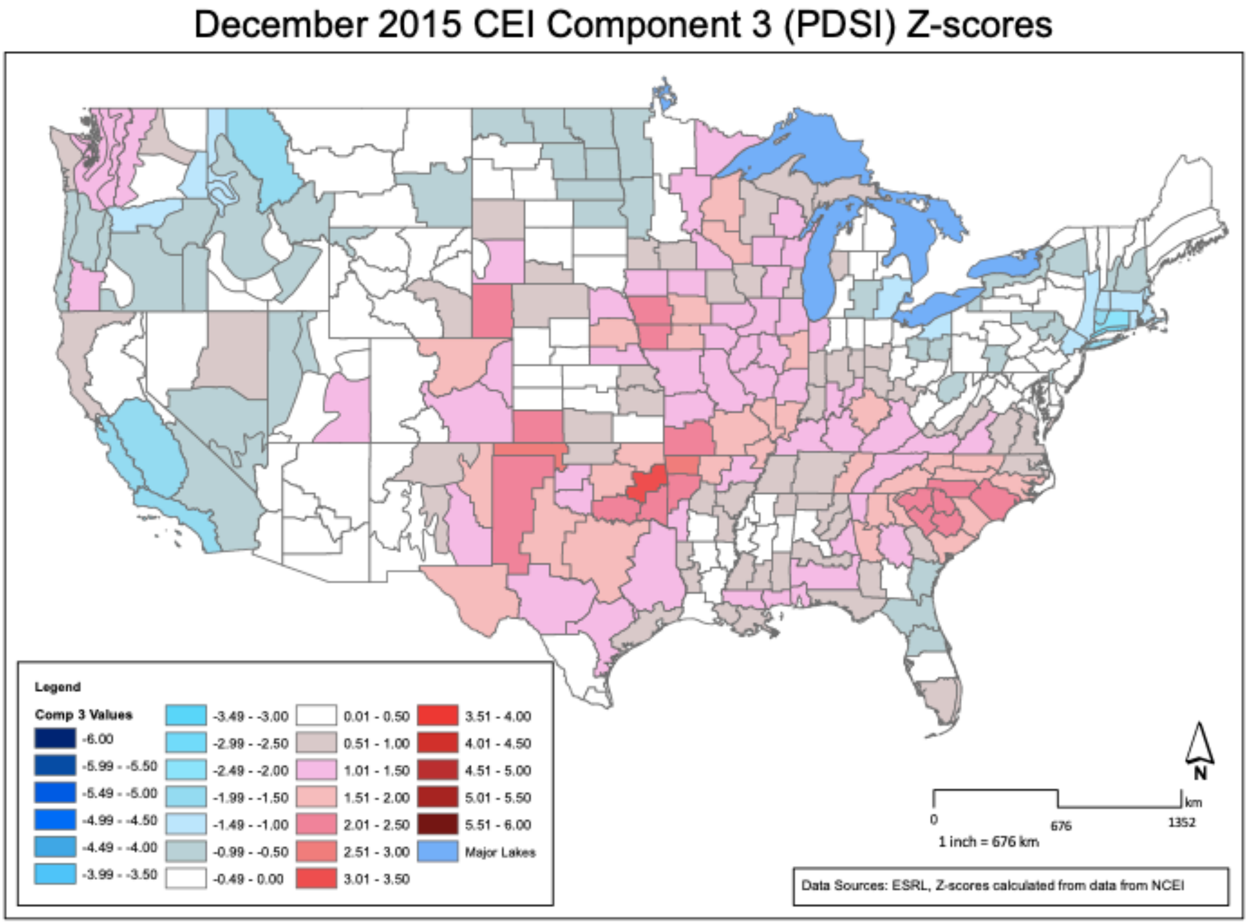

Figure 19: New December 2015 CEI Component 3 values for all 344 divisions of the United States.

4. Component 4

Finally, the distribution of the final component of the new CEI, monthly precipitation totals, was also analyzed in R to determine whether or not it was normal in relation to the 1981-2010 distribution. Analysis revealed that 72.67% of the Z-scores for this component fall within 2 standard deviations of the mean (Table 6). Because Chebyshev's Theorem states that at least 75% must lie within 2 standard deviations of the mean, this data distribution, in relation to the 1891-2010 values, is not normal. This analysis also revealed that 86.05% of values for this component lie within 3 standard deviations of the mean. This further confirms that the Z-score distribution of this component for December 2015 is not normal, because Chebyshev's Theorem states that 89% of the data must lie within 3 standard deviations. This means that just under 14% of the Z-scores for December 2015 from this component were extreme. This is the component with the highest percentage of extreme values for this month. The abnormality of this distribution further confirms how unusual the month of December 2015 was.

The histogram helps to illustrate the abnormality of this data distribution (Figure 20). This histogram shows that, like Components 1 and 2, the distribution of the Z-scores for this component is heavily skewed towards the right end of the distribution. This means that the majority of the Z-score values were greater than the mean value for this month. Because this data distribution is skewed, this further shows that, for the month of December 2015, it is considered to be abnormal. This histogram also helps identify where the extreme values are within the distribution for this month. This histogram shows that there are no Z-scores less than -1, and all of the extreme values have a Z-score greater than +3. The solid black in the histogram marks the location of the Z-score value of +3. This helps to illustrate that all extreme values for this month were positive extremes, meaning that many climate divisions this month reported precipitation values that were extremely greater than normal.

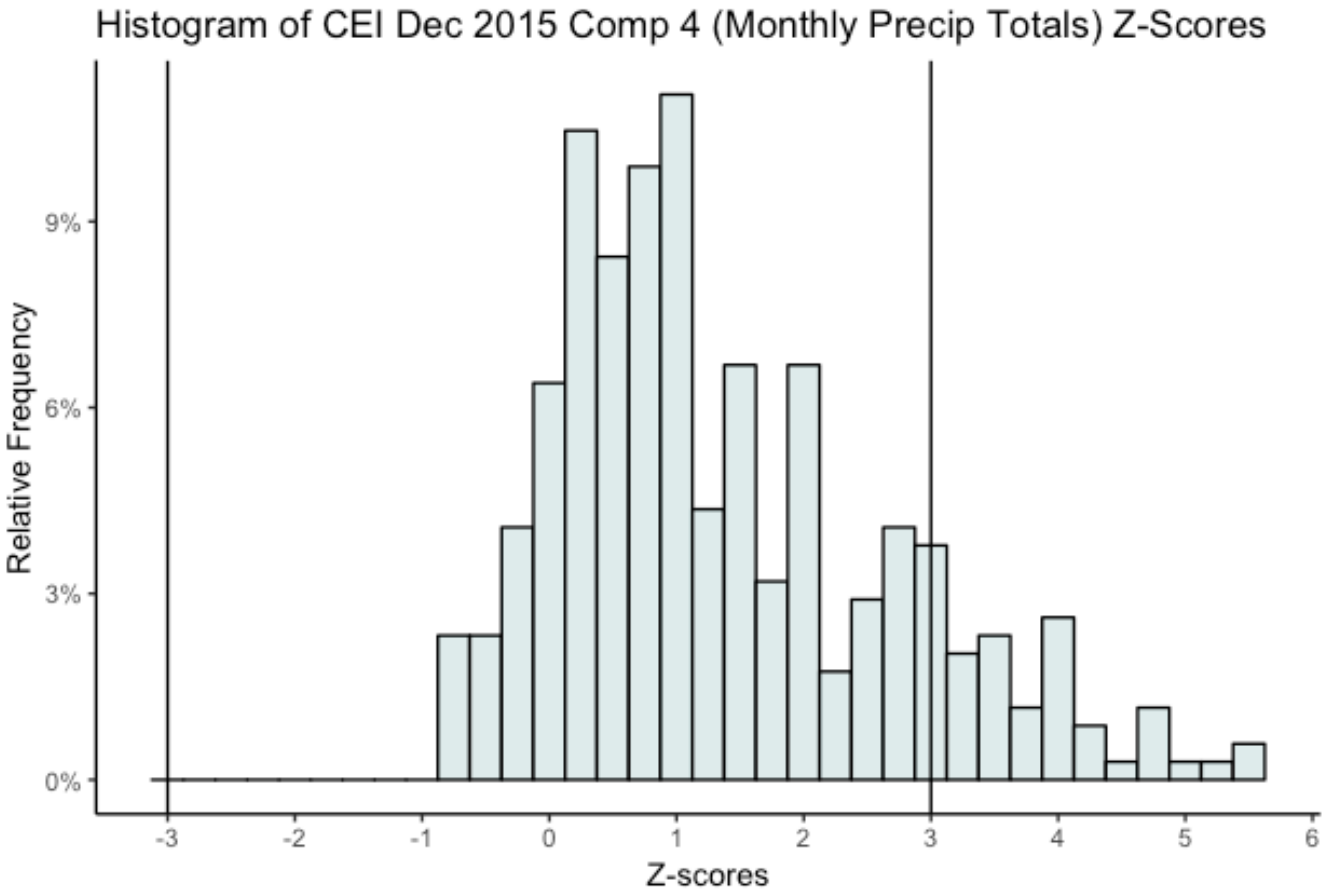

Figure 20: Histogram of the values for the CEI, Component 4 for December 2015

The comparison of different maps for this component also help to reveal where in the United States the extreme values for this component were located during December 2015, as well as illustrate the effectiveness of Z-scores at identifying values that are extreme. The map produced by NCEI (see Figure 21) gives only the precipitation totals per month per climate division. There is no indication on this map of values that are extreme during this month. However, the map created with the values for Component 4 of the new CEI actually helps to identify climate divisions that recorded totals for monthly precipitation values that are considered to be extreme (Figure 22). This map identifies two regions where the extreme values are concentrated. One region is the central United States, including Iowa, Illinois, Wisconsin, Missouri, and portions of surrounding states. The other region is the southeastern United States, with the extreme values concentrated in Alabama, Georgia, and the Carolinas. Both of these regions had climate divisions that reported precipitation values with a Z-score value with absolute values greater than 3. The histogram of this component shows that all of the extreme values were positive for this month, so all extreme values on this map were greater than +3.

As an example, the state that had the highest concentration of extreme values was Iowa. The Z-score values across this state range from 3.07-3.51, meaning that the entire state recorded extreme precipitation values for December 2015. The division that recorded the highest Z-score value for this month was Iowa's climate division 7, the southwestern-most climate division of the state. This Z-score corresponds to a total monthly precipitation value of 5.62 inches. This value, as compared to the 1981-2010 December mean value for this division, is positively extreme for this climate division in the month of December, meaning that it is much greater than the normal value. The remaining Z-score values in Iowa correspond to precipitation values ranging from 3.06-5.37 inches. All of these values, as compared to the 1981-2010 mean value for that climate division, are considered to be monthly totals that are extreme for the respective climate divisions.

Like the maps from the other components, this map helps to further illustrate the effectiveness of using Z-scores to identify values that are extreme. Unlike the current method used in the calculation of the CEI to determine extreme values that only gives the percent of the United States experiencing extreme values, Z-scores can be used to determine two things. Z-scores give not only where the extreme values are actually located, but also tell users how extreme that value is based on the number of standard deviations the datapoint is from the mean. This map also demonstrates how this statistic also correctly identifies raw data values that are extreme and ignores values that are not extreme.

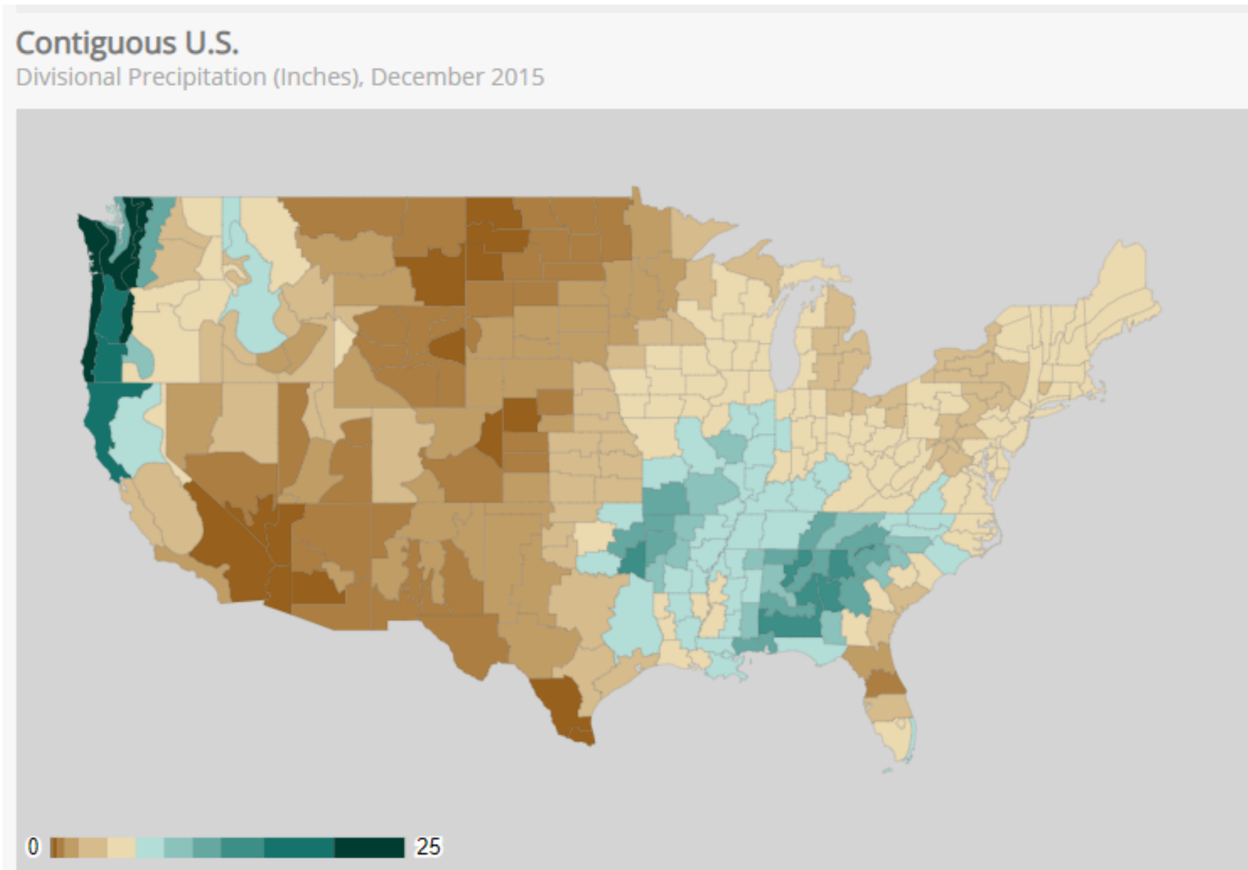

Figure 21: December 2015 Precipitation totals map from NCEI

(http://www.ncdc.noaa.gov)

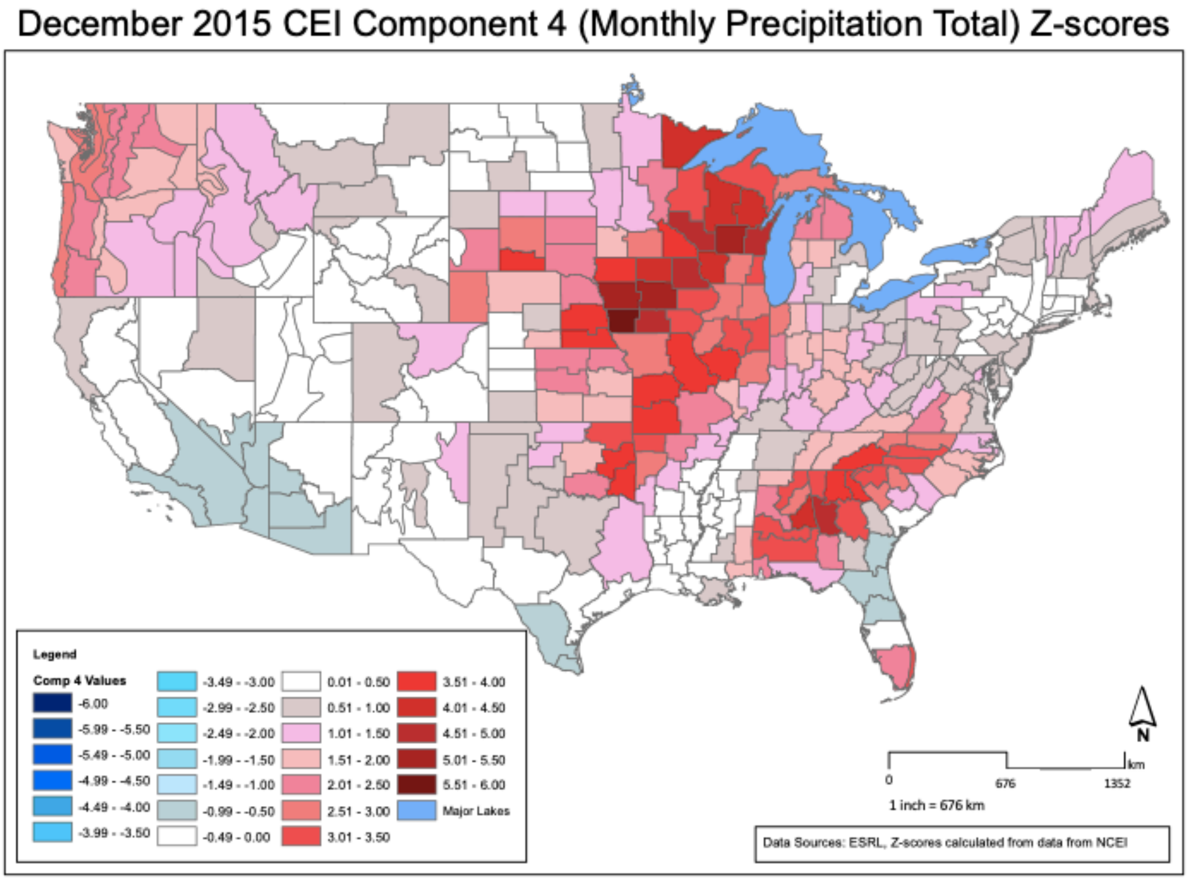

Figure 22: New December 2015 CEI Component 4 values for all 344 divisions of the United States.

5. Overall CEI

The values for the overall CEI for December 2015 are shown in Figure 23. While the majority of the United States did not have any extreme components during this month, several in the eastern and central United States had one, and there were five divisions that had two extreme components. One division that had 2 extreme components was Oklahoma’s climate division 2. The maps for each component reveal that the components that were extreme for this division were Components 3 (the PDSI) and 4 (total monthly precipitation). The positive PDSI value here means that this division had excess moisture during the month, and it also recorded a total precipitation amount that was extremely above average. The remaining climate divisions that had two extreme components during this month are located in the southeast. Climate division 5 in Georgia, climate divisions 4 and 5 in North Carolina, and climate division 2 in South Carolina all had two CEI components that were extreme during December 2015. All four of these divisions had the same extreme components for this month: Component 2 (minimum temperature) and Component 4 (total monthly precipitation). All four divisions also recorded upper-end extremes for this month, meaning that the raw data values were extremely above average. Since both minimum temperatures and precipitation amounts were extremely above normal, this supports the reports that this month was warmer and wetter than normal; it also confirms the abnormality of the El Niño that occurred during this month.

Even though these climate divisions only had two components each that were extreme, breaking down which components were extreme is important. Knowing which components were extreme for that month can have important implications when identifying potential risks associated with each extreme and also identifying populations that are vulnerable to being exposed to these extremes. For example, the National Weather Service in Tulsa, Oklahoma reported widespread major flooding that "devastated a large portion of eastern [Oklahoma] and northwest [Arkansas]" during the month (https://www.weather.gov). This can have devastating impacts for those living in the areas affected, especially if mobility or access to resources to repair the damage are limited.

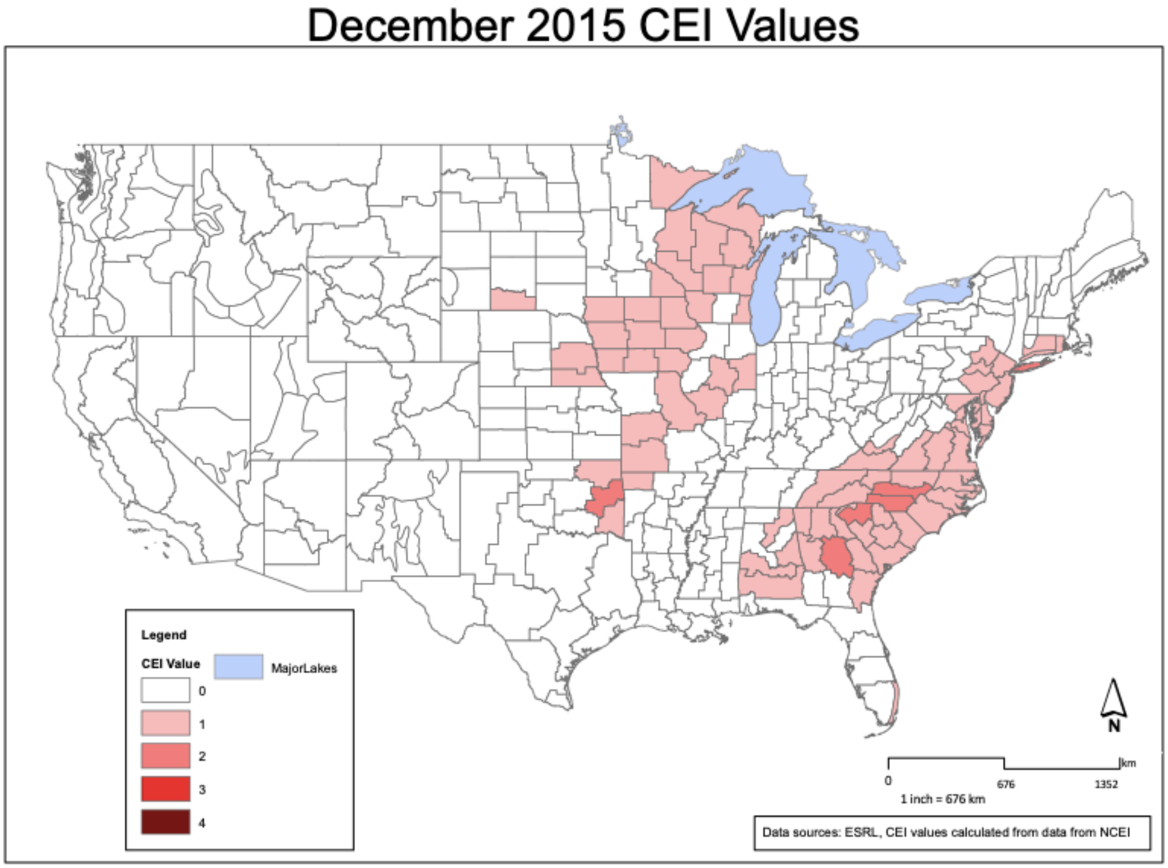

Figure 23: New December 2015 CEI values for all 344 divisions of the United States

C. Temporal Trends for CEI Components in December

In addition to analyzing the effectiveness of changing the spatial resolution at which the CEI is calculated, temporal trends in each component were examined. The month of December was selected for analysis to be consistent with the month selected for the case study discussed in Section B. Time series analysis was conducted in R with values from 1900-2017 for Georgia Climate Division 3. This division was selected because this is the division that Athens, Georgia is located in. For each image, the values are plotted in black, and a trend line, calculated with the Local Polynomial Regression Fitting method in R, is drawn in red. Z-scores are along the Y-axis and years are along the X-axis. The time series for Component 1 (Maximum temperature) is shown in Figure 24. This image reveals a steady increase in Z-score values beginning around 1980. This indicates an increase in maximum temperature values during the month of December deviating from the mean, and also an increased deviance from the mean. The time series for Component 2 (Minimum temperature), shown in Figure 25, reveals a similar trend. Like the Z-scores for Component 1, Z-scores for Component 2 have been increasing since 1980. This indicates that this component shows an increase in minimum temperature values that deviate from the mean, and that this deviation is increasing. Component 4 (Monthly Precipitation totals) displays a similar trend, but the increase in Z-scores does not occur until the 1990s (Figure 26).

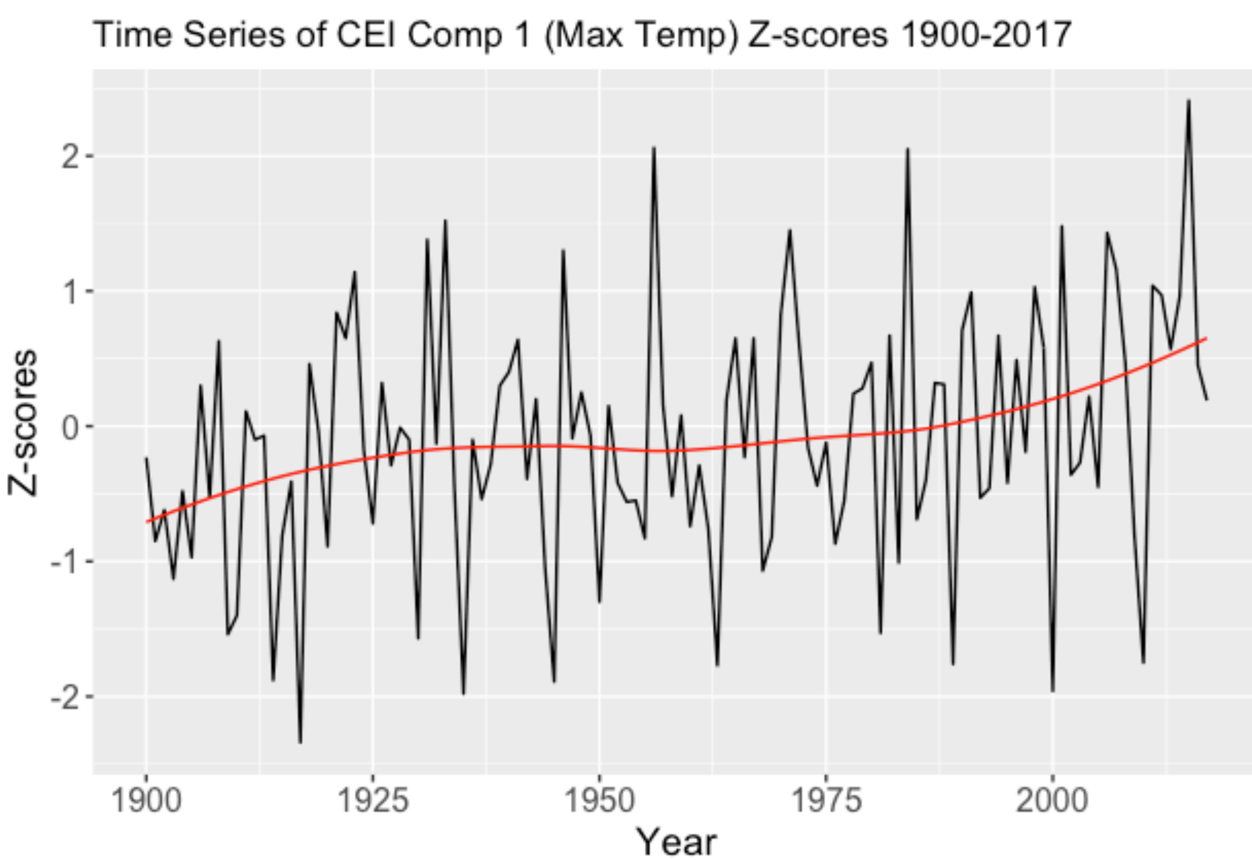

Figure 24: New December 1900-2017 CEI Component 1 time series for Georgia's climate division 3

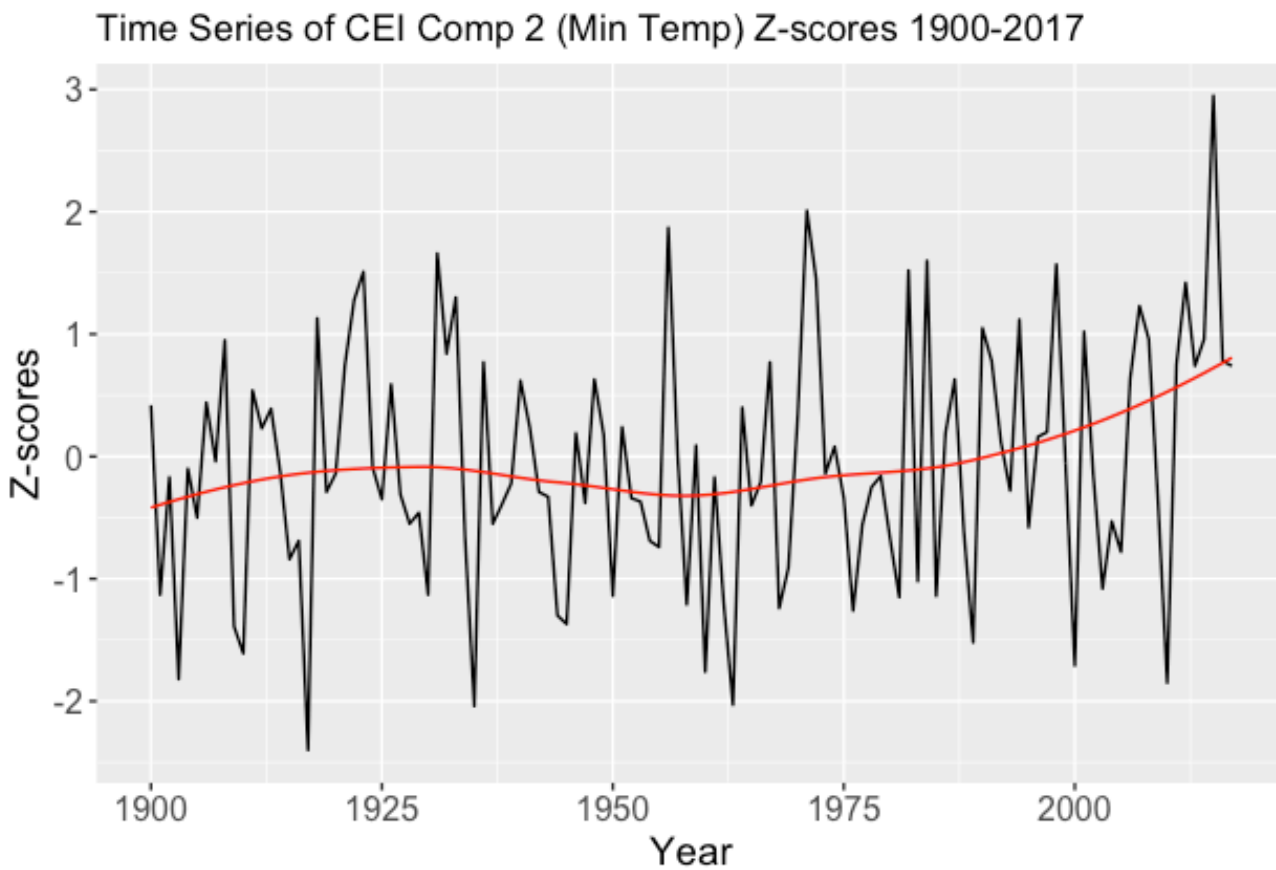

Figure 25: New December 2015 CEI Component 2 time series for Georgia’s climate division 3

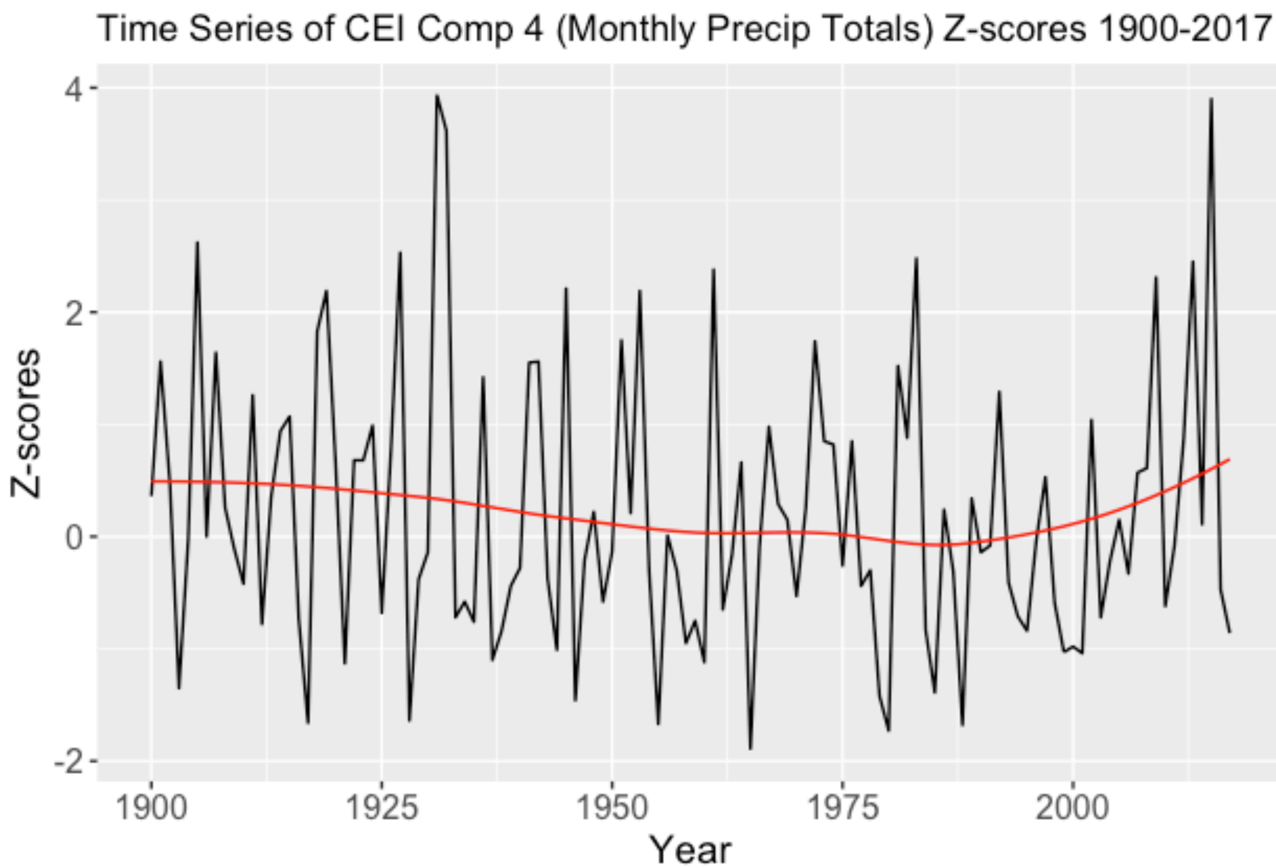

Figure 26: New December 2015 CEI Component 4 time series for Georgia's climate division 3

Component 3 demonstrates a trend different from that of the others. The trend line for Component 3 (PDSI, used for moisture availability), shown in Figure 27, reveals a decrease in Z-score values since the 1980s. Despite this decrease, the trend line remains fairly close to 0, indicating that values are not significantly deviating from the mean.

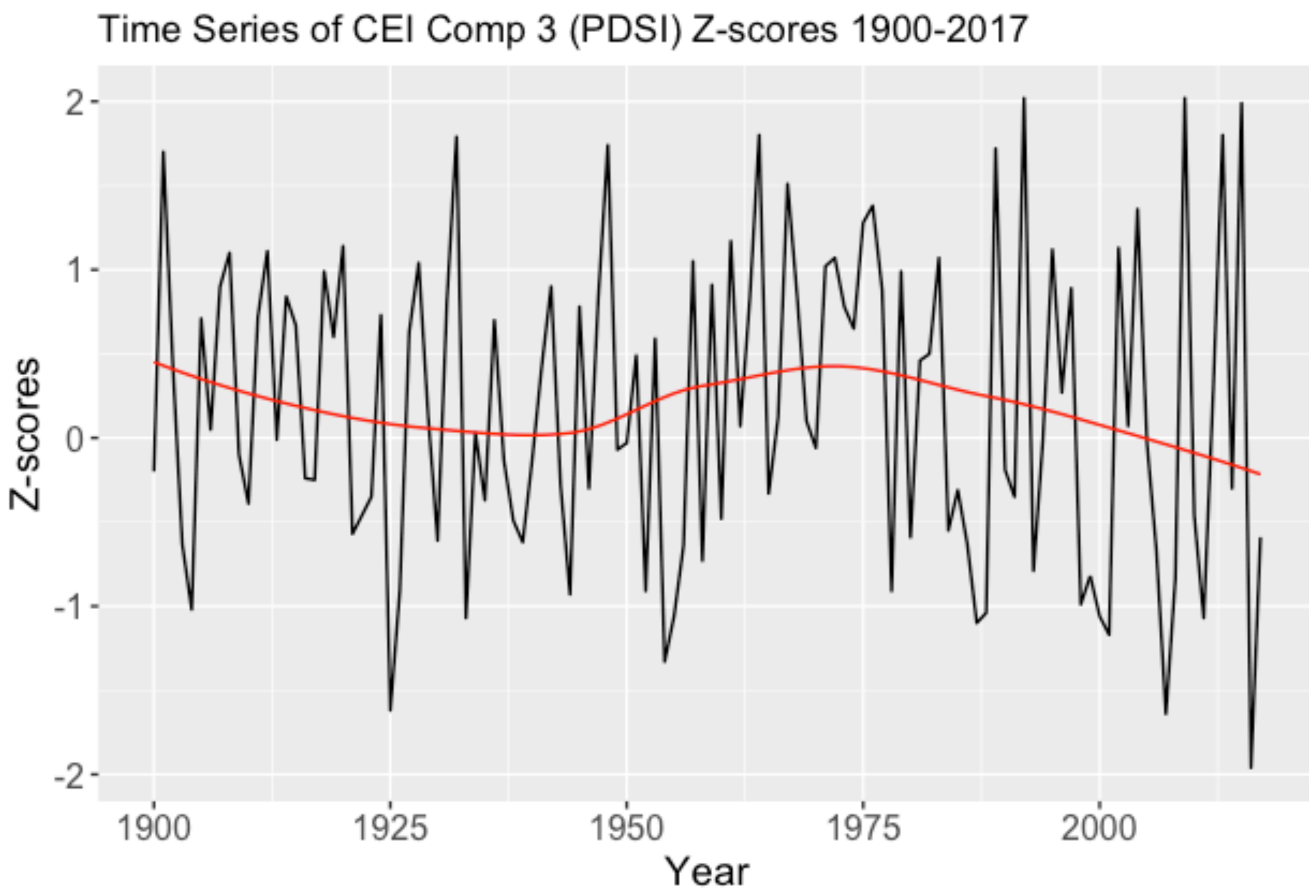

Figure 27: New December 2015 CEI Component 3 time series for Georgia's climate division 3

D. A Revised Social Vulnerability Index

Using ArcMap 10.6, maps of each of the four themes, as well as the final SVI, were produced. Again, the maximum Z-score value for each climate division is represented so that all divisions with vulnerable counties can be identified. Results from these maps are presented below.

1. *Theme 1: Socioeconomic Status*

Figure 28 shows re-scaled values for Theme 1: Socioeconomic Status. This theme includes variables like income and employment status. For 2016, the highest Z-score values are concentrated in the southern and southeastern United States. All these climate divisions had a re-scaled Z-score value of 2, meaning that the Z-score values were between 1 and 2. Since these were the highest Z-score values calculated for this theme, these climate divisions have the most vulnerable populations to socioeconomic variables. Practically, this means that these areas are low-income and have high poverty or unemployment rates. Reports from the Census Bureau confirm this; ranking all states by income and comparing these rankings to re-scaled Z-scores reveals that the ten lowest income states all had a re-scored Z-score value of 2, confirming that these are all low-income states (http://www.census.gov). Mississippi, Alabama, California, Louisiana, West Virginia, and New Mexico were among the ten states with the highest levels of unemployment in 2016 (http://www.bls.gov); all of these states had climate divisions with higher levels of vulnerability during this year, indicated by an adjusted Z-score value equal to 2. The climate division with the highest Z-score for this theme was Mississippi's climate division 4, located in west-central Mississippi, with a Z-score of 1.99. Since this climate division had the highest Z-score value as compared to the national mean for the 2016 data, this climate division has the highest population of socially vulnerable individuals for Socioeconomic variables. As a note, the two most vulnerable climate divisions for this theme were both located in Mississippi.

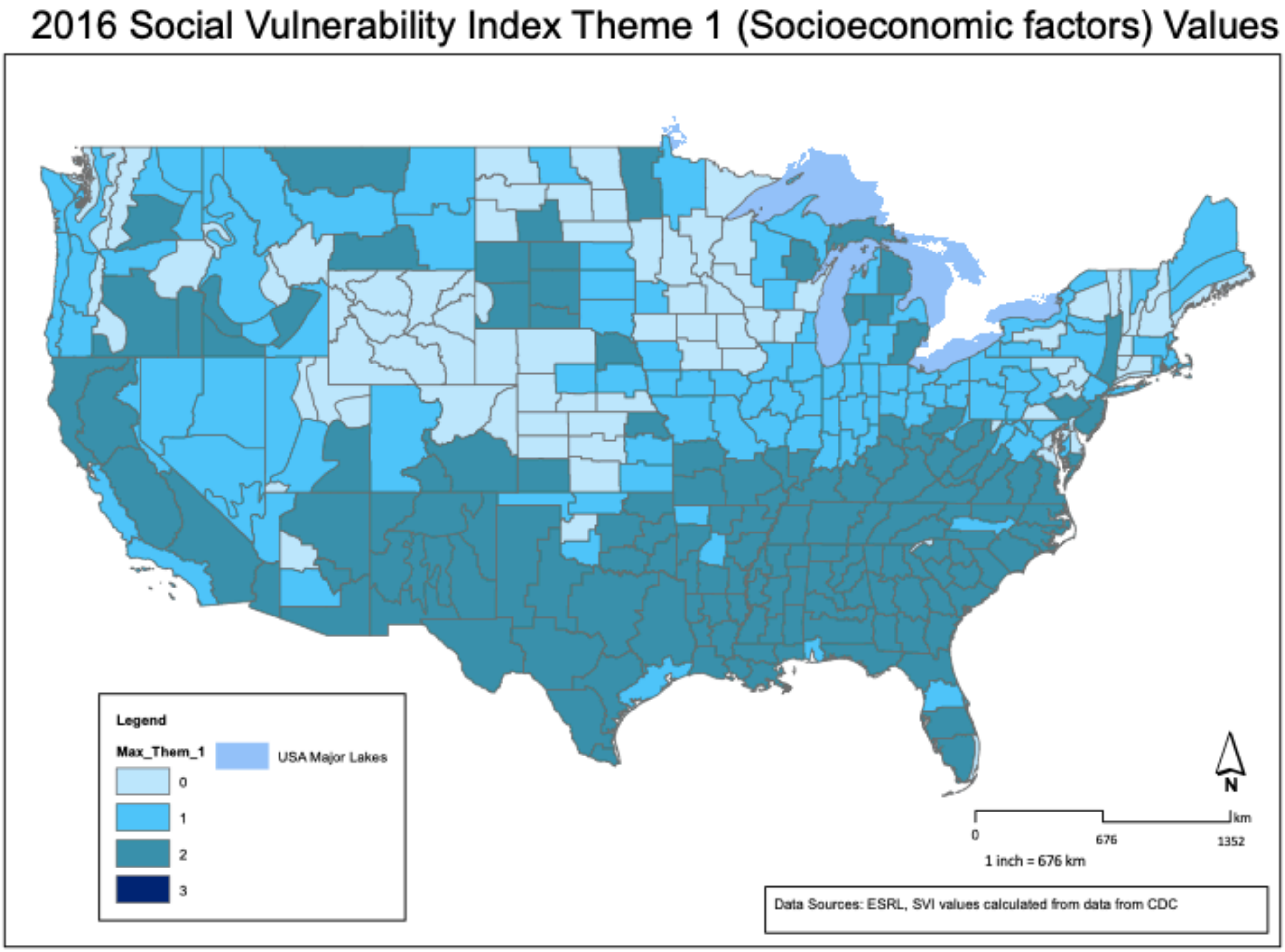

Figure 28: Theme 1 values for the new SVI for 2016 data

2. Theme 2: Household Composition and Disability

Resulting re-scaled Z-score values for Theme 2 (Household Composition) are mapped in Figure 29. This theme includes variables like race and an individual's disability status. Because negative Z-score values were eliminated, the re-scaled Z-scores plotted range from 0-3. For 2016, the areas with higher values are concentrated in the south and southeastern United States. The highest values, with a re-scaled value of 3, are located in select climate divisions in Texas, Oklahoma, Arkansas, Mississippi, Alabama, and Georgia. The re-scaled values correspond to a Z-score value between 2-3, and these are the climate divisions that deviated from the national mean value the most in 2016. Because they deviated the most, these divisions are considered to be the most vulnerable for this theme. Several statistics confirm this. A report from the Kaiser Family Foundation found that Texas has the second highest population of children (https://www.kff.org), and the Social Security Administration reported that the states with the highest percentages of individuals with disabilities are located in the south (https://www.ssa.gov); nearly all states in the southeast had at least one climate division with an adjusted Z-score value of 3. The climate division with the highest Z-score value for this theme was climate division 9 in Texas, located in south-southwest Texas, with a value of 2.85. This Z-score is the furthest from the mean value for this theme, which means that this climate division has the highest vulnerability value for Theme 2. This indicates that the highest population of vulnerable individuals in terms of Household Composition related variables is located here. Additionally, Texas had the three most vulnerable climate divisions for this theme.

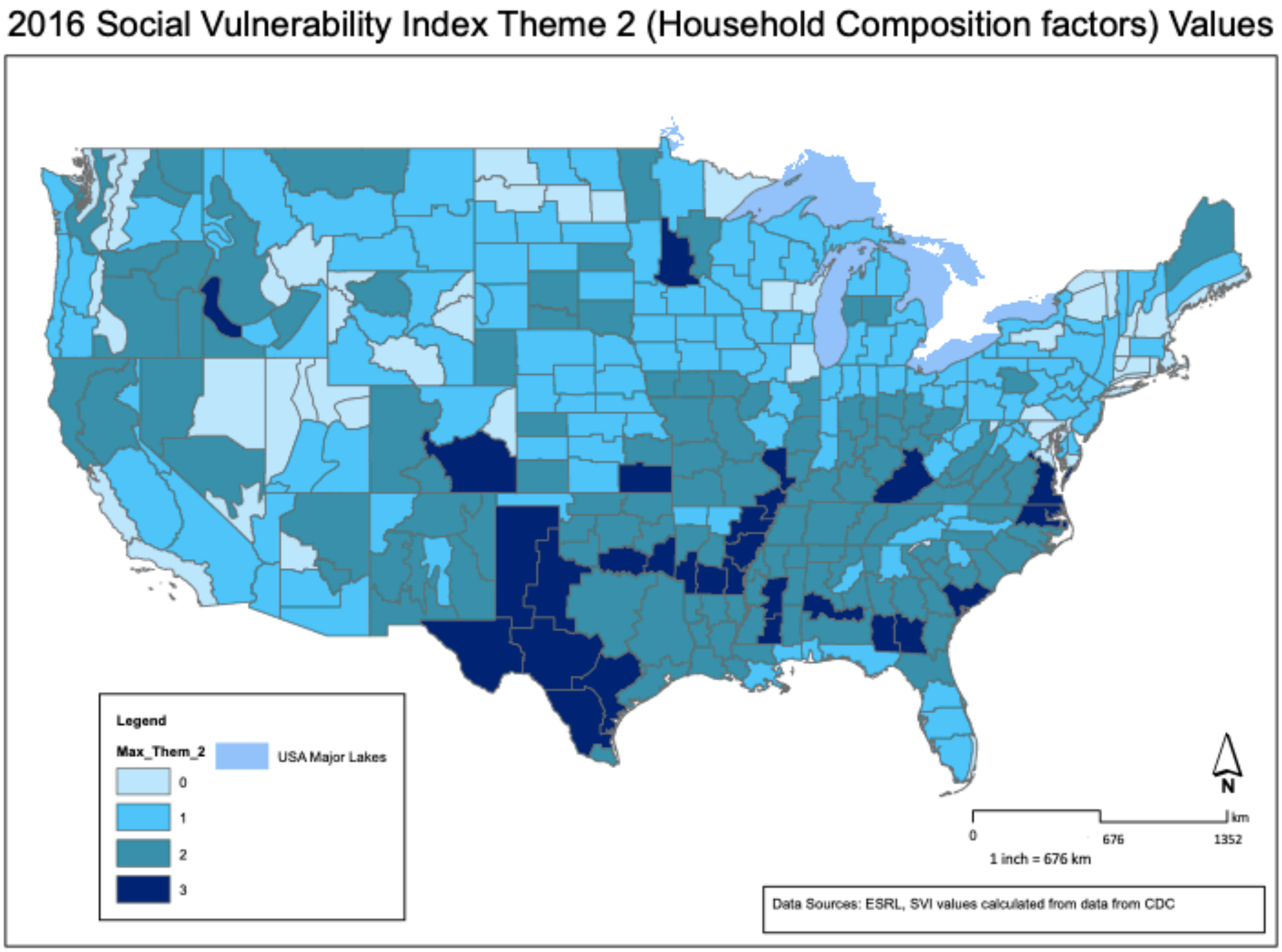

Figure 29: Theme 2 values for the new SVI for 2016 data

3. *Theme 3: Minority Status and Language*

Re-scaled Z-score values for Theme 3 (Minority Status) are displayed in Figure 30. Variables included in this theme are minority status and English-speaking-ability. Values displayed again range from 0-3 to highlight areas that are more vulnerable. Although areas with the highest re-scaled Z-scores are not as concentrated as the previous two themes, the highest re-scaled Z-scores are again located in states in the southeast and southwestern United States. There are also vulnerable climate divisions in the northwest.

These re-scaled values correspond to Z-scores that are between 1-2. Because these are the highest Z-score values recorded for this theme, these climate divisions deviate from the mean the most and represent the areas that have the highest vulnerability from the perspective of minority status. California, New Mexico, and Texas reported the highest minority populations (thoughtco.com), confirming what is seen in the map. These were also the states that reported the highest percentage of individuals that speak a foreign language (https://cis.org), helping to confirm the adjusted Z-score values seen in the map for all three states. Texas's climate division 9 recorded the highest Z-score value of 1.94, making it the climate division that deviated from the mean the most for 2016. This division was also the most vulnerable for Theme 2. Because of its higher Z-score, this division has the highest concentration of individuals that are vulnerable because of their minority status or English-speaking ability. Another climate division in Texas had the second highest Z-score for this theme, meaning that for Theme 3 the two most vulnerable climate divisions are both located in Texas.

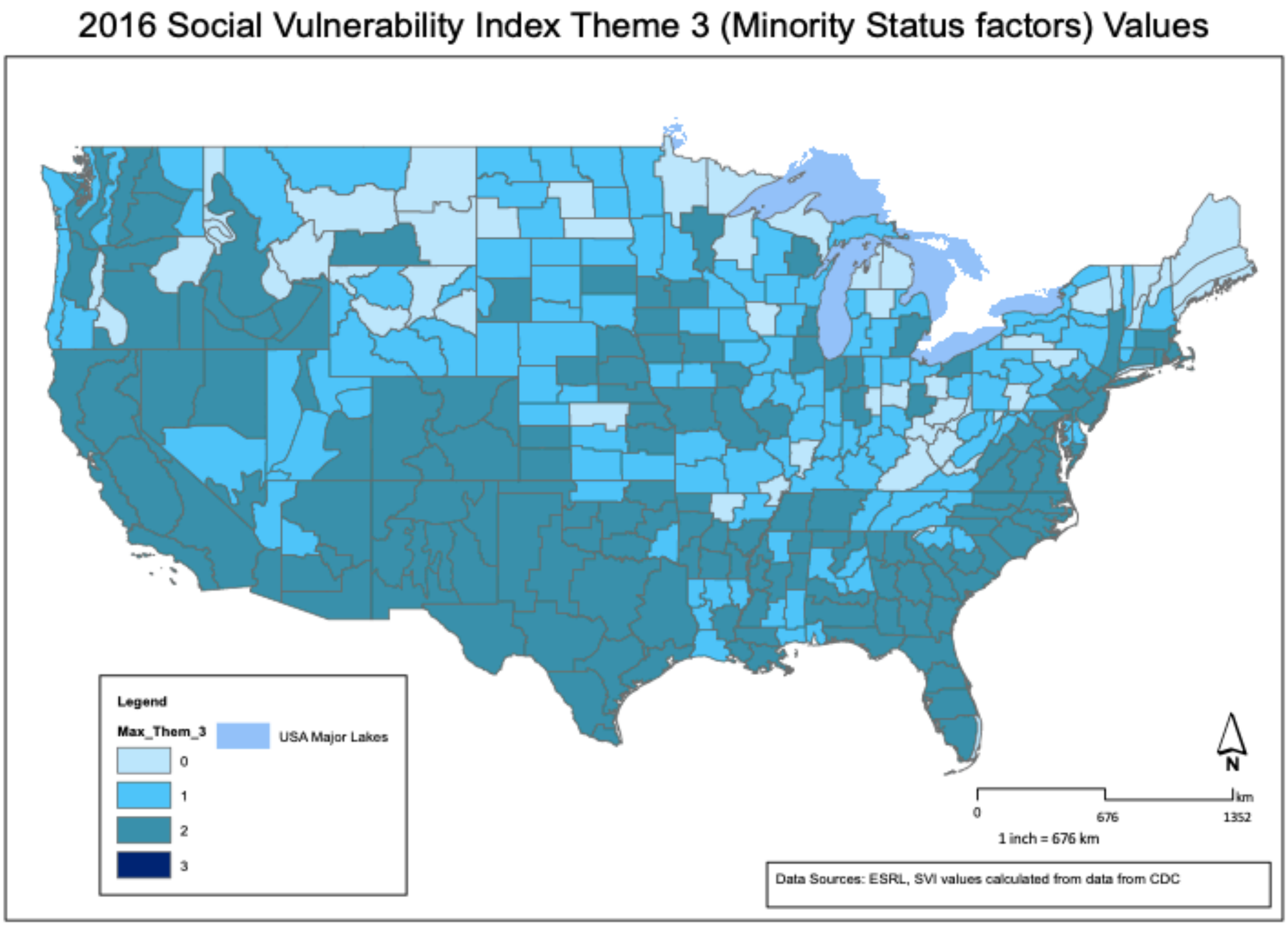

Figure 30: Theme 3 values for the new SVI for 2016 data

4. Theme 4: Housing and Transportation

Values for the final theme (Housing and Transportation) are displayed in Figure 31. Variables accounted for in this theme include number of mobile homes, number of homes without access to a vehicle, and crowding in a household. Values displayed again range from 0-3. Areas with the highest re-scaled Z-scores are concentrated in the southern and western United States. The value in these climate divisions is either 2 or 3, indicating a Z-score that is further from the 2016 national mean value for this theme. Climate divisions with a re-scaled value of 3 are located in Texas, Florida, Mississippi, California, Oregon, and Idaho, to name a few. Of all the states with at least one climate division with a re-scaled Z-score of 3, South Carolina, New Mexico, West Virginia, Alabama, Kentucky, Mississippi, and Arkansas are included in the list of the ten states with the highest percentage of mobile homes (http://www.statemaster.com). This is consistent with what is seen in the map. All of these divisions had a Z-score value between 2-3, meaning that these divisions deviated the most from the mean value for 2016. Because of this, these divisions can be considered to be the most vulnerable in terms of housing related variables. Climate Division 7 in Mississippi had the highest Z-score for this theme of 2.78. This means that this climate division was the most vulnerable division for this theme in 2016, indicating a higher population of individuals living in lower-quality homes or without access to reliable means of transportation.

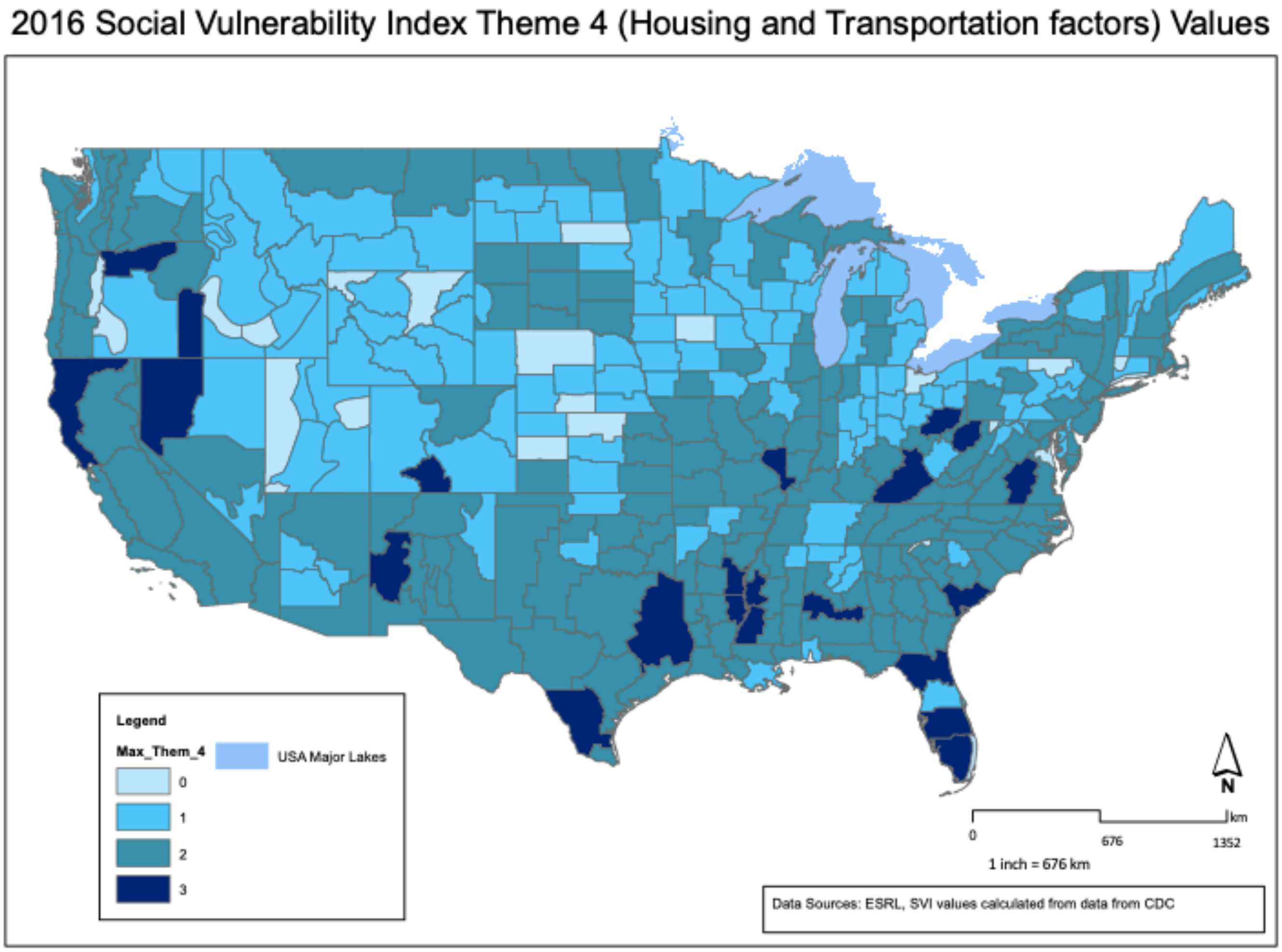

Figure 31: Theme 4 values for the new SVI for 2016 data

5. Overall SVI

Overall SVI values are shown in Figure 32. Overall values were calculated by adding the Z-score values of each theme per climate division together. The resulting index ranges from 0-10, due to the largest re-scaled values per theme observed (3 for two themes, 2 for two themes). Divisions with a lower overall SVI value are less vulnerable, and divisions with a higher overall SVI value are overall more vulnerable. To identify the themes that each division was most vulnerable to, one merely needs to look at the maps for each theme.

Looking at the Overall SVI map shows that areas of the United States with the highest levels of vulnerability, based on 2016 county-level data, are located in the south and west United States. Almost all climate divisions in these regions had an overall SVI value of at least 7, indicating that the re-scored Z-score value for at least three themes in these climate division was equal to 2. For two themes, this value represents the highest level of vulnerability. Four climate divisions had an overall SVI value of 10, meaning that these climate divisions had the highest re-scored Z-score value possible for all themes. These climate divisions are Climate Division 9 in Texas, Climate Division 9 in Arkansas, Climate Division 6 in Alabama, and Climate Division 7 in South Carolina. These four climate divisions are the most vulnerable based on 2016 census data, and indicate the areas that are the most predisposed to be vulnerable to climate extremes.

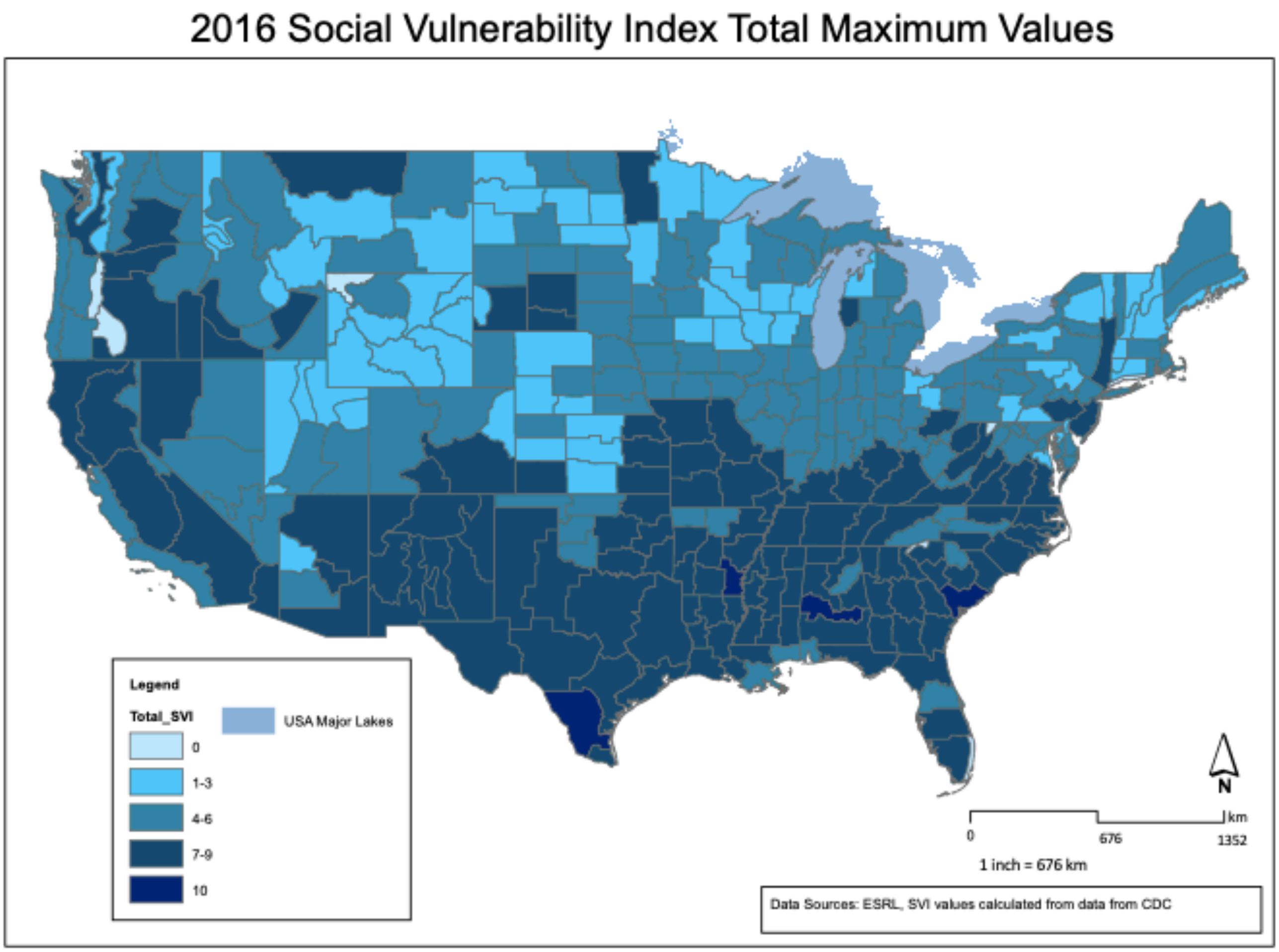

Figure 32: Overall Maximum SVI values for 2016 data

E. The Extremes Vulnerability Index

Values for overall vulnerability are shown below (Figure 33). This analysis identifies two regions that overall are more vulnerable to climate extremes than the rest of the United States for December 2015. The southeast United States and the central United States are both regions that had values of 5 on the numerical scale of vulnerability. However, this numerical scale does not indicate which index is contributing more to overall vulnerability. For this reason, the EVI was then created.

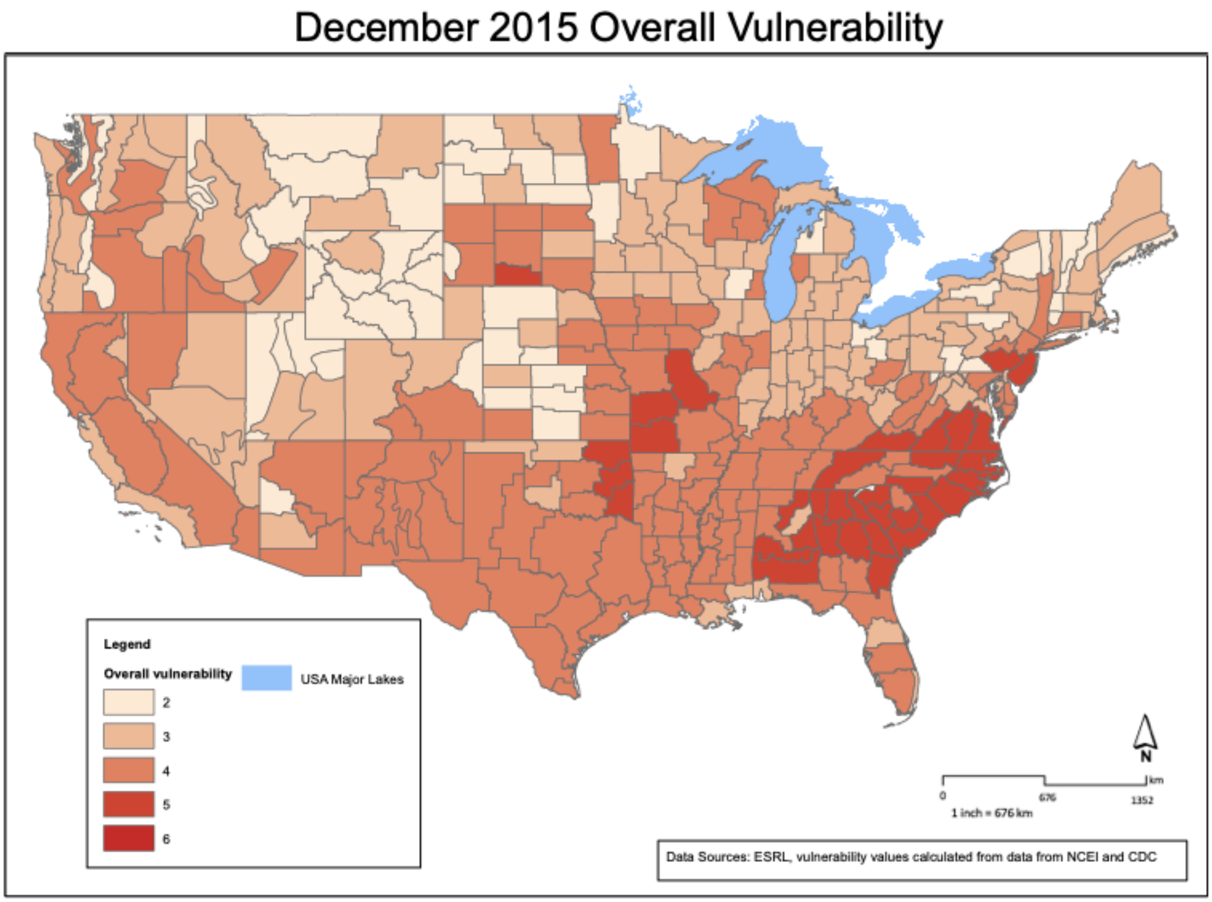

Figure 33: Overall Extremes Vulnerability for December 2015

Values from the CEI and SVI were combined for the first time to produce an Extremes Vulnerability Index (EVI) to identify climate divisions that are the most vulnerable to climate extremes. A bivariate scale was used to determine which index was contributing the most to overall vulnerability. The advantage of using bivariate analysis when comparing these two indices is being able to identify both the level of physical exposure that each climate division faces and the pre-existing conditions that can lead individuals living in these divisions to be vulnerable. For a full explanation of the calculation of the index and an explanation of each class, see Chapter 4, section E. Values of the EVI for December 2015 CEI values and 2016 SVI values are shown in Figure 34.

Classes 6, 8 and 9 indicate higher vulnerability, because this means that a climate division recorded a high value for at least one index. A value of 9 means that high levels of vulnerability were recorded for both indices, 8 means that a climate division had a high SVI value, and a 6 indicates that the climate division recorded a high CEI value for that particular month.

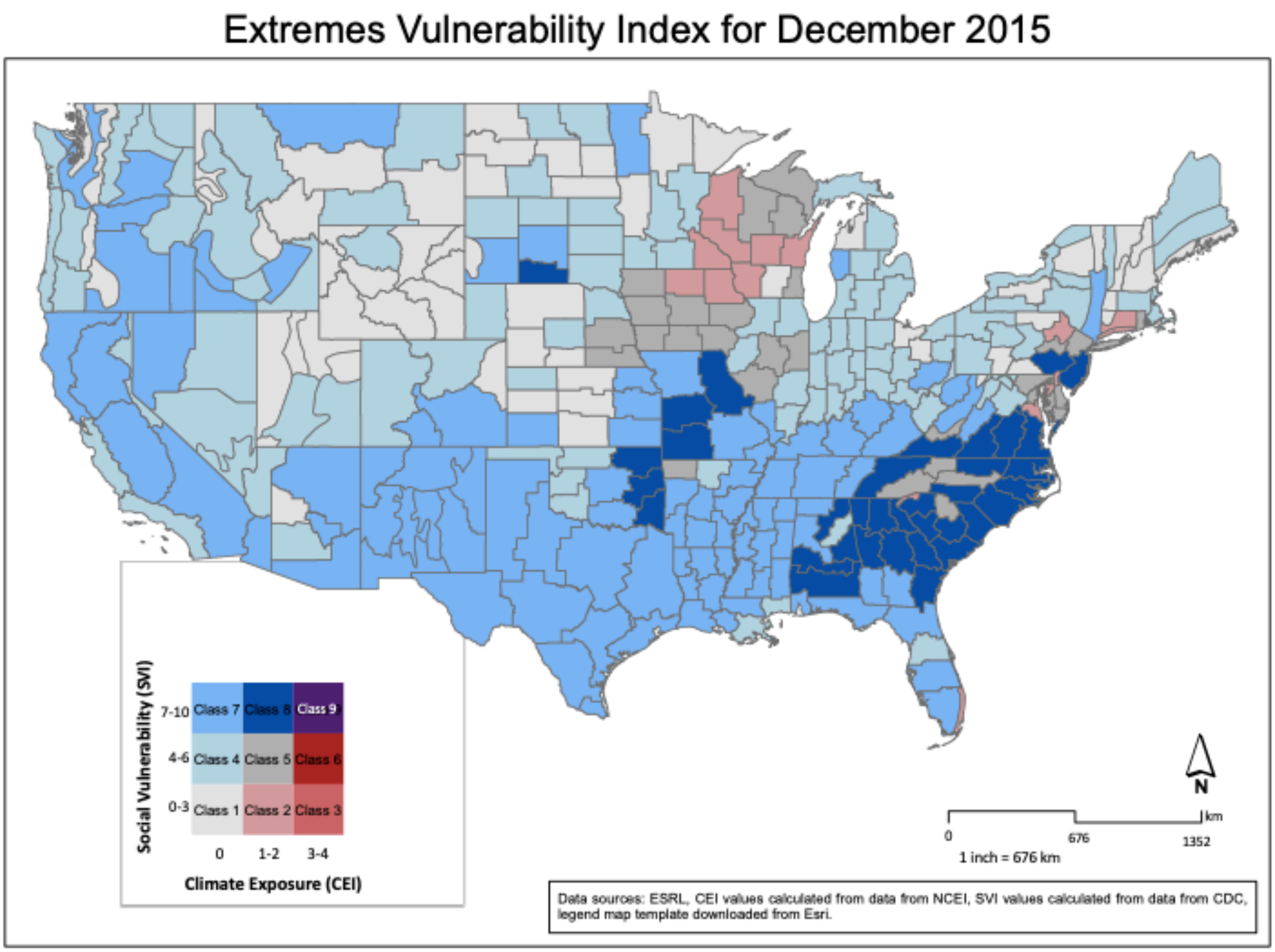

Figure 34: EVI values for December 2015

This figure identifies two regions of the United States that are overall more vulnerable during this month. One larger region is in the Southeast United States and includes climate divisions in Virginia, Tennessee, North and South Carolina, Georgia, and Alabama. All of these divisions had an EVI value of 8, indicating that the SVI value was high (between 7-10) and that they had a medium CEI value (either 1 or 2). Since the SVI value was ranked higher than the SVI value, this indicates that social variables are contributing to the overall vulnerability of these climate divisions slightly more than physical exposure. Knowing the factors that make a climate division more vulnerable can then be used to create policies to help mitigate risk and better equip these communities with resources to respond to climate extremes.

There are a number of practical applications of the information conveyed with this index. For example, looking at the maps for the SVI themes reveals that all of these climate divisions recorded high Z-score values for Themes 1 (Socioeconomic variables) and 2 (Household Composition) of the SVI. This could indicate a high percentage of the population of the climate division that is low income, and less likely to live in structures with air conditioning (Diem et al. 2017), as well as a higher number of individuals older than 65, who because of age are less efficient at regulating body temperature (https://www.cdc.gov). These areas also overlap with areas that recorded extreme values of minimum temperature for December 2015. This highlights regions that may be more vulnerable to heat extremes, meaning that these climate divisions do not have effective mechanisms and resources to limit heat exposure. This knowledge can then be used to create and implement policies and resources that help individuals in these climate divisions to mitigate their risk to heat exposure.

The second region of higher vulnerability is located in the central United States and includes climate divisions in Missouri and Oklahoma. These climate divisions also recorded EVI values of 8, meaning that the SVI value was high and that these divisions recorded a medium CEI value. Looking at the breakdown of values of each index can help to explain why these divisions are vulnerable and identify the risks that could most impact them. Looking at the maps for each theme of the SVI shows that these divisions in Oklahoma and Missouri had higher Z-score values for Themes 2 (Household Composition) and 4 (Housing and Transportation). These high SVI theme values overlap with climate divisions that had extreme precipitation values for December 2015 and experienced major flooding during this month. Many local National Weather Service offices reported major flooding during this month (https://www.weather.gov), and this would lead to the need to evacuate all residents from affected areas. If the population of these climate divisions has a high percentage of individuals over the age of 65, as well as a high percentage of individuals without access to reliable means of transportation, it may be harder for these individuals to evacuate. If these individuals are not able to evacuate, the result could be loss of life. Using this index to determine areas that have the most vulnerable populations can also be used to distribute resources and aid most effectively when an extreme event occurs.

An additional benefit of using this new combined index is that it can help to indicate potential health risks to vulnerable climate divisions. Since maximum and minimum temperatures across the vulnerable divisions of the southeast United States were higher than normal, with many of these areas recording extreme minimum temperatures for December 2015, these climate divisions could be more susceptible to vector borne diseases that require warm conditions to spread. For example, higher minimum temperatures and increased precipitation create an environment favorable for the growth of the *Aedes Aegypti* mosquito (Christophers 1960), the mosquito that carries the Zika virus. In 2016, there was an increase in the number of Zika cases reported in the

United States, and many states in the southeast recorded high numbers of cases (https://www.cdc.gov). Using an index like the EVI to identify regions that are the most vulnerable to diseases like this can be instrumental in the development of effective warning strategies. Knowing the regions that are the most vulnerable can ensure that safety warnings are adequately distributed to these regions and that extra resources to help in the treatment of the disease can also be directed to these areas. Ultimately, using an index like the EVI can be a powerful tool in saving human lives; identifying vulnerable and at-risk areas and populations can help to promote changes in either infrastructure or resources that increase the resilience of these areas and mitigate their risk so that in the future, overall vulnerability will decrease.

CHAPTER 6

CONCLUSIONS

There are several conclusions that can be taken from this project. First, using Z-scores is a much more accurate and effective way to identify extreme values within a dataset. For any dataset, this statistic emphasizes the mathematical relationship between the mean, the standard deviation, and the rest of the data. This means that one can see the amount by which a certain data point deviates from the mean of a dataset, which helps to identify the normality of that data value. In terms of extremity. This serves as a more effective metric to quantify extremes because it describes the deviation from the mean in a clear way. Additionally, calculating the CEI for each climate division of the United States allows for a more local index calculation and creates a more useful tool for the general public. This can be seen in the maps created for December 2015 CEI values. Using these maps allows for the user to see exactly where the extreme values were located, as well as compare those extreme values to the value in their own division to determine whether or not their division was extreme for a given component in a month.

Combining CEI values with values from the SVI created an Extremes Vulnerability Index that calculates overall vulnerability to climate extremes by combining physical exposure and social demographic factors and demonstrates the level to which each index is impacting overall vulnerability. After comparing December 2015 CEI values and 2016 SVI values, two regions of the United States stand out as being the most vulnerable overall: the Southeast United States and portions of the central United States. These regions both had a slightly higher vulnerability value from social factors than from physical exposure, but the bivariate scale indicates that physical exposure is still impacting overall vulnerability in these regions. This information can be used and applied to create and implement policies that help to increase resilience in these climate division and mitigate overall vulnerability.

This research adds to current knowledge in several ways. First, using Z-scores to determine values of temperature and precipitation that are extreme in the calculation of the CEI is an update that leads to a more accurate and efficient index. Representing this index as a numerical scale also makes it easier for users without a background in meteorology or climatology to interpret it. This project also made use of Z-scores to re-calculate values used in the calculation of the SVI. Using Z-scores allows for a better comparison between the values in each climate division, and also allows for the calculation of the SVI on a numerical scale. Finally, this project combined these two indices for the first time to produce an overall Extremes Vulnerability Index to identify how the factors contained in each interact to cause climate divisions to be more or less vulnerable to climate extremes, and this information can be used to identify divisions that are the most vulnerable to climate extremes. Policymakers can then use this information to create policies and infrastructure that mitigate a climate division's vulnerability.

REFERENCES

Actuaries Institute, 2018: Australian Actuaries Climate Index: Design Documentation. Accessed January 2019, https://actuaries.asn.au/microsites/climate-index.

Allison, E.H., and Coauthors, 2009: Vulnerability of national economies to the impact of climate change on fisheries. *Fish and Fisheries*, **10**, 173-196, DOI: 10.1111/j. 1467-2979.2008.00310.x.

American Academy of Actuaries, Canadian Institute of Actuaries, Casualty Actuarial Society and Society of Actuaries, 2016: Actuaries Climate Index: Development and Design. Accessed January 2018, http://actuariesclimateindex.org.

American Academy of Actuaries, Canadian Institute of Actuaries, Casualty Actuarial Society and Society of Actuaries, 2016: Actuaries Climate Index: Executive Summary. Accessed January 2018, http://actuariesclimateindex.org.

Bell, J.E., 2018: Double Threat. *The Actuary Magazine*, **April/May 2019**, accessed June 2018, https://theactuarymagazine.org/double-threat/.

Center for Immigrant Studies, 2014: One in Five U.S. Residents Speaks Foreign Language at Home, Record 61.8 million. Center for Immigrant Studies, accessed March 2019, https://cis.org/One-Five-US-Residents-Speaks-Foreign-Language-Home-Record-618-million.

Centers for Disease Control and Prevention, 2017: Heat and Older Adults. Accessed 25 March 2019, https://www.cdc.gov/disasters/extremeheat/older-adults-heat.html**.**

Centers for Disease Control and Prevention, 2019: Zika Virus. Accessed 25 March 2019, https://www.cdc.gov/zika/reporting/2016-case-counts.html.

Centers for Disease Control and Prevention/Agency for Toxic Substances and Disease Registry/Geospatial Research, Analysis, and Services Program, 2017: Social Vulnerability Index 2016 Database US, accessed February 2019, https://svi.cdc.gov/data-and-tools-download.html.

Christophers, Sir S. R., 1960: *Aedes Aegypti (L.), The Yellow Fever Mosquito*, Cambridge University Press, 739 pp.

Cutter, S.L., B.J. Boruff, and W.L. Shirley, 2003: Social Vulnerability to Environmental Hazards. *Social Science Quarterly*, **84**, 242-261.

Diem, J. E., C. E. Stauber, and R. Rothenberg, 2017: Heat in the southeastern United States: Characteristics, trends, and potential health impact. *PlOS ONE*, **12(5)**, e0177937, doi:10.1371/journal.pone.0177937.

Esri, 2010: USA Major Lakes layer file. Esri_dm, accessed November 2018.

Flanagan, B.E., E.W. Gregory, E.J. Hallisey, J.L. Heitgerd, and B. Lewis, 2011: A Social Vulnerability Index for Disaster Management. *J. Homeland Security and Emergency Management*, **8**, Article 3, DOI: 10.2202/1547-7355.1792.

Gallant, A.J.E., and D.J. Karoly, 2010: A Combined Climate Climate Extremes Index for the Australian Region. *J. Climate*, **23**, 6153-6165, DOI: 10.1175/2010JCLI3791.1

Gleason, K.L., J.H. Lawrimore, D.H. Levinson, and T.R. Karl, 2008: A Revised U.S. Climate Extremes Index. *J. Climate*, **21**, 2124-2137, DOI: 10.1175/2007JCLI1883.1.

Godber, E., and R. Wall, 2014: Livestock and food security: vulnerability to population growth and climate change. *Global Change Biology*, **20**, 3092-3102, doi: 10.1111/geb.12589.

Huang, H. H., J. Kerstein, and C. Wang, unpublished manuscript: Climate Risk and Bank

Loan Contracting.

Karl, T.R., R.W. Knight, D.R. Easterling, and R.G. Quayle, 1996: Indices of Climate Change for the United States. *Bull. Amer. Meteor. Soc.*, **77**, 279-292.

Karl, T.R., R.W. Knight, D.S. Ezell, and F.T. Quinlan, 1986: Probabilities and Precipitation Required to End/Ameliorate Droughts. National Oceanic and Atmospheric Administration/National Climatic Data Center, 307 pp.

Karl, T. R., C. N. Williams Jr., F. T. Quinlan, and T. A. Boden, 1990: United States Historical Climatology Network (HCN) serial temperature and precipitation data. Department of Energy, Oak Ridge National Laboratory Tech. Rep. ORNL/CDIAC-30, NDP-019/R1, 389 pp.

Karl, T. R., C. N. Williams Jr., F. T. Quinlan, and T. A. Boden, 1990: United States Historical Climatology Network (HCN) serial temperature and precipitation data. National Centers for Environmental Information, accessed October 2018 from ftp file.

KC, B., J.M. Shepherd, and C.J. Gaither, 2015: Climate change vulnerability assessment in Georgia. *Applied Geography*, **62**, 62-74.

Kaiser Family Foundation, 2017: Population Distribution by Age. Kaiser Family Foundation, accessed March 2019, https://www.kff.org/other/state-indicator/distribution-by-age/?currentTimeframe=1&sortModel=%7B%22colId%22:%22Children%200-18%22,%22sort%22:%22desc%22%7D.

Larson, R., and B. Farber, 2006: *Elementary Statistics: Picturing the World.* Pearson Prentice Hall, 615 pp.

Lim, B., E. Spanger-Siegfried, I. Burton, E. Malone, and S. Huq, 2004: *Adaptation Policy Frameworks for Climate Change: Developing Strategies, Policies and Measures*, Cambridge University Press.

Lund, R., and L. Seymour, 1999: Assessing Temperature Anomalies for a geographical region: a Control Chart Approach. *Environmetrics*, **10**, 163-177.

Mattingly, K.S., L. Seymour, and P.W. Miller, 2017: Estimates of Extreme Precipitation Frequency Derived from Spatially Dense Rain Gauge Observations: A Case Study of Two Urban Areas in the Colorado Front Range Region. *Annals of the American Association of Geographers*, **107 (6)**, 1499-1518.

National Centers for Environmental Information: Climate Division Shapefiles. Earth Systems Research Laboratory, accessed November 2018.

National Weather Service, St. Louis, Missouri, 2016: Historic Heavy Rainfall and Flooding Event, December 26-28th 2015. Accessed 27 March 2019, https://www.weather.gov/lsx/12_26_2015.

National Weather Service, Tulsa, Oklahoma, 2016: December & Year 2015. Accessed 13 March 2019, https://www.weather.gov/tsa/climo_summary_dec2015.

Nittle, N.K., 2018: Which U.S. States have the Highest Minority Populations. Thought Co, accessed March 2019, https://www.thoughtco.com/states-with-majority-minority-populations-2834515.

NOAA National Centers for Environmental Information, 2016: State of the Climate: National Climate Report for December 2015. Accessed 7 March 2019, https://www.ncdc.noaa.gov/sotc/national/201512.

NOAA National Centers for Environmental Information, 2019: Climate at a Glance:

Divisional Mapping. Published February 2019, Accessed February 25 2019, https://www.ncdc.noaa.gov/cag/

O'Brien, K., S. Eriksen, A. Schjolden, and L. Nygaard, 2004: What's in a Word? Conflicting interpretations of vulnerability in climate change research. Center for International Climate and Environmental Research Working Paper 2004, 19 pp.

Owen, R., 2019: Actuaries are Paying Attention to Climate Data [in "Explaining Extremes of 2017 from a Climate Perspective]. *Bull. Amer. Meteor. Soc.*, **100(1)**, SV-SVIII.

Oxfam America, 2018: Mapping social vulnerability in Southeastern states and the Gulf Coast. Accessed 16 March 2018, https://policypractice.oxfamamerica.org/work/poverty-in-the-us/mapping-social-vulnerability-in-southeastern-states-and-the-gulf-coast/

Pachauri, R.K. and L.A. Meyer, Eds., 2014: IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, Geneva, Switzerland.

Palmer, W.C., 1965: Research Paper No. 45: Meteorological Drought. U.S. Department of Commerce, 58 pp.

Parry, M.L., O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson, Eds., 2007: IPCC, 2007: Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, UK.

Reid, C.E., M.S. O'Neill, C.J. Gronlund, S.J. Brines, D.G. Brown, A.V. Diez-Roux, and

J. Schwartz, 2009: Mapping Community Determinants of Heat Vulnerability. *Environmental Health Perspectives*, **117**, 1730-1736.

Ropelewski, C.F., and M.S. Halpert, 1986: North American Precipitation and Temperature Patterns Associated with the El Niño/Southern Oscillation (ENSO). *Mon. Wea. Rev.*, **114**, 2352-2362.

Shafer, D. and Z. Zhang, 2017: Descriptive Statistics, *Introductory Statistics*, LiebreTexts, https://stats.libretexts.org/Textbook_Maps/Map%3A_Introductory_Statistics_(Shafer_and_Zhang)/02%3A_Descriptive_Statistics/2.5%3A_The_Empirical_Rule_and_Chebyshev's_Theorem

Social Security Administration, 2017: Annual Statistical Report on the Social Security Disability Insurance Program, 2016. Social Security Administration, https://www.ssa.gov/policy/docs/statcomps/di_asr/2016/di_asr16.pdf.

State Master, 2005: Housing Statistics. State Master, accessed March 2019, http://www.statemaster.com/graph/hou_per_of_hou_uni_tha_are_mob_hom-housing-percent-units-mobile-homes.

United States Bureau of Labor Statistics, 2017: Unemployment Rates for States, 2016 Annual Averages. United States Department of Labor, accessed March 2019, https://www.bls.gov/lau/lastrk16.htm.

United States Census Bureau, 2018: Median Household Income by State. Census Bureau, accessed March 2019, https://www.census.gov/search-results.html?q=income+ranked+by+state+2016&page=1&stateGeo=none&searchtype=web&cssp=SERP&_charset_=UTF-8

Vose, R.S., S. Applequist, M. Squires, I. Durre, M. J. Menne, C.N. Williams, Jr., C.

Fenimore, K. Gleason, and D. Arndt, 2014: Improved Historical Temperature and Precipitation Time Series for U.S. Climate Divisions. *J. Appl. Meteor. Climatol*. **53**, 1232-1251, DOI: 10.1175/JAMC-D-13-1248.1.

Vose, R.S., S. Applequist, M. Squires, I. Durre, M.J. Menne, C.N. Williams, Jr., C. Fenimore, K. Gleason, and D. Arndt, 2014: NOAA's Gridded Climate Divisional Dataset (CLIMDIV), version 2. NOAA National Centers for Environmental Information, accessed July and December 2018, doi:10.7289/V5M32STR.

Wilks, D.S., 1989: Rainfall Intensity, the Weibull Distribution, and Estimation of Daily Surface Runoff. *J. Appl. Meteor.*, **28**, 52-58.

Yang, L., L. Mahoney, and S. Hurley, 2017: Northeast US Cross-Cutting: Developing Annual, Seasonal, and Monthly Temperature Indices over the Northeast United States to Represent Recent Temperature Trends using NASA and NOAA Datasets. *NASA DEVELOP National Program Fall 2017 Technical Paper*.

APPENDIX A

THE PALMER DROUGHT SEVERITY INDEX

Due to the complex nature of the PDSI, a brief explanation of the PDSI is contained in this appendix. The original index was released around the year 1965, and an explanation was given in Palmer (1965). This paper describes the PDSI as an "index of drought or wet spell severity". In Palmer (1965), this value is represented by the variable "X". PDSI values are ultimately calculated using the following equation:

$$X_i = X_{i-1} + (z_i/3) - 0.103 * X_{i-1}.$$

Within this equation, X_i represents the PDSI value of the area of interest for the current month, X_{i-1} represents the PDSI value from the previous month, and z is a calculation of moisture anomaly. The final index is a numerical scale ranging from -10 to 10, but the range of values most commonly seen, and the range most referenced by Palmer, is from -4 to 4. This scale is used to represent departures from normal, where 0 is the numerical value that is considered to be normal.

Palmer's definition of normal is the amount of precipitation required for near-normal operation of a particular region. This amount is dependent on the hydrological climate of that region, measured by variables like potential evapotranspiration, as well as the meteorological conditions during and preceding the time period in question. The difference between received and required precipitation represents the departure from the amount needed for normal operation. This difference is the anomaly that is used in the calculation of the PDSI. So, "normal" is when the moisture supply (i.e., the amount of precipitation that an area receives) meets the demand (i.e., the amount of precipitation required for normal operation). This would indicate that a positive value means that an area has more moisture than it requires, and a negative value means that an area has less moisture than it requires.

Another important note is that the final calculation of the PDSI is dependent on the duration and the magnitude of the moisture anomaly. One of the goals of the index is to make time and space comparisons of moisture anomalies to determine the severity of the anomaly. The amount of time that an area experiences a departure from normal moisture availability and the intensity of the anomaly are both important factors considered in the calculation of the PDSI. Because the goal of the PDSI was to be able to compare anomalies from different places and different time periods, the final index itself is dimensionless, meaning that it is represented by a unit-less number.

To get the range of -4 to 4, Palmer arbitrarily fit a line to data points of moisture anomalies for Iowa and Kansas. Since these were the driest time periods, these values were assigned a value of -4 on the scale. He then subdivided the extreme drought (-4) value and 0 into another three regions, hence the -3, -2, and -1 values representing drought. For the purpose of comparison, this method helped to standardize moisture anomaly values. Again, the full extent of the index is from -10 to 10, but range of values most commonly seen is from -4 to 4.

For the purpose of the CEI, the developers needed to analyze these values to determine which were extreme. Karl et al. (1986) examined the frequency of each type of moisture anomaly that occurred, and ultimately determined that values that were greater than 3 and values that were less than -3 corresponded to the 90th and 10th percentiles, respectively. This means that values with an absolute value of 3 or larger only occurred during 5-10% of the period of record. Hence, for many years, these values were used as the means of determining whether or not a climate division was extreme. The updated, and current CEI, does not use this criterion anymore; values exceeding the 90th percentile value and those less than the 10th percentile value are now considered to be extreme values. For a full explanation see Gleason et al. (2008).

APPENDIX B

NEW CEI COMPONENTS CODE

A portion of the code that was written to re-calculate the CEI is contained below. This code is from the script for Component 1 (maximum temperature), and this section of the code calculated the Z-score for each monthly value per climate division. First in this section of code is a series of do loops that create the array that the final Z-scores will be written to. The program loops through each state (represented by i), climate division (represented by j), year (represented by k), and month (represented by l). The program must loop through each of these variables to reserve a place in the rZscore array for each monthly value per climate division. The array is initially filled with -99.99 to represent missing values so that missing values not recorded as 0. After creating the rZscore array, final Z-score values are calculated. The program loops through states, climate divisions, years, and months to calculate the Z-score of each monthly maximum temperature value. The Z-score value is calculated using the array with the monthly maximum temperature value (contained in the rMaxAvg array) per climate division, the 1981-2010 monthly maximum temperature mean value, and the 1981-2010 monthly standard deviation value (both per climate division). Once calculated, the Z-score value is stored in the rZscore array and written to a text file. The file that this value will be written to is contained in the write statement. The unit number 15 is used to represent the name of the text file, and the format of the values in the file is contained in the format statement under the write statement. This is the section from the script used to calculate Component 1 values, but a similar format was used in the code to calculate Z-scores for all components.

```
*******************************************************************
c       Now, do the Z-score thing. Write final values to file.

        print*,'Calculating Z-scores'
        print*,' '

c       Initialising array
        do i=1,48
          do j=1,iDivisions(i)
            do k=1900,2017
              do l=1,12
                rZscore(i,j,k,l)=-99.99
              enddo
            enddo
          enddo
        enddo

        do i=1,48
          do j=1,iDivisions(i)
            do k=1900,2017
              do l=1,12
                rZscore(i,j,k,l) = (rMaxAvg(i,j,k,l) -
rMean30(i,j,l)) /
     +             rstdev(i,j,l)
              enddo

              write(15,555) i,j,k,(rZscore(i,j,k,l),l=1,12)
555           format(i2,1x,i2,1x,i4,1x,12(f5.2,1x))
            enddo
          enddo
        enddo
        print *,rZscore(1,1,1900,1)
        close(15)

        end
```

The following section is discussed in Chapter 4, Section C. In this section changes made to the code for the first two components that allow the calculation of these components on a climate division basis are discussed. Below is a portion of the "regread" subroutine.

```
        do m=1,48
          do j=1,10
            iSeg(m,j)=0
          enddo
        enddo
```

```
      iSeg(25,4)=0

      iKounter(1)=0
      do iState=2,48
        iKounter(iState) = iKounter(iState-1) +
iDivisions(iState-1)
      enddo

      do m=1,48
        do j=1,iDivisions(m)
          if((m.eq.25).and.(j.eq.4)) then
            goto 51

          else
            write(cM,'(i2.2)') m
            write(cJ,'(i2.2)') j
            iSeg(m,j)=iSegments(j + iKounter(m))

            do iSegNumber=1,iSeg(m,j)
              if(iSegNumber.eq.1) then
                cInFile = cM//cJ//'.poly'
                write(*,*) 'Reading: ',cInFile

              else
                write(cE,'(i2.2)') iSegNumber
                cInFile = cM//cJ//cE//'.poly'
                write(*,*) 'Reading: ',cInFile
              endif

            open(unit=13,file=cInFile)
            jCount(m,j,iSegNumber) = 0 !replace with e
88          read(13,*,end=99) rLon,rLat
            jCount(m,j,iSegNumber) = jCount(m,j,iSegNumber) + 1
            xx(m,j,iSegNumber,jCount(m,j,iSegNumber)) = rLon
            yy(m,j,iSegNumber,jCount(m,j,isegNumber)) = rLat
            goto 88
99          close(13)
            enddo
          endif
51      enddo
      enddo
```