Heat as a hazard to human health: A multiple dataset assessment of extreme heat indices relevant to human health.

by

Evan M. Oswald

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Atmospheric Sciences)
in the University of Michigan
2013

Doctoral Committee:

Professor Richard B. Rood, Chair
Professor Frank Marsik
Professor Marie S. O'Neill
Professor Allison Steiner

Dedications

To my wife Emma and kids; to my parents: David and Rose; to my brothers, sister and relatives; to all my friends. Thank you for your invaluable support.

To George Richard “Little Dickey” Oswald, in memoriam.

To my friend Andrew Jureziz for listening to me think aloud about nearly all aspects of this thesis research, as they came to my mind while we watched sports.

Acknowledgements

This PhD work would have been impossible without the support and help provided by my advisory committee.

My special thanks go out to my advisor Dr. Richard "Ricky" Rood for his guidance, support, and mentorship throughout my graduate studies and part of my undergraduate studies; to Dr. Marie S. O'Neill who provided invaluable research opportunities, great insight into the workings of applications community professionals and invaluable personal growth of my interdisciplinary communication skills; to Dr. Sushil Atreya who provided crucial funding during my first year of graduate studies as well as valuable learning in the field of planetary atmospheres; to Drs. Frank Marsik and Allison Steiner for their roles in my advisory committee and providing valuable feedback.

I am also thankful to my research collaborators for the help, feedback and great work they did: Mr. Shannon Brines, Drs. Kai Zhang, Carina Gronlund, Jalonne White-Newsome and Daniel Brown.

I also thank all faculty, staff and fellow students from the Atmospheric, Oceanic and Space Sciences department at the University of Michigan. Especially Kristen Mihalka, Kevin Reed and Ahmed Tawfik for advice; Sandra Hicks and Margaret Reed for administrative support; and Dr. Mary Anne Carroll for first suggesting I pursue a doctoral degree.

Funding for this research was provided by the Graham Environmental Sustainability Institute at the University of Michigan, the U.S. Environmental Protection Agency Science to Achieve Results (STAR) under grant R832752010, the Centers for Disease Control and Prevention under grant R18EH000348 and by NOAA Climate Program Office grant NA10OAR4310213.

Table of contents

List of tables

List of figures

List of acronyms

Abbreviation	Phrase
110DM	110-day mean
AWOS (station)	NWS's Automated Weather Observing Station
ASOS (station)	NWS's Automated Surface Observing Station
CDIAC	Carbon Dioxide Information Analysis Center
CONUS	Continental United States
Co-Op (network)	the national weather service's cooperative climate observing network
CPE(s)	Cumulative percentile exceedences, which are described in section 2.2.3.
CRS (instrument type)	Cotton region shelter with liquid-in-glass thermometers inside
D2CC (variable)	distance to city center, with the city center defined as grand circus park downtown Detroit, MI
D2H2O (variable)	distance to water, with water defined as either Lake St. Clair, the Detroit River and Lake Erie.
DAYMET (dataset)	dataset described in the Thornton et al. (1997) study
Early season EHEs	An EHE or EHE day before July 1^{st}
EHE(s)	extreme heat event(s)
EHE-days	the number of calendar dates considered to be within an EHE
HL (method)	method of creating high resolution gridded datasets that loosely match that of a homogenized dataset (typically the USHCN network)
HOBO (instrument/station)	Pro V2 External Temp/RH data loggers (Onset Computer Corporation 2010) (Onset Computer Corp., Pocasset/MA)
Late season EHEs	An EHE or EHE day after June 30^{th}
LCLU	Land cover/land use
Maurer (dataset)	dataset described in the Maurer et al. (2002)
MDEQ	Michigan Department of Environmental Quality
MI	Michigan (the state of)
MMTS (instrument type)	minimum/maximum temperature sensor
MTR	metropolitan temperature range
NARR (dataset)	North American Regional Reanalysis
NCDC	National Climatic Data Center
NWS	National Weather Service

OK	Ordinary Kriging interpolation method
OLS	Ordinary Least Squares linear regression method
PIS (variable)	percent impervious surface
PRISM/Diluzio (dataset)	the temporally downscaled version of the Daly et al. (2008) PRISM dataset described in the Diluzio et al. (2005) study
RMSE	root mean squared error
SSC/SSC2	Spatial synoptic classification system 1st and 2nd versions. More details at http://sheridan.geog.kent.edu/ssc.html
std	standard deviation
Surfacestations.org	the project by Anthony Watts found at http://surfacestations.org
SR	Spearman Rank
Tmax	the daily maximum temperature
Tmax type EHE	An EHE that meets requirements in Tmax
Tmin	the daily minimum temperature
Tmin type EHE	An EHE that meets requirements in Tmin
Tmnx type EHE	An EHE that meets requirements in both Tmin and Tmax
TOB (non-climatic discontinuity)	Time of Observation bias
UHI	urban heat island
USHCN	United States Historical Climatology Network
USHCNv1-daily dataset	dataset described in the Menne et al. (2012) paper
USHCNv2-monthly dataset	dataset described in the Menne et al. (2009) paper, most recent version of the dataset
USCRN (network)	National Oceanic and Atmospheric Agency's United States Climate Reference Network

Thesis abstract

Oswald, Evan. Ph.D., University of Michigan, May 2013. Heat as a hazard to human health: A multiple dataset assessment of extreme heat indices relevant to human health.

Advisor: Richard B. Rood.

In the United States more people die due to complications of extremely hot near surface air temperatures than any other weather type. This is true despite a systematic underestimation of accrediting heat to death. Global climate change may raise temperatures substantially by mid century but even more so in some regions - particularly in high latitudes. This situation coupled with resource issues caused by urbanization and population increases, generates a need for research in this field to be relevant and informative to the climate field, as well as by scientists within other fields.

Research in this field can be difficult as surface meteorological data can be compromised by non-climatic factors (e.g. built environment surrounding station, observing practice variability), especially in areas were people live. Thus often datasets require corrections of such influences. This, in part, generates datasets with different strengths and limitations, and occasionally the differences between them can also be informative. One of our datasets consisted of station data that corrected non-climatic factors, another was a gridded dataset that did not correct those factors and yet another was constructed and the non-climatic factors were adjusted and/or accounted for. Ultimately, this work investigated trends in extreme heat across the United States, how they were represented in a group of high-resolution gridded datasets, as well as how temperatures varied across an urban region.

Part of this research used a dataset created by combining the United States Historical Climatology Network monthly and daily dataset versions into a dataset both with accurate trends while also being at the scale required to resolve extreme heat events (daily). Trends in the characteristics of extreme heat events (EHEs) were investigated as well as EHEs based on different daily temperatures. Additionally studied was how those trends related to the more predictable, summer average

trends. Subsequently it was desired to quantify those trends on a spatially continuous surface able to resolve the smaller scales (e.g. the Upper Peninsula of Michigan), but a literature review suggested that the current generation of gridded datasets could be of insufficient quality. Thus the aforementioned dataset was used to evaluate the ability of three high-resolution datasets to reproduce trends in extreme heat indices. In order to investigate the climate within an urban region with respect to spatial variability, an observational network across an urban region (Detroit, MI) was constructed by integrating multiple observing networks.

The extreme heat analysis produced maps of linear trends in EHE characteristics as well as trends on the regional and continental scale. It also quantified the relationships with both early season (before June 30) trends in EHEs and summer average temperature trends. It was found that EHEs have increased in the recent history, however this was partially offset by decrease during 1930-1970 period. Notably the center of the United States had a decrease/lack of increase. This study also indicated trends were sensitive to which daily temperature extreme the definition was based on or what region was being described, and showed differences between trends in both EHEs and summer average temperatures and between early and late summer EHE trends.

The evaluation of high-resolution gridded climate observational datasets indicated that non-climatic biases (such as urbanization, instrument changes, station relocations) caused temporal discontinuities that led to statistically significant differences between the datasets. The differences between the datasets changed through time and varied in space. The differences existed at interannual and long-term (e.g. trends) time scales as well as at the continental, regional and small spatial scales. The bulk of the errors operate at the small scale. The differences were loosely empirically linked to proxies for non-climatic biases at those stations. The results also indicated that in gridded datasets, the errors caused by non-climatic biases in time series extend away from the data locations and integrated with errors from other data locations to form time series with intricate differences from a homogenized dataset.

The research that focused on the urban spatial variability indicated that spatial variability was more pronounced during the daily minimum than daily maximum temperatures. That is important to the heat-health discussion because the spatial variability in temperature exists during hot weather. Also it was demonstrated that the spatial variability could be diagnosed, and possibly forecasted, from the city-average wind and sky conditions. Finally, a method for diagnosing the spatial pattern of temperatures was demonstrated in Detroit. Informing heat health decision makers on the tools to protect people from hazardous heat was important as was demonstrating a framework for observationally studying urban regions.

Several things learned from these studies have implications moving forward. Trends in EHEs were unexpectedly sensitive to which daily temperature extreme met the requirements of the EHE definition. This means future studies should explicitly quantify EHEs based on different daily temperature extremes and both daily temperature extremes. It was learned that trends in EHEs during the earlier portion of the summer can differ from those of the later portion of the summer. A further investigation is needed into which mechanisms enable independent changes of the earlier portion from the later portion of the summer, and how those mechanisms will change in the future. Lastly, establishing that the existence of differences between the trends in summer mean daily temperature extremes and the corresponding EHEs, implies caution should be used when assuming summer temperature trends equate to EHE trends. The next question addressed should be regarding the role of the mean temperature changes in the changes in EHEs (as opposed to the variability).

On the large scale, the high-resolution gridded datasets had statistically significant differences in both trends and temporal averages. This was significant because previous conclusions were that the issues caused by non-climatic biases would cancel out at large scales. The differences however were much larger at the small scales, which is important as it means uncertainty likely exists in statistical downscaling products. Conclusions that the non-climatic biases that occurred at the stations were not as strongly related as expected to the differences between the datasets at those locations implied that the gridding process effectively merges the effects of non-climatic biases across space, creating time series at all locations affected by many biases. Ultimately the next generation of this type of climate dataset needs to homogenize the underlying network (prior to the creation of grids) without removing so many stations as to make high-resolution grids impossible.

Observational evidence that the previous afternoon's radiative characteristics were a useful predictor of the following overnight spatial variability in temperatures is new to the field and should be followed up by a more concrete analysis of that relationship. Findings that reanalysis data could force a statistical model of the spatial variability across the region suggests the next work should lead to building the relationship with forecast models so the amount of spatial variability during EHEs can be forecasted. This study represents a framework to observationally study an urban climate. This framework might be applied to other cities in a similar fashion or could be expanded with the goal of converting the "weather underground" observing network into a tool that could provide quality quantitative studies of the urban climate. Lastly, the confirmation of the spatial variability specifically during hot weather implies that spatial variability in temperatures should be considered during EHEs when assessing vulnerability

CHAPTER 1. INTRODUCTION

1.1. Motivation

Extremely high temperatures contribute to death (Baccini 2008; Basu 2009) and between 2001 and 2010 the average United States fatalities for heat was behind hurricanes by only one fatality per year (NWS 2011). This decade included the extraordinary toll of Hurricane Katrina, and looking over several decades, extreme heat caused more deaths than any other weather-related event. That is impressive considering it is generally agreed that heat as a cause of death is systematically underestimated (Donoghue et al. 1997). The soon to be released fifth climate assessment by the Intergovernmental Panel on Climate Change will show that surface air temperature changes by mid-century in certain regions are predicted to be over 2 °C. Furthermore, the number of heat-related deaths will likely increase due to resource limitations in response to population growth (Kapitza 2006) and urbanization (United Nations 2012). Thus understanding how extreme temperatures vary in time and space and which datasets properly describe that variability is, and will remain, important in the future. This thesis explores questions regarding the near surface air temperatures with a focus on extreme heat.

Currently, there exist many questions yet unanswered concerning extreme heat. Exploring such questions builds knowledge of the physical mechanisms of the climate at large and small scales, which is a chief objective of atmospheric science. However beyond exploring the physical system, this research also strives to reduce the gap between the provision of information about the physical climate and the usability of that data by those charged with making decisions regarding weather, climate and climate change.

The question of whether EHEs have changed in the past is not a new one, but neither has it been fully explored. For instance most EHE trend analyses focusing on the CONUS start their analysis in the middle or end of a period of temperature decrease (temperature decreased from about 1935 to 1975 in the CONUS (Menne et al. 2009; Hansen et al. 2010)) instead of capturing the entire period of cooling (i.e.

starting at or slightly before 1935). Trends in EHEs have never been quantified using the latest dataset version of the United States Historical Climatology Network. How the trends in EHEs relate to the trends in the mean summer temperature trends has not been explicitly examined to our knowledge. Furthermore, EHE trend analyses have never been undertaken with a focus to transfer knowledge to the heat-health discussion (discussed in section 2.1).

The differentiations between high-resolution gridded observational climate datasets and more traditional climate datasets that lacks published evaluation. These datasets originate out of the need for high-resolution datasets by various modeling sectors outside of the atmospheric science sector, for example hydrology, and have not been robustly evaluated for their application in extended climate studies. An older study touched on the subject and concluded that issues averaged out on larger temporal and spatial scales (Moberg and Alexandersson 1997). Questions need to be addressed regarding which spatial and temporal scale(s) the differences between these gridded datasets and more trusted climate datasets exist at. Questions need to be tackled regarding the role of inhomogeneities in the underlying ungridded data and what effect the gridding process had. This information in turn can guide the changes made to the next generation of these datasets. Also the answers to these questions contribute to the informed use of these datasets for evaluations and of downscaled global climate model products produced with these datasets help. Often scientists from other communities (public health, forestry, fire risk, biogeochemical modeling) use these data products as guidance for decision-making. This evaluation was tailored to the heat-health discussion.

Improving knowledge in the urban climate field is also beneficial for both climatologists and scientists in the heat-health discussion. While the differences in temperatures between outside the city and inside the city have been examined, the temperatures differences throughout developed regions were much less examined. Most previous studies focus on the time of day where the temperatures are more influenced by the built environment, in other words the time with the largest temperature difference (or "urban heat island effect"). However since the heat-health discussion regularly focuses on daily minimum and maximum temperatures, examining the variability in those daily temperature extremes was how this study proceeded. Additionally, examination of the spatial variability in temperatures during hot or stressful weather was a previously underexplored are that was investigated.

1.2. Scientific questions

This thesis set out to answer a handful of scientific questions that would be helpful to the respective fields. These questions were approached in a manner that would also be helpful to the heat-health discussion. These questions are succinctly provided below.

In regards to EHEs, were there linear trends in the characteristics of EHEs since the 1930’s (the last "hot decade")? Did the trends generally follow the air temperatures decreasing pattern until the mid 1970s and increasing pattern through 2010? Furthermore, did those trends depend on which daily temperature extreme the EHEs were based on, or the portion of the summer (e.g. timing within the summer)? Were the trends the same as the trends in the summer mean temperatures?

In regards to the gridded datasets, were there differences in linear trends in extreme heat climate indices? Were there differences in the temporal average temperatures? Did the differences in linear trends and temporal averages exist at the continental and regional spatial scales as well as the smaller scales? Did the trends relate well to the non-climatic biases that occurred at the observing stations?

In regards to the urban climate of the Detroit, MI metropolitan region, the first question was whether there existed spatial variability in both daily temperature extremes. What were the differences, in that variability, between the daily minimum and maximum temperatures? Did that variability exist during dangerous weather? Was the amount of variability controlled by the larger-scale weather conditions? Were the locations of hotter and cooler temperatures related to the land attributes? Did those locations and/or relationships with land attributes, change during weather favoring above average amounts of spatial variability (as opposed to during all weather types)?

1.3. Datasets

The longest standing meteorological observing network in the continental United States (CONUS) is called the Co-Operative Observer Network Program (Co-Op) (McCarthy 2007); this network has roughly 7600 stations reporting daily maximum and minimum temperatures for at least 10 of the network's 120-plus year history. This network observes daily minimum and maximum temperatures as well as precipitation. Volunteers operate the network and subsequently the locations are typically located in areas accessible from where they live (i.e. near where people live). Most CONUS climate datasets are based on observations made by the Co-Op network.

The gridded Maurer dataset has a spatial resolution of roughly 12 km, daily temporal resolution and is serially complete over the 1949-2010 period. The spatial domain is the CONUS and parts of Mexico and Canada. Its widespread use has risen due to the dataset being used to downscale global climate model output (e.g. Maurer et al. 2010; Hayhoe et al. 2010) through the bias-correction and spatial disaggregation downscaling technique (Wood et al. 2004). It has also been used to assess climate signals, for example summer nighttime temperature trends in the California's Central Valley (Bonfils et al. 2007), trends in annual maximum temperatures in Florida (Waylen et al. 2012) and the U.S. spatio-temporal patterns in surface temperature caused by the El-Niño/Southern Oscillation (Zhang et al. 2012).

The DAYMET dataset spans only 1980-2008 but is at a higher spatial resolution (1 km). The domain of publicly available data is the United States, Mexico and Canada; however data in other parts of the globe have been made available to other studies. The temperature values in the DAYMET dataset are widely used in several application communities such as those modeling past and current fire hazard and risk (Keane et al. 2010), modeling productivity of forests (Turner et al. 2011; Littell et al. 2010), modeling biogeochemical cycling rates (Hartman et al. 2011; Pan et al. 2009), mapping past and future corn pest risk (Diffenbaugh et al. 2008) and modeling the transmission risk of human diseases (Konrad et al. 2011; Wimberly et al. 2008).

The final dataset evaluated is a daily version of the PRISM dataset, which has a 4 km resolution from 1895-2010 over the CONUS. Climate scientists did participate in the design of the PRISM dataset. The PRISM dataset has relatively sophisticated consideration of geographical features and quality control (Daly et al. 2008). For example, it takes distance to coast into consideration when creating its grids for temperature. The PRISM dataset is frequently used for its precipitation fields but studies also use it for its temperatures fields such as the projected impact the 21st century climate changes will have on tree growth (Williams et al. 2010), river basin crop yields (Srinivasan et al. 2010) and riverine nitrogen flows (Schaefer et al. 2009). The temporally downscaled version of the PRISM dataset described in a study by Di Luzio et al. (2005) spans 1960-2001 over the CONUS.

Within the climate community a number of datasets have been designed for determining trends, oscillations and the behavior of temperature at multiple spatial and temporal scales. Popular datasets include the National Climatic Data Center's (NCDC) Global Historical Climatology Network (GHCN)/United States Historical Climatology Network (USHCN), (Lawrimore et al. 2011; Menne et al. 2009), the Hadley Centre and University of East Angila's Climate Research Unit (CRU) datasets (Mitchell and Jones 2005), the National Aeronautics and Space Administration's Goddard Institute for Space Studies (NASA-GISS) dataset (Hansen et al. 2010). These datasets have monthly resolution and are either ungridded (GHCN, USHCN) or low resolution (CRU and GISS are 0.5° and 2.0°, respectively). These datasets generally agree with one another (Rohde et al. 2013; Hansen et al. 2010) and in some regards are thought of as "standard" or "trustworthy" climate datasets.

The United States Historical Climatology Network (USHCN) is a high-quality dataset of daily and monthly meteorological observations from 1218 observing stations across the CONUS. These stations were selected due to a balance of geographic distribution with other factors such as the length of temperature records, amount of missing data, and stability of a station's location and other measurement conditions (GAO 2011). Daily data include observations of maximum and minimum temperature, precipitation amount, snowfall amount, and snow depth; monthly data consist of monthly-averaged maximum, minimum, and mean temperature and total monthly precipitation. The homogenization process for the version 2.0 monthly dataset (Menne and Williams 2009) has been defended as robust (Menne and Williams 2005, 2009; Hausfather et al. 2013), and will be discussed in section 1.4.1.

The automated airport weather stations are a network of "sensor suites" devised to assist in effective aviation operations, climatology and weather forecasting. The observations are at the hourly (sometimes sub-hourly) resolution, and include several meteorological variables. These stations include the Automated Weather Observing System (AWOS), the Automated Weather Sensor System (AWSS) and the Automated Surface Observing System (ASOS). These are operated and maintained by the Federal Aviation Administration, National Weather Service (NWS) and Department of Defense. While there are differences between these networks, the way the observations (of temperature, cloud cover, wind speed) were used in this study, discussion is not essential.

The North American Regional Reanalysis dataset (Mesinger et al. 2006) is a high-resolution climate dataset blending observations and the NCEP Eta Model (a 3-D numerical model used to forecast weather). The data covers 1979-2003 and provides 3-hourly data at a 32km resolution through 29 atmospheric levels, 5 soil layers and the surface and near surface (2 meters above ground level). Reanalysis datasets have full the suite of atmosphere of information, while also being in overall agreement with the meteorological observations. The differences between reanalysis and observational datasets reflect the sensitivity of surface temperature trends to land use land cover change (Fall et al. 2010), but is outside the scoop of this work.

1.4. Working definitions and concepts

1.4.1. Homogenization

The concept of homogenization is a reoccurring one throughout this research. Homogenization is the process of making observations uniform through time and across numerous stations throughout an observing network. In climate datasets this is done through the removing of erroneous temporal discontinuities (e.g. jumps, drifts) in time series; often referred to as non-climatic biases. Non-climatic biases come in 5 general forms: instrument changes, urbanization, time of observation change ("time of observation bias" or TOB), micro-climate changes (e.g. a change within the relevant vicinity) and station relocations. These causes of non-climatic temporal discontinuities in climate records are known and there currently exist methods of fixing them at the monthly, seasonal and annual levels. Daily homogenization methods were recently developed but have their own limitations and to date have not produced datasets for the CONUS (Mestre et al. 2011). Another way to look at homogenization is the adjustment of different station's observations so that they agree with other station's observations. For example, correcting for elevation, instrument differences or sampling methods between different stations or networks (e.g. the observing network in chapter 4).

The "pairwise comparison method" is a popular method of homogenization, and fully described elsewhere (Menne and Williams Jr. 2009). This method uses the time series of differences between a station and its highly correlated neighbors to detect the time and magnitude of discontinuities. This method was used to homogenize the USHCNv2.0-monthly dataset (Menne et al. 2009) and has been defended as robust and objective at the monthly level (Menne and Williams 2005, 2009; Hausfather et al. 2013).

1.4.2. Definition of EHEs

There are numerous ways to quantify extreme heat and EHEs. One reason for the lack of consensus is the spatial variability of the temperature-mortality relationship: it varies by geographical region (Kalkstein 1989). Another is that EHEs impact many systems and thus definitions tend to be constructed with a certain impact in mind (e.g. human health, livestock health, wildfire, agriculture, electricity). However the epidemiological literature (i.e. tailored to the human health impact) does suggest attributes closer linked to mortality such as duration, timing within the summer and the use of percentiles. This study used a definition of two consecutive dates exceeding a threshold percentile to start any EHE, and when the EHE-mean percentile dips below the threshold percentile the EHE ended. The threshold in chapter 2 was 92.5 and the threshold in chapter 3 was 90.0. We chose to lower the threshold for the chapter 3 analysis because having a lower threshold allowed for less instances of a zero value, and thus more confidence in the ordinary least squares method of linear trend estimation. This increased confidence was more important when analyzing another dataset.

This EHE definition was applied in three different ways to produce three distinct types of EHEs. This was done in order to represent the differences between EHE definitions based on different daily temperatures. Periods where the requirements of the EHE definition were satisfied based only on the daily maximum temperatures were diagnosed "Tmax EHEs", periods where the requirements of the EHE definition were satisfied based only on the daily minimum temperatures were diagnosed "Tmin EHEs", and periods where the requirements of the EHE definition were satisfied based only on the daily maximum temperatures were diagnosed "Tmnx EHEs".

1.4.3. Mathematical techniques

There were a number of mathematical tools used repeatedly throughout this thesis, and they are described here. The first was a resampling method called *bootstrapping*. Bootstrapping is the practice of approximating properties of an estimator (e.g. median, standard deviation, correlation coefficient, regression coefficient) by measuring those properties when sampling from an approximating distribution. This is accomplished by constructing a number of resamples of the observed sample (and of equal size to the sample), each of which is obtained by random sampling with replacement from the original sample. Then the estimator is calculated with each sample, and subsequently the median value of the sample of estimators (of equal size to the number of resamples) is considered a stronger approximation of that estimator. Random resampling means that the values obtained each time vary a little. However, throughout the analysis the number of resamples was adjusted to create estimates that varied at a more fine precision than the numbers we wanted to present.

The next tool was that of *percentiles*. Percentiles were used instead of conventional temperatures because the way they were calculated allowed them to be specific to the location of observations and account for the typical changes of temperature throughout the summer season. This is akin to adapting to human heat tolerance (i.e. people in Florida can tolerate hotter temperatures than in Michigan, and people tolerate hotter temperatures in August than in June). All percentiles were calculated without assuming a statistical distribution. In other words, all values from the climate base period were used to calculate the empirical cumulative distribution function (Kaplan and Meier, 1958), which assigned a percentile (y_o) to each temperature value (x_o) in the climate base period sample. Subsequently, bi-linear interpolation was used to find the value (y_i) of the aforementioned percentile function (y_o) at the target temperature value point (x_i) in the climate base period temperature function (x_o). If the target temperature value (x_i) was larger (smaller) than any value in the climate base period sample (x_o) then it (y_i) was assigned a 1.0 (0.0).

Throughout the analysis a statistical hypothesis test, the *Student's t test*, was used to test the mean or mean(s). Student's t tests can be used with one sample or two samples. Within this work a single sample was tested for its mean being equal (or greater, lesser than) to the zero value. Two samples were tested for their means being equal to one another. A Student's t test rejects or accepts the hypothesis and also provides a P-value for interpretation of the statistical significance of the decision.

Lastly, there were choices to be made throughout the analysis between parametric and non-parametric types of statistical methods. In all cases tests of both types were used and the sensitivity examined (often not shown). It was found that the amount of difference usually indicated differences from normality of the statistical distribution. However, often the statistical distributions were too small for normality tests to determine if the samples were well described by a normal distribution. Moreover, in time series it is generally unclear what missing values do to non-parametric statistical tests, and overall parametric tests are more common and thus facilitate comparisons better. Thus typically parametric tests were used in this work, unless there was disagreement or no need to strive for comparability.

1.5. Organization

This thesis was presented in three chapters (in addition to an introduction and a conclusions chapter), each of them corresponding to a full-length journal-article. However, each chapter was slightly extended from those articles as they had content limitations, were tailored to the journal scope and goals of the funding agencies. The citations of the journal articles will be provided at the beginning of each chapter.

Chapter 2 investigates the trends of EHE characteristics within the USHCN version 2 monthly (USHCNv2-monthly) dataset. This chapter corresponds to a paper recently accepted (with revisions) and resubmitted to the Journal of Applied Meteorology and Climatology. Chapter 3 deals with the agreement between three high-resolution gridded observational climate datasets and a climate community standard dataset. This study is near submission to the International Journal of Climatology. Chapter 4 evaluates the spatial variability in daily temperature extremes across the Detroit, MI region. This study has been published in the Journal of Applied Meteorology and Climatology.

1.6. Commonality between chapters

The work in chapter 2 evaluated the CONUS linear trends in several characteristics of EHEs. Two products of the USHCN network were combined to provide a daily dataset more suitable for trend analysis. Its suitability was due to homogenization at the monthly level. A series of scientific questions were addressed relating to EHEs and their relationships to other variables.

The primary connection between the research in chapter 2 and chapter 3 was their focus on CONUS trends in extreme heat climate indices in publicly available and widely used climate datasets. The primary incentive for the evaluation of the datasets in chapter 3 was confirmation that the chapter 2 trends within highly resolved gridded datasets agreed with a more trusted climate dataset. If yes, then more robust trend analyses should be performed (e.g. trends in the spatial area of EHEs) using those gridded datasets; if not, a characterization of the disagreements might allow insight into how/if they can be appropriately used, or how to correct them. The reason for the disconnect with the more trusted climate datasets was a lack of homogenization in the underlying network of the gridded datasets.

The research in chapter 4 was similar to the previous two chapters because it also focused on daily temperature extremes and extreme heat. The work in this chapter was an extension of the work in chapter 2 and 3; while chapter 2 examined extreme heat climate risk on larger scales (decadal trends, regional variability), this chapter examines that risk on smaller scales (days, intra-city variability). The primary allure of the datasets evaluated in chapter 3 was the spatial resolution that allows extreme heat risk to be studied on smaller scales, but the analysis of chapter 4 was at even smaller scales than the datasets used in chapter 3. This network was based on integrating three observing networks in the urban environment, which required a type of homogenization since there were biases using observations from different networks and from monitoring in an urban environment. Lastly, the chapter 3 datasets fall short of being excellent tools for such a purpose, in part, because they were influenced by urbanization, which was a primary driver of the small-scale spatial variability in chapter 4.

1.7. References

Baccini, M., A. Biggeri, G. Accetta, T. Kosatsky, K. Katsouyanni, A. Analitis, H. R. Anderson, L. Bisanti, D. D'Ippoliti, J. Danova, B. Forsberg, S. Medina, A. Paldy, D. Rabczenko, C. Schindler, P. Michelozzi, 2008: Heat effects on mortality in 15 European cities. *Environmental Health*, **9**, 37, doi: 10.1186/1476-069X-9-37.

Basu, R. 2009: High ambient temperature and mortality: a review of epidemiologic studies from 2001 to 2008. *Environmental Health,* **8**:40, doi:10.1186/1476-069X-8-40.

Bonfils, C., P. B. Duffy, B. D. Santer, T. M. L. Wigley, D. P. Lobell, T. J. Phillips, C. Doutriaux, 2008: Identification of external influences on temperatures in California. *Climatic Change* **87**: S43-S55. doi: 10.1007/s10584-007-9374-9.

Daly, C., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, P. P. Paseris, 2008: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, **28:** 2031-2064.

Diffenbaugh, N. S., C. H. Krupke, M. A. White, C. E. Alexander, 2008: Global warming presents new challenges for maize pest management. *Environmental Research Letters* **3**: 044007. doi: 10.1088/1748-9326/3/4/044007.

Di Luzio, M., G. L. Johnson, C. Daly, J. K. Eischeid, J. G. Arnold, 2008: Constructing retrospective gridded daily precipitation and temperature datasets fo the conterminous United States. *Journal of Applied Meteorology and Climatology* **47**: 475-497. doi: 10.1175/2007JAMC1356.1.

Donoghue, E. R., M. A. Graham, J. M. Jentzen, B. D. Lifschultz, J. L. Luke, H. G. Mirchandani, 1997: Criteria for the diagnosis of heat-related deaths: National Association of Medical Examiners: position paper. *The American Journal of Forensic Medicine and Pathology*, **18**, 11-14.

Fall, S., D. Niyogi, A. Gluhovsky, R. A. Pielke Sr., E. Kalnay, G. Rochon, 2010: Impacts of land use land cover on temperature trends over the continental United States: assessment using the North American Regional Reanalysis. *International Journal of Climatology*, **30**: 1980-1993. doi: 10.1002/joc.1996.

Hansen, J., R. Ruedy, M. Sato, K. Lo, 2010: Global surface temperature change, *Rev. Geophys.,* **48**: RG4004, doi: 10.1029/2010RG000345.

Hartman, M. D., E. R. Merchant, J. W. Parton, M. P. Guttmann, S. M. Lutz, S. A. Williams, 2011: Impact of historical land-use changes on greenhouse gas exchange in the U.S. Great Plains, 1883-2003. *Ecological Application* **21**: 1105-1119.

Hamlet, A. F., D. P. Lettenmaier, 2005: Production of temporally consistent gridded precipitation and temperature datasets for the continental United States. *Journal of Hydrometeorology* **6**: 330-336. doi: 10.1175/JHM420.1.

Hausfather, Z., M. J. Menne, C. N. Williams, T. Masters, R. Broberg, D. Jones, 2013: Quantifying the effect of urbanization on U.S. Historical Climatology Network temperature records. *Journal of Geophysical Research: Atmospheres*, **118:**

481-494, doi: 10.1029/2012JD018509.

Hayhoe, K., J. VanDorn, I. I. T. Croley, N. Schlegal, D. Wuebbles, 2010: Regional climate change projections for Chicago and the U.S. Great Lakes. *Journal of Great Lakes Research* **36**: 7-21. doi: 10.1016/j.jglr.2010.03.012.

Kapitza, S. P., 2006: Global population blow-up and after. *Report to the global marshall plan initiative.* Global Marshall Plant Initiative Eds.

Keane, R. E., S. A. Drury, E. C. Karau, P. F. Hessburg, K. M. Reynolds, 2010: A method for mapping fire hazard and risk across multiple scales and it's application in fire management. *Ecological Modeling* **221**: 2-18. doi: 10.1016/j.ecolmodel.2008.10.022.

Konrad, S. K., S. N. Miller, W. K. Reeves, 2011: A spatially explicit degree-day model of Rift Valley fever transmission risk in the continental United States. *Geojournal* **76**: 257-266. doi: 10.1007/s10708-010-9338-x.

Lawrimore, J. H., M. J. Menne, B. E. Gleason, C. N. Williams Jr., D. B. Wuertz, R. S. Vose, J. Rennie, 2011: An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3. *Journal of Geophysical Research* **116**: D19121, doi: 10.1029/2011JD016187.

Littell, J. S., E. E. Oneil, D. McKenzie, J. A. Hicke, J. A. Lutz, R. A. Norheim, M. M. Elsner, 2010: Forest ecosystems, disturbance, and climatic change in Washington State, U.S.A. *Climatic Change* **102**: 129-158. doi: 10.1007/s10584-010-9858-x.

Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, 2002: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *Journal of Climate* **15**: 3237-3251. doi: 10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2.

McCarthy, D., 2007: Cooperative station observations. National weather service manual 10-1315, pp. 122. (Available at http://www.docstoc.com/docs/977267/Cooperative-Station-Observations)

Menne, M. J., C. N. Williams, R. S. Vose, 2009: The United States Historical Climatology Network Monly Temperature Data – Version 2. *Bullentin of the American Meteorology Society*, **90**, 993-1007, doi:10.1175/2008BAMS2613.1.

Menne, M. J. and C. N. Williams Jr., 2009: Homogenization of temperature series via pairwise comparisons. *Journal of climate* **22**: 1700-1717. doi: 10.1175/2008JCLI2263.1.

Menne, M. J. and C. N. Williams, J., 2005: Detection of undocumented changepoints using multiple test statistics and composite reference series. *Journal of Climate*, **18**: 4271-4286.

Mesinger, F., G. DiMego, E. Kalnay, K. Mitchell, P. C. Shafran, W. Ebisuzaki, D. J and J. Woolen, E. Rogers, E. H. Berbery, M. B. Ek, Y. Fan, R. Grumbine, W. Higgins, H. Li, Y. Lin, G. Manikin, D. Parrish, W. Shi, 2006: North American Regional Reanalysis. *Bulletin of the American Meteorological Society*, **87**: 343–360, doi: 10.1175/BAMS-87-3-343

Mestre, O., C. Gruber, C. Prieur, H. Caussinus, S. Jourdain, 2011: SPLIDHOM: A method for homogenization of daily temperature observations. *Journal of Applied Meteorology and Climatology*, **50**: 2343-2358.

Mitchell, T. D. and P. D Jones, 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology* **25**: 693-712. doi: 10.1002/joc.1181.

Moberg, A., H. Alexandersson, 1997: Homogenization of Swedish temperature data. Part II: Homogenized gridded air temperature compared with a subset of global gridded air temperature since 1861. *International Journal of Climatology* **17**: 35-54. doi: 10.1002/(SICI)1097-0088(199701)17:1<35::AID-JOC104>3.0.CO;2-F.

NWS, 2011: Office of Climate, Weather and Water Services: Natural Hazard Statistics. [Available online at http://www.weather.gov/os/hazstats.shtml]

Pan, Y., R. Birdsey, J. Hom, K. McCullough, 2009: Separating effects of changes in atmospheric composition, climate and land-use on carbon sequestration of U.S. Mid-Atlantic temperate forests. *Forest Ecology and Management* **259**: 151-164. doi: 10.1016/j.foreco.2009.09.049.

Rohde, R., R. Muller, R. Jacobsen, S. Perimutter, A. Rosenfeld, J. Wurtele, J. Curry, C. Wickham, S. Mosher, 2013: Berkeley Earth temperature averaging process. *Geoinformatics and Geostatistics: An overview* **1**:1. doi: 10.4172/gigs.1000103.

Schaefer, S. C., J. T. Hollibaugh, M. Alber, 2009: Watershed nitrogen input and riverine export on the west coast of the U.S. *Biogeochemistry* **93**, 219-233. doi: 10.1007/s10533-009-9299-7.

Srinivasan, R., X. Zhang, J. Arnold, 2010: SWAT ungauged: hydrological budget and crop yield predictions in the upper Mississippi river basin. *Transactions of the ASABE*, **35**, 1533-1546. (Available at http://ddr.nal.usda.gov/handle/10113/46714)

Turner, J., S. R. Colwell, G. J. Marshall, T. A. Lachlan-Cope, A. M Carleton, P. D. Jones, V. Lagun, P. A. Reid, S. Iagovkina, 2004: The SCAR READER project: toward a high-quality database of mean Antarctic meteorological observations. *Journal of Climate* **17**: 2890-2898. doi: 10.1175/1520-0442(2004)017<2890:TSRPTA>2.0.CO;2.

United Nations, 2012: World Urbanization Prospects: The 2011 Revision, Department of Economic and Social Affairs, Population Division, (Available at http://esa.un.org/unpd/wup/index.htm)

United States Government Accountability Office, 2011: Climate monitoring: NOAA can improve management of the U.S. Historical Climatology Network. GAO-11-800, 42 pages.

Waylen, P., D. Keellings, Y. Qiu, 2012: Climate and health in Florida: Changes in risks of annual maximum temperatures in the second half of the twentieth century. *Applied Geography* **33**: 73-81. doi: 10.1016/j.apgeog.2011.06.007.

Williams, A. P., J. Michaelsen, S. W. Leavitt, 2010: Using tree rings to predict the response of tree growth to climate change in the continental United States during the twenty-first century. *Earth Interactions* **14**, 1-20. doi: 10.1175/2010EI362.1.

Wimberly, M. C., A. D. Baer, M. J. Yabsley. 2008: Enhanced spatial models for predicting the geographic distributions of tick-borne pathogens. *International Journal of Health Geographics* **7**: 15, doi: 10.1186/1476-072X-7-15.

Wood, A. W., L. R. Leung, V. Sridhar, D. P. Lettenmaier, 2004: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change* **62**: 189-216. doi: 10.1023/B:CLIM.0000013685.99609.9e.

Zhang, Y., Y. Qian, V. Dulière, E. P. Salathé Jr., L. R. Leung, 2012: ENSO anomalies over the Western United States: present and future patterns in regional climate simulation. *Climatic Change* **110**: 315-346. doi: 10.1007/s1058-011-0088-7.

CHAPTER 2. A TREND ANALYSIS OF EXTREME HEAT EVENTS IN THE CONTINENTAL U.S. SINCE 1930

Full citation of corresponding manuscript

Oswald, E. M. and R. B. Rood: A trend analysis of extreme heat events in the continental U.S. since 1930. *Resubmitted post revisions, Journal of Applied Meteorology and Climatology March 1st 2013,* manuscript number JAMC-D-13-071.

Abstract

Extreme heat events (EHEs) are linked to mortality rates, making them an important research subject in both the climate and public health fields. This study evaluated linear trends in EHEs using the United States Historical Climatology Network (USHCN) version 2.0 dataset and quantified the longer-term EHE trends across the continental United States (CONUS).

The USHCN-daily version 1 dataset was integrated with the homogenized USHCN-monthly version 2.0 dataset to create daily data more appropriate for trend analysis. Time series and estimated trends in multiple characteristics of EHEs (e.g. mean duration, intensity) were calculated as were the continental means and maps. In order to focus on warming and cooling periods the trends were also estimated separately over the first half and second half of the study period (1930-2010). The differences between EHEs based on daily maximum temperatures, minimum temperatures and both minimum and maximum temperatures were explored. Furthermore the differences between EHE trends and trends in the summer mean temperatures, and between EHE trends in the early and mid-late summer were both explored.

The results indicated the trends for different EHE characteristics were coherent (e.g. temporally correlated, similar spatial pattern of trends). Conversely trends for EHEs based on different daily temperature extremes were much less coherent. Continental scale increases between 1970 and 2010 were mostly offset by the decreases between 1930 and 1970. Several daily maximum (minimum) EHEs near the 1930's (2000's) led to 1930-2010 trends of daily maximum (minimum) EHEs decreasing (increasing). Maps indicated negative trends in the interior of the CONUS and positive trends in coastal and southern areas. Summer mean temperature trends and trends in EHEs were generally similar (moreso between daily maximum and daily maximum based EHEs) but regions of discrepancies existed. Lastly, trends in early and late summer were dissimilar during the 1930-1970 period, but not the 1970-2010 period.

Follow up work should focus on the differences between early and late season EHE trends, and what roles trends in the mean and synoptic variability play in EHE trends. Lastly, regional, instead of continental-scale, analyses should be used as guidance due to the amount of spatial variability.

2.1. Introduction

Knowledge of past extreme heat event (EHE) trends assists in the prediction of future EHEs trends. This is particularly true in light of trends in more certain variables such as seasonal average temperatures. Recent studies have quantified past continental United States (CONUS) annual daily mean near-surface air temperatures (Hansen et al. 2010) and annual daily minimum and maximum temperatures (Menne et al. 2009; Shen et al. 2011). Such studies are valuable to both the climate and applications communities but similar studies tailored to be more inline with the heat-health discussion could be further useful in protecting lives. This study attempts to fill gaps between the epidemiological literature (linking mortality and weather) and the climate community (exploring extreme temperatures over time), while still producing new knowledge about EHEs.

In the epidemiological literature, features of hot weather linked to mortality rates have been identified. Trend analyses of EHE characteristics related with mortality generate knowledge more usable to scientists within the heat-health field. For example, the epidemiological literature has established the importance of duration in elevated temperatures (e.g. Kalkstein 1991; Díaz et al. 2002; Hajat et al. 2006; Anderson and Bell 2009; Ostro et al. 2009). Also important is the event's sum of cumulative degree-days over a heat stress-relevant threshold (i.e. intensity) (Díaz et al. 2002; Díaz et al. 2006; Fouillet et al. 2007; Fouillet et al. 2008; Gershunov et al. 2009). Also timing within the summer is important, as studies have indicated EHEs earlier in the season have a larger impact on mortality (Kalkstein and Smoyer 1993; Rooney et al. 1998; Hajat et al. 2002; Páldy et al. 2005; Baccini et al. 2008).

From the climate community, extreme temperature trend analyses that include a duration requirement (e.g. Gaffen and Ross 1998; New et al. 2006; Gershunov et al. 2009; Kuglitsch et al. 2010) are much less common than studies of trends in single-day extremes (e.g. 90th and 10th percentiles). Studies without duration requirements exist on the global scale (Frich et al. 2002; Alexander et al. 2006), as well as regional studies including focus on the CONUS (Gaffen and Ross 1998; DeGaetano and Allen 2002; Peterson et al. 2008; Portmann et al. 2009), the northeastern CONUS (Griffiths and Bradley 2007; Brown et al. 2010), Italy (Tomozeiu et al. 2006), Europe (Klein

Tank and Können 2003; Moberg and Jones 2005; Della-Marta et al. 2007), South Africa (New et al. 2006), and south and central Asia (Klein Tank et al. 2006).

EHEs with unusually elevated temperatures in both daily minimum temperatures and daily maximum temperatures have been associated with significant mortality impacts. Fouillet et al. (2006) showed in France during the 2003 EHE, mortality was linked to simultaneously elevated minimums and maximums. Similarly Karl and Knight (1997) in the 1995 Chicago EHE, Henschel et al. (1969) in the 1966 St. Louis EHE, and Grumm (2011) in the 2010 Russian EHE observed extreme daily minimums and maximums. Most studies only impose requirements on the daily minimums (e.g. Tamrazian et al. 2008) or maximums (e.g. Tomozeiu et al. 2006; Kyseĺý et al. 2010; Huth et al. 2000; Meehl and Tebaldi 2004). This could give the impression that trends of EHEs focusing on different daily temperature extremes, or both daily temperature extremes, are interchangeable. Kuglitsch et al. (2010) is the only study we're familiar with that required both extremes to be simultaneously elevated over a threshold.

Epidemiological studies have linked mortality to daily minimum temperatures (e.g. Kalkstein 1991; Grize et al. 2005; Schwartz 2005; Hajat et al. 2006; Fouillet et al. 2007; Fouillet et al. 2008; Basu et al. 2008), daily maximum temperatures (e.g. Hajat et al. 2002; Díaz et al. 2002; Tan et al. 2007; Baccinni et al. 2008; Anderson and Bell 2009) and daily mean temperatures (Hajat et al. 2002; Hajat et al. 2006; Basu et al. 2008; Anderson and Bell 2009; Ostro et al. 2009). While it is not clear which daily temperature is closer related to mortality, the heat-mortality relationship likely varies with region (Kalkstein and Davis 1989). For example, in Texas duration may play a larger role or in Michigan maybe the daily minimum temperatures are more important. Some studies use biometeorological indices (e.g. heat index) as predictors (Grize et al. 2005, Hajat et al. 2006), but sensitivity analyses have not found substantial differences with conventional temperatures (Anderson and Bell 2009; Vaneckova et al. 2011). Within the climate literature many of the analyses do not examine the attributes found to be important in the epidemiological literature.

Many trend analyses employ popular extreme heat indices such as the "Warm Spell Duration Indicator" (e.g. Frich et al. 2002; Alexander et al. 2006; New et al.

2006; Brown et al. 2010), Heat Wave Frequency index (e.g. Wu et al. 2012) and Heat Wave Duration Index (e.g. Griffiths and Bradley 2007). While these indices can be good for comparing to other studies, they can also be seen as too simplistic and/or inflexible for some purposes. For instance, none of these indices require the daily minimum temperatures to be elevated, quantify the number of separate EHEs or recognize EHE's with shorter durations (e.g. three days).

Instead a multi-aspect approach describing the different characteristics was taken in the current study. Quantifying the trends in only one aspect only partially explains the trends, and trends in indices sensitive to numerous aspects might not be specific enough. For instance trends in duration, intensity and frequency all provide different information. Nevertheless, some scientists specifically want indices that incorporate numerous aspects. Perkins and Alexander (2013) recently took a similar multi-aspect approach in their analysis of Australian EHEs. Another study that did a robust job of EHE characterization was a study by Gershunov et al. (2009), which evaluated the California region's trends in EHE intensity, duration and spatial extent. The Kuglitsch et al. (2010) study quantified the EHE intensity as well as duration and number of EHEs per summer in the Mediterranean region.

As detailed in section 2.2, the current study was designed to address the aforementioned gaps. For instance, the current study requires duration and examines trends in duration. Also inspected were the trends in EHEs with both daily temperature extremes simultaneously meeting the EHE requirements, and furthermore the differences were quantified in EHE trends based on different daily temperature extremes. Lastly, the current study examined multiple aspects of EHEs in addition to indices encompassing multiple aspects (e.g. the number of EHE days per summer). An examination of EHE timing within the season was also accomplished, through evaluating the difference between trends based on the early portion of the summer (i.e. before July 1) and those based on the mid-late summer (i.e. after June 30).

Similar to this study, DeGaetano and Allen (2002) focused on the CONUS trends starting in 1930. The EHE required either two or three days of thresholds to be met, at various percentiles, for either (not both) the daily minimums or maximums.

The current study has numerous differences from the DeGaetano and Allen (2002) study, for instance the EHEs were specific to summertime and used dynamic (with calendar date) percentiles instead of annual percentiles.

Another similar study, by Gaffen and Ross (1998), quantified modern trends in EHEs across the CONUS. The authors presented linear trends since 1949 in apparent temperature (Steadman 1984) based EHEs. The current study has differences than Gaffen and Ross (1998) such as independence from the hydrological trends, quantifying multiple EHE characteristics and including the hot 1930's decade. As described in section 2.2, conventional temperatures facilitated this analysis. They simplify interpretation of the results, are easier to project into the future and remove the issue of needing a homogenized long-term time series of water vapor observations.

The analysis aims of the current study focused on some specific questions about past EHEs. Did CONUS average linear trends exist in the characteristics of EHEs from 1930 to 2010? Was there spatial structure in those trends over the CONUS, and what did it look like? Did the trends differ within that time period, and how? Did discrepancies exist between the trends of EHEs with definitions based on different daily temperature extremes? Were the trends in EHE characteristics during the earlier portion of the summer similar to those of the whole summer? Lastly, did the summer average temperature trends behave differently than EHE trends?

The organization of this thesis chapter is as follows: In the next section 2.2 the datasets and definitions were described. In section 2.3 the results were presented and in section 2.4 discussed. In section 2.5 the conclusions were discussed and suggestions for future work provided.

2.2. Methods

2.2.1. Dataset and temporal downscaling

This investigation required a daily dataset that was appropriate for trend analysis (i.e. homogenized). The process of homogenization removes non-climatic biases (e.g. urbanization, microclimate changes, station relocations, etc.) that cause temporal discontinuities in time series of climate observations (as described in section 1.4.1). To form this dataset a temporal downscaling method similar to that used by Di Luzio et al. (2008) and Hamlet and Lettenmaier (2005) was used that mapped day-to-day variability onto monthly resolution data. This method provided a daily dataset more suitable for trend analysis. This approach was chosen because a dataset with an effective daily adjustment scheme has not been devised and made available yet.

This combined information from two products of the United States Historical Climatology Network (USHCN). The USHCN version 1 dataset (USHCNv1-daily; Menne et al. 2012) is a quality-controlled (Durre et al. 2008) dataset with daily variability, but it is not homogenized. The USHCN version 2.0 monthly dataset (USHCNv2.0-monthly; Menne et al. 2009, 2011) utilizes the same 1218 stations as the USHCNv1-daily dataset, but is both homogenized and has serially complete data at each station over that station's respective operational period. The USHCNv2.0-monthly dataset is ideal for trend analyses (Menne et al. 2010; Williams et al. 2012) because the homogenization process regarding urbanization is robust (Menne and Williams 2005, 2009; Hausfather et al. 2013).

The USHCNv1-daily dataset was acquired online from the Carbon Dioxide Information Analysis Center (CDIAC), but only the data from 1930 through 2010 and from May though September. During that period 9.78% (9.81%) of the station-months from the Tmax (Tmin) variable were missing. The focus on the summer months is important to heat-health scientists; it is the season elevated temperature manifests itself as a hazard to human health. The starting date 1930 was chosen because our confidence in Co-Op based data was lower prior to 1930 than after it (J. A. Andresen 2012, personal communication). Furthermore the performance of the homogenization process of the USHCNv2.0-monthly dataset (Menne and Williams 2009) was low before 1930, but high after (Hausfather et al. 2013).

The daily temperature values first needed to be converted into anomalies relative to their respective monthly means. The monthly mean temperature ($\overline{T_{m,y}}$) from the USHCNv1-daily dataset was calculated in the typical manner:

$$\overline{T_{m,y}} = \sum_{Tot\_day_m}^{d=1} Tdaily_{d,m,y} \Big/ Tot\_day_m$$

Tdaily represents the USHCNv1-daily temperature, **Tot_day** the total number of days in each month, **d** the day of the month, **m** the month and **y** represents each year. This was done for months with no consecutive missing/flagged days and less than four total missing/flagged days. Months with more missing/flagged days were not used because missing/flagged data points complicate the ability to estimate the monthly mean temperature accurately. Accurate monthly means are required for accurate calculation of anomalies. That meant 6.07% (6.19%) of the months for the Tmax (Tmin) variable did not pass these requirements.

Missing/flagged days were infilled via bi-linear interpolation using the values one day before and after the date. This infilling method might affect the accuracy of single-day percentile exceedence count trends (DeGaetano et al. 2002), however this study examines multiple-day events and thus the infilling of a single day should be small. Quantifiably, the error this infilling introduces into estimating the monthly mean was 0.03 multiplied by the degrees the estimated sum was over (or below) the actual monthly sum. For example if you infilled three days and those estimates were a sum of 16 degrees over the actual temperatures, then the monthly average temperature would be just shy of 0.5 degrees over the actual monthly average temperature. However, this added uncertainty is not a function of time and thus likely has a minimal effect on quantifying trends. For the Tmax (Tmin) variable, 7.88% (8.52%) of the total months had one missing/flagged date, 1.76% (1.94%) of the months had two missing/flagged dates and 0.64% (0.67%) had three missing/flagged dates. Then each day's temperature was converted into an anomaly with respect to its monthly mean temperature in this manner:

$$Anomaly_{d,m,y} = Tdaily_{d,m,y} - \overline{T_{m,y}}$$

d represents the day of the month, **m** the month and **y** the year.

The daily anomalies were combined with the USHCNv2.0-monthly dataset. First the fully adjusted (F52), September 2011 version of the dataset was acquired from the CDIAC over the same time-season span as the daily data. To create the new temperature time series the anomaly time series was combined with the USHCNv2.0-monthly mean temperature time series in the following manner:

$$NewTemperature_{d,m,y} = Tmonthly_{m,y} + Anomaly_{d,m,y}$$

Tmonthly represents the USHCNv2.0-monthly temperatures. The resulting time series at each station comprised the resulting dataset, and retained both the homogenization of the USHCNv2.0-monthly dataset and the daily variability of the USHCNv1-daily dataset. This process has its limitations and is at best an empirical method. The assumption that the monthly homogenization-adjustments made in the USHCNv2.0-monthly dataset apply equally to all daily values (within that month) may be weak, but its accuracy likely does not change with time. Thus the uncertainty in estimating trends is probably minimal.

Three time periods: 1930-1970, 1970-2010 and 1930-2010 were explicitly evaluated. This allowed quantification of the entire period as well as two periods of equal length. This isolated the general CONUS warming since 1970 and the general cooling between the mid 1930s and 1970 (Menne et al. 2009; Hansen et al. 2010), which informs the description of uncertainty for end users. Splitting the total period into separate sections allowed improved spatial density, as compared with the full time span. Even when accounting for spatial variability of the sampling density in the calculations of the CONUS spatial mean (*CONUS mean*) trends, some uncertainty still exists from sampling at different locations. So recalculating half-period trends using the 161 overlapping stations enabled confirmation of the results.

Since the resulting downscaled dataset was not serially complete, a series of quality criteria were needed to filter out inappropriate stations. These requirements were applied separately for each study period, and all stations meeting the requirements during that study period were retained. Stations employed to determine the overall 1930-2010 period were required to pass the requirements for both the 1930-1970 and 1970-2010 periods.

The first criterions insured each station would have enough data to assume uncertainty due to missing data was negligible. Specifically, each station was required to a) have available data for at least 80% of the years in the period, b) have available data at least 70% of the years within each half of each period and c) have available data for both the first or second and last or second-to-last years in the period. A year's data was only considered *available* if all five summer months were present for both daily temperature extremes. This left 581 stations for the 1930-1970 period, 312 stations for the 1970-2010 period and 171 stations for the 1930-2010 period. The discrepancy between older and newer periods was due to the amount of available data within the resulting dataset. Each station was sequentially tested for meeting the aforementioned a, b and then c requirements; and resulted in 844, 33 and 29 disqualifications (respectively) in the 1970-2010 period; and 564, 34 and 39 disqualifications in the 1930-1970 period. The amount of available data in the resulting dataset was controlled by both the completeness of the USHCNv1-daily dataset and the station density of the USHCNv2.0-monthly dataset (the USHCNv2.0-monthly dataset is serially complete but not all stations span the entire dataset). Interestingly, the USHCNv2.0-monthly dataset station density was lowest at the end of the 1930-2010 period (Menne et al. 2009) and played a major role of the newer periods shortage of appropriate stations.

The next criterion focused on station siting quality. A list of station rankings was acquired from surfacestation.org (http://surfacestations.org/fall_etal_2011.htm on Feb. 14, 2012) and any station with the lowest ranking was excluded. This further reduced the number of stations for the 1930-1970 period to 546, the 1970-2010 period to 303 and the 1930-2010 period to 165.

The next criterion was sufficient data for a climate base sample (which is discussed in section 2.2.2). Here the availability was determined at the monthly level and it was not dependent on the other daily temperature extreme. It was required that at least 125 (out of 150 possible) months during the climatological period be available and at least 25 (out of 30 possible) years be available for each month. This brought the number of stations for the 1930-1970 period to 543, for the 1970-2010 period to 302 and for the 1930-2010 period to 165. The last criterion was based on originality of the monthly values from the USHCNv2.0-monthly dataset. At least 80% of the present months in the climate base period were required to be original values (i.e. not infilled/estimated within the USHCNv2.0-monthly dataset). This reduced the number of stations for the 1930-1970 period to 541, for the 1970-2010 period to 295 and the number of stations in both periods to 161. These sets of stations can be seen in figures 2.3-2.5.

2.2.2. Percentiles, extreme heat events

Each station's time series of temperature was turned into a time series of percentiles relative to both calendar date and station. These percentiles were sensitive to the geographical location and the annual cycle. Thus they accounted for human adaption, which is important to scientists of the heat-health discussion. The percentiles at each station were determined from the values at that station during a common period called the climate base period. Here the climate base period for the 1930-1970 period was 1941-1970 and the climate base period for both the 1970-2010 and 1930-2010 periods was 1970-1999. These climate base periods exclude the hot decades of the 1930s and 2000s.

To avoid the climate base period-related inhomogeneity described by Zhang et al. (2005); the current study replaced such years in the climate base sample by a time series that was constructed by averaging the time series' of the other 29 years within the climate base period. The method of sampling used to select the values within the climate base period for calculating the percentiles was a window size of 15-consecutive days centered on the date (Figure 2.1). This size was chosen as Zhang et al. (2005) explained a 5-consecutive day window could be problematic and a 25-consecutive day window was big enough to have no issues, except a general lessened sensitivity to the peak of the seasonal cycle in mid-late July. Thus a 15-consecutive day window was between the two.

The percentiles were calculated empirically; first the empirical cumulative distribution function (Kaplan and Meier, 1958) was calculated, which assigned a percentile (y_o) to each temperature value (x_o) from the climate base period sample. Subsequently, bi-linear interpolation was used to find the value (y_i) of the aforementioned percentile function (y_o) at the target temperature value point (x_i) in the climate base period temperature function (x_o). If the target temperature value (x_i)

was larger (smaller) than any value in the climate base period sample (x_o) then it (y_i) was assigned a 1.0 (0.0). For example, if the empirical cumulative function indicated that 20 °C corresponded to the 50th percentile and 22 °C to the 54th, then 21 °C would be the 52nd percentile.

The exceedence threshold in EHE definitions varies throughout the literature but is usually between the 81st and 99th percentiles. This study chose the 92.5th percentile because a) the 90th percentile was exceeded too frequently (Zhang et al. 2012), b) the 99th percentile was exceeded too infrequently (Zhang et al. 2012) and c) the higher the percentile the more uncertainty exists in the mathematical estimation of that percentile.

EHE duration requirements also vary throughout the literature. Here trends in EHE duration were explicitly quantified, so this study only required *any* duration, and thus two consecutive dates with percentiles over the 92.5th percentile were required to start an EHE. Requirements of EHE continuation vary in the literature, some studies require consecutive exceedences of the threshold (Alexander et al. 2006; New et al. 2006; Griffiths and Bradley 2007), and others require an EHE-mean temperature to stay above a threshold (e.g. Huth et al. 2000; Meehl and Tebaldi 2004). The current study terminated an EHE when the EHE-mean percentile no longer exceeded the aforementioned 92.5th percentile threshold. It was assumed that humans would consider a time period as one continuous event if an isolated inner date marginally failed to exceed the threshold, and thus it was felt that this flexibility might relate better to human heat stress. The EHE-mean percentiles ($\overline{P}$) were calculated as follows:

$$\overline{P_i} = \left. \sum_{i}^{n=1} P_n \middle/ i \right.$$

n represents each day within an EHE and **i** the day number within the EHE. For example, if you had four consecutive dates of 95, 95, 92.5 and 60 percentiles, then the EHE-mean percentile was 95 on the first two dates, 94.2 on the third date and 85.6 on the fourth date. Thus the EHE lasted three dates and had an intensity of 5 CPE.

Which daily temperature extremes are required by the EHE definition to exceed the threshold also varies, but is rarely discussed in the literature. This study separately quantified EHEs requiring a) only daily minimum compliance (Tmin type), b) only daily maximum compliance (Tmax type) and c) both minimum and maximum compliance (Tmnx type). In this study EHE *type* refers to one of these three EHEs, and was used to differentiate between EHEs with definitions based on the different daily temperature extremes. Discriminating Tmnx EHEs from Tmax and Tmin type EHEs was essential because in our dataset on a Tmin or Tmax type EHE date the opposing daily temperature extreme percentile was over the same threshold (92.5 percentile) roughly 35% of the time, over a lower threshold (80.0 percentile) around 60% of the time and above normal (50.0 percentile) about 85% of the time.

The first two EHE types were diagnosed using the aforementioned requirements, but Tmnx type EHEs were diagnosed using the time series of Tmin and Tmax type EHEs. Tmnx EHEs were triggered when either a) two calendar dates overlapped of Tmin type and Tmax type EHEs or b) the two Tmax type EHE dates were one single date behind the Tmin type dates (Figure 2.1). A Tmnx EHE event continued forward in time until a date was reached that was not diagnosed as an EHE by either Tmin or Tmax type. Thus, half days were effectively the temporal resolution of the Tmnx EHE diagnosis and whole days for Tmin and Tmax type EHEs (Figure 2.1). Since this study was focused on the summertime the earliest (latest) possible EHE start (end) date was May (September) 15.

2.2.3. EHE characteristic trends, significance and spatial averages

The current study investigated multiple aspects of EHEs, ranging from objective to inclusive. A value was calculated for every year that the station had available data for the following five specific aspects, or *EHE characteristics*: the number of EHEs, the total number of days categorized as EHEs (EHE-days), the mean EHE duration, the mean EHE intensity and the sum of all EHE intensities. The EHE mean duration, number per year and intensity were considered objective characteristics because they focused on specific aspects of EHEs. EHE-days and sum intensity were considered inclusive because they describe the combination of multiple aspects of EHEs. The Pearson's correlation coefficients between the different EHE characteristics were calculated over the 1930-2010 period using the CONUS-mean time series (discussed later in this section).

The intensity for each EHE was calculated in a cumulative degree-day manner, similarly to Díaz et al. (2006). However, here percentiles replaced the role of temperatures. Calculating the intensity (w.r.t. either daily temperature extreme) was determined by:

$$Intensity_x = \sum_{dur_x}^{n=1} (P_n - 92.5)$$

x represents each EHE**, dur** the duration in days for that EHE**, n** each day within that EHE, **P** each dates percentile. Percentiles below the 92.5 value were set to 92.5. The intensity was based on the daily maximum percentiles only for the Tmax EHE type, on the daily minimum percentiles for the Tmin EHE type and on both daily minimum and maximum percentiles for the Tmnx EHE type. The intensity was calculated for each EHE and the units were referred to as cumulative percentile exceedences (CPEs). Physically, CPEs depend on how far the percentiles exceeded the threshold and how long the event lasted. The sum intensity then depends on both mean intensity and the number of events in a summer.

The linear trends of each EHE characteristic, for the three time periods were calculated for each EHE type and station. Both the magnitude and significance of trends were determined using the often-used least squares regression method. Significance here refers to non-inclusion of the zero value within the 90 percent confidence interval. While the results were verified with non-parametric alternatives (not shown), the study was not primarily undertaken using them because questions exist about the effect of missing values on non-parametric tests.

When calculating the CONUS means, for added confidence two separate methods were employed. The first was used to create time series, per region and EHE type, from 1930-2010. This method first divided the country into six geographical regions (north west, south west, north central, south central, north east and south east; delineated using the 100-degree longitude line, the 39.72-degree latitude line and either the 83 or 87-degree longitude line (in the north and south, respectively)). Then stations from the 161 stations that spanned 1930-2010 within the regional boundaries produced arithmetic means for each region (*regional means*). The CONUS mean was then the weighted (by spatial area) mean of those regional means.

The second method determined the CONUS-mean linear trends for each period. Grids over the CONUS at roughly 100 km resolution were filled by the estimated linear trends (at each station) within their boundaries using the nearest neighbor spatial interpolation method (Gold 1991), which places randomly spaced data onto the nearest locations of a grid. Grids without stations in their boundaries had values infilled by solving Poisson's equation via relaxation over the input domain using NCL (Brown et al. 2012). The CONUS means were calculated with all the retained grid points by weighting grid values by the cosine of the latitude. This was done to compensate for the smaller spatial areas represented by grid points at higher latitudes.

2.2.4. Spatial variability

The spatial characteristics of the different trends were examined via maps showing the trend magnitudes at each station and their significance. Visual analysis was chose as it allows one to look beyond the spatial coverage biases and differences when assessing spatial patterns and scales of spatial variability.

2.2.5. Sensitivity to daily temperature extreme and timing

To explore the relationships between the trends in Tmin, Tmax and Tmnx EHE types, the Pearson's correlation coefficients were calculated across all stations between the Tmin and Tmax, the Tmin and Tmnx and between the Tmax and Tmnx trend magnitudes. To further explore the relationships between the EHE type trends, Student's t tests determined if the arithmetic mean trend magnitudes of the different EHE type trends were statistically different at the 0.01 significance level. The spatial coverages were identical across the comparisons and thus uncertainty due spatial coverage differences should be minimalized.

To explore the relationships between the trends in EHE characteristics of early and whole season EHEs, the number of EHE days per summer between 15 May and 30 June (early season), 1 July and 15 September (mid-late season), 15 May and 15 September (whole season) EHEs was compared. EHE-days was the EHE characteristic of choice due to its high correlation with the other EHE characteristics and its inclusive nature. Comparisons using other characteristics were executed and confirmed our findings, but the results are not shown. The early period is 38% the length of the total summer length, and thus if the early part of the summer is changing the same as the whole summer, then the trend magnitude (in the early part of the summer) will be 38% of that of the whole summer season magnitude.

First, CONUS-mean trends were calculated and compared across all three seasons (early, mid-late, whole). To gauge statistical significance, Student's t-tests were used to test whether the arithmetic means (over all locations) of the early season trends were different than 38% of the arithmetic means in the whole season trends (the early season is 38% of the whole season). This testing was for a difference at the 0.1 statistical significance level between the sample averages. Next, maps of the trends in all three seasons were visually compared. Lastly, correlation coefficients over all locations were calculated between trends in early, mid-late and whole seasons to quantify the relationships between the spatial patterns.

2.2.6. Relationship with mean temperature trends

The last relationships explored were between the summer average temperatures and (whole season) EHE characteristic trends. Again, the number of EHE days was the EHE characteristic used for comparison with the summer average temperatures. Summer average temperatures were calculated for daily maximums, daily minimums and daily means (constructed by averaging the two time series together). The summer averages were calculated using the same calendar dates (15

May to 15 September) and the trends and magnitude were calculated the same as the EHE trends were.

Examination of the summer average temperature trends at the continental scale was by calculating the percentages of stations with positive and negative trends (and corresponding significant trends) in each time period for each daily temperature extreme (daily average included) and comparing those percentages with those percentages in the EHE trends. Then, Pearson's correlation coefficients were calculated through time (at each station) to assess the relationships between the temporal variability (e.g. speaks to the predicative ability on shorter time scales). To assess the spatial pattern similarities, maps were created and visually compared with the EHE trends. Furthermore, correlation coefficients over the spatial domain were calculated between trends in average temperature and the trends in EHE-days.

2.3. Results

The arithmetic mean values for all EHE characteristics over all stations and years are listed in table 2.1 for both the 1930-1970 and 1970-2010 periods. These indicate that EHEs based on the daily maximum temperatures (Tmax type) and those based on the daily minimum temperatures (Tmin type) had roughly twice the number of events per summer, 1.5 times the duration but only .75 of the intensity of EHEs based on both daily temperature extremes (Tmnx type). Albeit, the technique of calculating the intensity for each EHE type potentially plays a large role in these differences in intensity. These results also indicate Tmax type EHEs were typically slightly longer, more frequent and intense than Tmin type EHEs.

Time series of the CONUS mean EHE-days for all three EHE types (Figure 2.2) demonstrated both temporal and regional variability of EHEs. They showed parallels in the behavior of different EHE types (e.g. the deficiency of EHE-days during the 1960's and 1970's) and differences (e.g. a stronger recent increase in Tmin type EHEs and stronger past decrease in Tmax type EHEs). Correlation coefficients between time series of different EHE characteristics of the same EHE type (Table 2.2) indicated they behaved coherently; the lowest coefficient being 0.85

(mean duration, number of EHEs per year) and highest being 0.99 (sum EHE intensity, EHE-days).

2.3.1. Continental scale trends

The results of CONUS-mean linear trends are provided (Figure 2.3, Table 2.2). These results show that all EHE type trends were negative in the 1930-1970 period and positive in the 1970-2010 period. However, for the Tmin (Tmax) type EHE, the increase during the 1970-2010 period was larger (smaller) in magnitude than the 1930-1970 decrease. This led to negative (positive) trends in the 1930-2010 period for Tmax (Tmin) type EHE characteristics. For the Tmnx type of EHE, the two time periods had increasing and decreasing trends of similar magnitude, which led to small positive trend magnitudes during 1930-2010. The 1930-2010 CONUS-mean EHE duration trends appeared to have stronger trends in the number of EHEs, and mean EHE intensity trended stronger in Tmax type EHEs than in the Tmin type. This resulted in stronger EHE-day trends in Tmin but stronger sum EHE intensity trends in Tmax. While uncertainty exists from using different stations during different periods, the results of trends re-calculated only using the overlapping 161 stations (not shown) indicated it to be negligible.

2.3.2. Spatial variability

Trend maps indicated both small-scale (i.e. sub regional) spatial variability as well as regional-scale variability. The small-scale variability might arise physically in part from the local level where LCLU play larger roles (Gallo et al. 1996); Pielke Sr. et al. (2011) note LCLU has significantly changed since 1920. Additionally, because the majority of stations have some amount of missing data, differences in which years were missing between stations also can cause small-scale variability in linear trend estimations. Differences in the regional-scale spatial patterns were not distinguishable between different EHE characteristics (Figure 2.4), and so only the EHE characteristic-wide spatial patterns are described here; the individual maps are available to be analyzed (Figures 2.5-2.12). To assist in visual comparison, figures with maps grouped by time period were also provided (Figures 2.13-2.17). Albeit small, the largest discrepancies in spatial patterns between EHE characteristics were observed between the mean duration and number per year. Differences in regional spatial patterns between different EHE types were noticeable, but not as large as the differences between time periods.

The first maps described are those of the trends in the EHE characteristics during the 1930-1970 period (Figures 2.5-2.7). Here the stations with decreasing trends out numbered those with increasing trends. The common pattern was that of decrease from Montana eastward as far as New England and southward through most of the south east. The area from roughly Texas westward and northwestward all the way to the west coast was mixed to generally increasing. The areas of decrease were more convincing in the Tmax type EHE trends, and the areas of mixed sign were more positively biased in the Tmin type EHE. The Tmnx type EHE trend spatial pattern appeared roughly a blend of the other two EHE types.

The maps of the trends in EHE characteristics during the 1970-2010 period (Figures 2.4, 2.8-2.9) demonstrated a common spatial pattern of a loose horseshoe shape of increase (e.g. increase in the west, south and east parts of the CONUS). The northern areas were mixed but favored decreasing trends. There existed two areas of decrease; one centered on roughly Idaho and another on Minnesota. Areas of increase and decrease varied in strength depending on EHE type, with the Tmin (Tmax) type showing more increase (decrease). The area from central Texas north through Missouri and Arkansas generally changed dominant sign based on EHE type. The Tmnx type EHE trend spatial patterns were a mix of the other two EHE types.

The maps of trends during the 1930-2010 period (Figures 2.10-2.12) showed both the most prominent regional-scale variability and the most balanced number of stations with increasing and decreasing trends. The general spatial pattern was that of decreasing trends in the continental center (extending north and east of the center), and increasing trends in the west, north east and south east and south areas. Consistent with the other time periods, the areas of increase (decrease) were prominent in the Tmin (Tmax) EHE type trends, and the Tmnx EHE type trend spatial pattern was a mix between the two.

The results of testing for regional variability by separating the CONUS into six regions are given in table 2.3 (region's sample average number of EHE days) and table 2.4 (results of the Student's t-tests between regions). Table 2.3 confirms what can be seen on the maps and time series: the northern central region had the smallest positive trend magnitude during the period of increase and largest trend magnitude during the period of decrease, resulting in a negative trend magnitude during the overall 1930-2010 period. There were also other regions of interesting behavior such as the southeastern region (largest Tmin type EHE trend magnitudes in all periods). Table 2.4 confirms that the regions were fairly independent of one another (75 out of 135 comparisons were determined as different); as well that the northern central region behaves the most independently of the other regions.

2.3.3. Relationships between EHE type trends

The results of the correlation coefficients and Student's t-tests between the EHE types are shown in table 2.5. The correlation coefficients between the EHE characteristics of the Tmin and Tmax EHE types were notably smaller than those between either types and the Tmnx type. Student's t test results showed that the arithmetic means were usually statistically different between EHE types. While about 20% of the Tmin and Tmnx, and Tmax and Tmnx means were statistically dissimilar, all means were dissimilar between Tmax and Tmin types. Although modest, correlation coefficients indicated stronger relationships between Tmin and Tmnx type EHE trends than Tmax and Tmnx type EHE trends.

2.3.4. Relationship between early and whole/late season trends

First, the results of the comparison between the trends in the number of EHE days during the different portions of the summer season at the continental level were given (Table 2.6). The results suggest the 1930-1970 period consistently had early season trends statistically dissimilar from the whole and mid-late season trends.

Generally, early season trends and mid-late season trends were rarely close (+/- 5%) to the 38% and 62% respective values of the corresponding whole season trends. The early and mid-late season trends do not always equal the whole season trends because all trends were estimates of the true trends.

The spatial structure of the early season EHE trends was first evaluated by comparing maps created of the trends in the number of early season EHE days with the corresponding maps of the whole and mid-late seasons (Figures 2.18-2.20). The general patterns seemed similar, but regions of disagreement between maps did exist. For instance the west and north west regions were often in disagreement between early and mid-late season trends. Preliminary examination (not shown) of the lack of stations with statistically significant trends in the early season suggested this result was possibly due to statistical distribution effects (e.g. there were more zero values) on the estimated confidence (in the linear trends), rather than a physical signal in the climate system. While the patterns were never wildly different, there did exist regions of differences.

For a more quantifiable relationship, the results of the Pearson's correlation coefficients over the spatial domain were given (Table 2.7). The correlation coefficients between early and whole season spatial patterns ranged from 0.77 and 0.90. All coefficients between early and late season patterns ranged from 0.55 to 0.78. There were not obvious patterns between the coefficients from EHE type or time period.

2.3.5. Relationship between mean and EHE trends

The relationships between the trends in the number of EHE days and summer average temperature trends were evaluated mostly by calculating the percentages of positive and negative trends and comparing them with the EHE trends (Table 2.8). Compared to the EHE trends, the summer average temperature trends often had larger discrepancies in the percentages of positive and negative trends, both significant and insignificant. The 1970-2010 and 1930-2010 period trends in Tmax were exceptions to that, however. The later may have induced the daily mean being positive over the 1930-2010 period, rather than the balanced negative and positive trends in the Tmnx EHE characteristics in the 1930-2010 period.

The average coefficients of temporal correlation (Table 2.9) displayed the temporal variability of each EHE type was most strongly correlated with its summer average counterpart (e.g. summer mean daily minimum with Tmin type EHEs). In other words Tmin (Tmax) EHE types correlated best with summer average daily minimum (maximum) temperatures and Tmnx EHE types correlated best with summer average daily mean temperatures. The correlations between summer average daily mean temperatures and either Tmin or Tmax type EHE-days were stronger than those between opposing daily temperature extremes (e.g. daily minimum temperatures and Tmax type EHE days). The correlations between the Tmax (Tmnx) type EHE and its summer average counterpart had the largest (smallest) correlation.

Next the similarities between the spatial patterns of the trends were assessed. Maps again facilitated a visual analysis (Figures 2.21–2.23). It was notable that in the 1930-2010 period the daily average trends (Figure 2.23) show more positive trends and less significant negative trends than the Tmnx type EHEs, which agreed with table 2.8. Also it seems that there was more consistency (i.e. less small scale variability) in the trends in the summer average maps. Albeit in general the spatial patterns were quite similar and have comparable numbers of stations with statistically significant trends, but there again exist regions with systematic dissimilarities. Particularly notable, the patterns in Tmin summer averages were different than those in the Tmin type EHEs in the upper Midwest region (e.g. Wisconsin) during the 1970-2010 period.

Lastly Pearson correlation coefficients (Table 2.10) enabled quantifiable comparisons of the spatial patterns of trends in summer mean temperature and EHEs. Each EHE type's spatial pattern correlated with its corresponding summer mean daily temperature extreme pattern the best, and similarly the Tmnx type EHE trends correlated best with the summer average daily mean temperature trends. Also the spatial pattern of trends in Tmin and Tmax type EHEs correlated much stronger with the summer average daily mean temperature spatial patterns than the opposite daily temperature extreme (e.g. daily minimum and Tmax type EHEs) did. The largest coefficients between EHE days and summer mean daily mean temperatures were between the Tmax EHE days and daily maximums temperatures. Lastly, these results, in conjunction with table 2.7, indicated that the spatial patterns between summer mean temperatures and the number of EHE days were roughly as correlated as those between total and early season EHEs (excluding Tmax type EHEs which correlated strongly to summer mean Tmax).

2.4 Discussion

Empirically linking physical drivers to the continental scale trend variability was beyond the scope of this study, nonetheless several possibilities are herein considered. Several of the possible driving mechanisms are categorized as climate system oscillations at decadal scales and longer. Teleconnections between sea surface temperatures (SSTs) in the Atlantic and Pacific ocean basins allow oscillations such as the Pacific Decadal Oscillation (PDO), Interdecadal Pacific Oscillation (IPO), El Niño/Southern Oscillation (ENSO) and the Atlantic Multi-decadal Oscillation (AMO) (McCabe et al. 2004) to play major roles in CONUS-mean summertime daily minimum and maximum temperatures (Robinson et al. 2002; Alfaro et al. 2006; McCabe et al. 2004). For example during 1930-1970 the PDO index values generally decreased and during 1970-2010 they increased; this could explain the 1930-1970 (1970-2010) decrease (increase) in EHE activity.

Many potential explanations for the differences in trends between Tmin and Tmax type EHE trends were land cover/land use (LCLU) related, as called for in the Pielke Sr. et al. (2011) study. For instance, the United States irrigation statistics indicate substantial increases over during 1930-1970 compared to that during 1970-2010 (USGS 2012, Segal et al. 1998). Misra et al. (2012) showed agricultural irrigation affects both daily temperature extremes oppositely, and subsequently in the 1930-1970 period these irrigation trends would work against the daily minimum decreasing trends and aide the decreasing trends in daily maximum temperatures. Another possible factor influencing the large Tmin increase during 1970-2010 relative to 1930-1970, was that during 1970-2010 more land area was developed (urban, exurban) than during 1930-1970 (Brown et al. 2005). Any urbanization biases the homogenization algorithm (Menne et al. 2009) failed to eliminate would increase the daily minimum temperatures preferentially (Hausfather et al. 2013).

Atmospheric water vapor trends also should impact Tmin and Tmax EHE type trends. Gershunov et al. (2009) showed high water vapor levels to be conducive (non-conducive) to Tmin (Tmax) type EHEs. Brown et al. (2013) showed CONUS-mean dewpoint temperature and specific humidity generally decreased (increased) during the 1930-1970 (1970-2010) period. This should have encouraged Tmin (Tmax) type EHE decrease (increase) in the 1930-1970 period and Tmin (Tmax) type EHE increase (decrease) in the 1970-2010 period. This could account in part for the discrepancies in the trends of Tmin and Tmax type EHEs during the 1970-2010 period, but not during the 1930-1970 period.

The spatial patterns of EHE characteristic trends could be influenced by several physical mechanisms that are mentioned below. LCLU types and LCLU changes can affect regional-scale trends; Fall et al. (2010) demonstrated that surface temperature trends can both be a function of LCLU type and be impacted by LCLU changes (e.g. agricultural to urban). For instance Fall et al. (2010) demonstrated the temperature trends in agriculture LCLU types were shown to have decreased during the 1979-2003 period. The agriculture LCLU type spatially occupies much of the area of decrease during the 1970-2010 period found in this study (i.e. the central, east central CONUS).

Climate system oscillations likely influenced the observed spatial patterns as well. McCabe et al. (2004) demonstrated that oscillations of the IDO/PDO and AMO could account for 52% of the spatial and temporal variance over the CONUS. Other works (e.g. Pan et al. 2004; Robinson et al. 2002) have also linked the decrease in the central CONUS (i.e. the “warming hole”) to Pacific and Atlantic Ocean basin oscillations.

The changes in moisture (soil, atmospheric) could also potentially impact the observed spatial patterns in EHE trends. McCabe et al. (2004) suggested 22% of the spatial and temporal variance was explainable by drought indices (which are often related to SST oscillations). Brown et al. (2013) observed that the central CONUS experienced atmospheric moistening since 1930 while other regions experienced drying. This moistening was similarly located (spatially) with the region of EHE decrease in the central CONUS. The noted area of disagreement between EHE types from central Texas north through Missouri and Arkansas, is an area where Steiner and Owen (2013, unpublished manuscript) propose an aerosol-circulation mechanism has induced more precipitation during 1998-2007.

The differences between trends in Tmin and Tmax type EHEs might suggest relevancy of physical mechanisms that act oppositely on different daily temperature extremes (e.g. urbanization, atmospheric water vapor (Gershunov et al. (2009), soil moisture (Misra et al. 2012)). The spatial patterns of Tmin EHE type trends having slightly closer relationships to Tmnx, than Tmax did, was sensible since Tmin type EHEs were observed to be less common, and thus could be considered a limiting factor.

The differences between EHE trends in the early portion of the summer and the later (and whole) portion of the summer suggested the physical system is being driven, or forced, differently during the early and later portions of the summer. One possible physical cause within the climate system for the lack of a closer relationship could be associated with trends in winter snow accumulation or spring precipitation affecting soil moisture and subsequently spatial patterns in the early summer EHEs. Another possible cause could be the North American Jet Stream; it is typically positioned differently over North America in the spring-early summer than it is in the mid-late summer, and recent studies have indicated recent changes in the tendencies of the North American Jet Stream positions (e.g. Woollings and Blackburn 2012; Francis and Vavrus 2012).

In regards to the discrepancies between the mean temperature and the EHE trends, these indicate that the mean and the tails of the statistical distributions of daily temperature extremes changed differently. Such incoherency is possible because the mean is sensitive to all parts of the statistical distribution, and the tails are not. This likely occurs for both daily temperature extremes, but these results suggest the mean of the daily maximum temperatures acted more coherently with the statistical left tail than the daily minimums did.

To speculate on a possible physical driver, storm paths changing (e.g. Yin 2005) could increase nighttime cloud cover or soil moisture. That moisture could then reduce the frequency of extremely cold nights without necessarily affecting the extremely warm nights, and subsequently affect the mean temperatures differently than the number of EHEs.

2.5. Conclusions

Over the 1930-2010 period the characteristics of EHEs based on daily minimums (Tmin type) increased, the characteristics of EHEs based on daily maximum (Tmax type) decreased and the characteristics of EHEs with respect to both daily temperature extremes (Tmnx type) increased a small amount. Such conclusions were cemented by CONUS-mean linear trends that showed during the 1930-2010 period all characteristics increased for the Tmin type EHEs, decreased for the Tmax type EHEs and increased by an order of magnitude less than the other EHE types for the Tmnx type EHEs. The CONUS-mean time series of EHE-days for all three EHE types supported these conclusions. In contrast, DeGaetano and Allen (2002) indicated decreasing continental trends in both Tmax EHEs and Tmin EHEs in 1930-1996 (Tmin trends were smaller). The 14 additional (recent) years in our study potentially explain why our conclusions in the Tmin type EHEs were of increase (instead of decrease).

The linear trends in EHE characteristics in 1930-1970 decreased (Tmax type EHEs most prominently), but the 1970-2010 trends of all EHE types increased in EHE characteristics (Tmin type EHEs most prominently). The results of the CONUS-mean EHE characteristics demonstrated that during 1930-1970 all EHE characteristics, regardless of EHE type, decreased and during 1970-2010 all EHE characteristics increased. Time series of the number of EHE-days for all EHE types supported these conclusions. Both times series and CONUS-mean EHE characteristic trends demonstrated a stronger decrease in Tmax type EHEs during the 1930-1970 period and a stronger increase in Tmin type EHEs during the 1970-2010 period. The next step for time series analysis of EHEs is likely one of principal component analysis in hopes of more fully describing the oscillations pertinent to EHEs.

Shen et al. (2011) demonstrated summertime mean daily minimum and maximum temperatures decreased during the 1946-1975 period and both increased during the 1976-2000 period. DeGaetano and Allen (2002) indicated a stronger decrease in extreme percentiles of daily maximums than minimums during 1930-1996, and a stronger increase in the minimums than maximums during 1970-1996. Peterson et al. (2008) showed that both daily minimum and maximum extreme percentiles decreased during 1950-1970 and increased during 1970-2005. DeGaetano and Allen (2002) demonstrated during 1960-1996 the percentage of increasing trend stations was larger than the percentage of decreasing stations. These support our conclusions. These conclusions implied that similar to the CONUS-mean temperatures (e.g. Hansen et al. 2010), EHEs decreased from roughly the mid 1930's until the late 1970's and then increased through 2010; notably Tmax type EHEs were prominent the 1930s and 1950s while Tmin type EHEs more dominant in the 1990s and 2000s.

Regional and small-scale spatial variability existed across the CONUS for all periods, EHE characteristics and EHE types. Small-scale variability was plainly seen in the maps, as was regional-scale variability. Examples of regional variability might be the lack of warming in the north central CONUS during the 1970-2010 period or the cooling in the central CONUS region during the 1930-2010 period. Conclusions of regional variability were also reinforced by the time series results.

Several other studies also showed regional-scale spatial variability (Gaffen and Ross 1998; DeGaetano and Allen 2002; Peterson et al. 2008; Portmann et al. 2009; Wu et al. 2012). The existence of substantial regional variability in the past insinuates differences between regions will also exist in the future, thus CONUS-mean trends were not representative of trends in each region and thus regional analyses would have been substantially more helpful for scientists in the heat-health field. Likewise, several studies confirmed the existence of small-scale spatial variability (e.g. Gaffen and Ross 1998; DeGaetano and Allen 2002; Portmann et al. 2009). The presence of small-scale variability might serve as notice to future EHE trend analyses calculating regional and sub-regional spatial means. The next step in exploration of the drivers behind the spatial patterns in EHEs is likely an empirical orthogonal function analysis of EHE characteristics. However, first a gridded climate dataset appropriate for long-term trend analysis must be created.

The different EHE characteristics behaved coherently. This was evidenced by the different EHE characteristic trends having nearly indistinguishable spatial patterns, strong temporal correlations with one another and by the CONUS-mean linear trends showing the same signs. That is not to say they were identical, however. Other studies that assessed multiple aspects of EHEs (Kuglitsch et al. 2010; Gershunov et al. 2009; Perkins and Alexander 2013) also showed general agreement across trends in the different aspects of EHEs. Such conclusions reduce some of the uncertainty in using single aspects of EHEs to study EHEs. This might also suggest similar coherency in future changes in EHE characteristics.

Conversely the different EHE types behaved less coherently. Relationships found between EHE characteristics of Tmin and Tmax types were modest. By definition EHE characteristics in Tmnx type trends are connected to Tmax and Tmin type EHEs and the relationships were strong, but short of very strong. The correlation coefficients and maps confirmed such conclusions regarding spatial structure, and Student's t tests confirmed them with respect to the large-scale arithmetic mean trends. These conclusions should reaffirm to the climatologist that the different daily temperature extremes respond differently to changes in the climate system. Mathematically speaking, examination of both daily temperature extremes relationships to the forcing variables controlling the same climate system is akin to have two solution sets. These conclusions were consistent with Portmann et al. (2009) that showed different spatial patterns and CONUS means in the 90^{th} percentile exceedence occurrences for minimum and maximum temperatures. Also the Gaffen and Ross (1998) results support these conclusions; they showed meaningful differences in both CONUS and regional-mean trends in the daily minimum and maximum apparent temperature percentile exceedences.

The climate literature could do a better job of emphasizing EHEs based on different daily temperature extremes are not interchangeable. Particularly to scientists outside the climate community, it is not obvious that the trends of EHEs depend on which daily temperature extreme(s) is (are) required to be elevated. Which extreme is optimal depends on both the region and purpose, and not explored here. These conclusions imply that Tmin and Tmax EHE trends could be used to get a good, but not a great, estimate for how EHEs with requirements on both daily temperature extremes (Tmnx) changed or will change. Trends in EHEs with both daily temperature extremes exceeding thresholds should be quantified in trend analyses more often. An epidemiological assessment linking mortality to events requiring both daily temperature extremes to be over a threshold could further knowledge about this type of EHEs. The next step was likely assessing both the trends in EHEs that meet requirements in both daily temperature extremes and the relationships between all three EHE types moving forward in time using global climate model output.

The trends in EHE characteristics during the earlier portion of the summer had differences from those of the mid-late and whole summer. For example, continental-mean results showed the 1930-1970 period to have statistically different early and whole season trends. Spatial correlation coefficients indicated in general the patterns were similar, but not indistinguishable. Maps of the trends indicated that some regions were dissimilar, most notably the western U.S. region. Such conclusions were important because they demonstrated the existence of differences in extreme heat trends between early summer and the other parts of the summer, which hah not previously been explored. These conclusions will hopefully inform heat-health scientists not to rely too heavily on early summer trends behaving the same as whole summer trends. To climate scientists this indicates either a) differences in which factors were influential, or b) the influential factors were behaving differently at different parts of the summer. The next steps in this area are better quantifications of the disparity in trends by time within summer, and linking changes in the physical system (e.g North American Jet Stream position, spring precipitation) to the different trends.

The trends in the characteristics of EHEs and the summer mean temperature were also similar yet had some notable differences. On the large scale the summer temperature trends indicated the same signs but seemed to have stronger signals (increasing or decreasing). The spatial correlations and maps of trends both indicated similar overall spatial structure but that there existed regions of disagreement. These conclusions were important to the heat-health discussion because it warns heat-health scientists not to rely too heavily on summer mean trends for guidance in EHEs. These conclusions were also important because no other studies were found explicitly discussing the relationships between summer mean temperatures and EHE characteristics.

These conclusions first demonstrate that the different portions of the statistical distribution were changing differently, at least at the regional level. Thus the often-seen figure of a probability distribution function of temperatures merely shifting (as a result of climate change) towards warmer temperatures should be retired. Instead the probability distribution is probably going to change shape as well as shift. Shen et al. (2011) demonstrated this as well. The next step in regards to the study of EHEs is exploration into what roles the mean, variability, skew and other higher statistical moments have in the trends of EHEs. From a climatologist's perspective, physical mechanisms of the climate system may control different statistical moments (e.g. the jet stream controlling synoptic variability of temperatures at the surface (Francis and Vavrus 2012)).

Comparing results between extreme temperature trend analyses with different periods is problematic because linear trends are highly sensitive to many factors. For example, they are considerably sensitive to the time period, seasonal focus and daily temperature extreme. Generally, linear trends are very limited in their ability to describe the temporal behavior of any variable in a climate system that behaves non-linearly. Despite this, maps were created to match the time periods of other studies such as 1950-2006 in the Portman et al. (2009) study (Figure 2.24), 1960-1996 in the DeGaetano and Allen (2002) study (Figure 2.25), 1950-2004 in the Peterson et al. (2008) paper (Figure 2.26), and the 1950-1999 in the Meehl et al. (2012) paper (Figure 2.27). It should be noted that the stations used to create these maps were not chosen with the data availability of these time periods in mind, but rather the stations selected for the 1930-2010 period; thus additional spatial noise arises in these maps due to the effect of missing years. It was felt the disparities between our results and comparable studies' results were explainable by the differences in the data being presented in the figures.

Additionally, our results show the so-called "warming hole" investigated by multiple studies (Kunkel et al. 2006; Pan et al. 2004; Robinson et al. 2002; Meehl et al. 2012). There is an amount of debate what mechanisms cause the warming hole, and we felt that this study's results might lend insight. Pan et al. (2004) demonstrated changes in the low-level jet (from the gulf to the Plains) lead to soil moisture changes that suppressed daily maximum temperatures. Regardless of mechanism, the moisture levels have been observed to increase in this region (Brown et al. 2013). Misra et al. (2012) showed increased irrigation statistics to increase (decrease) daily minimum (maximum) temperatures. However a decrease was seen in both Tmin and Tmax type EHEs (albeit Tmin EHE type showed a weaker decrease). Thus, a hypothesis that works with these results is while soil moisture efficiently suppressed Tmax it did not increase Tmin as much as the Tmax-Tmin coupling mechanism (Gershunov et al. 2009) decreased Tmin. An analysis of the Tmax-Tmin coupling mechanism (e.g. which daily temperature extreme is better linked with its trailing opposite daily temperature extreme) would provide relevant knowledge to the EHE field.

2.6. Acknowledgements

This research was sponsored by NOAA Climate Program Office grant NA10OAR4310213 through the Great Lakes Regional Integrated Sciences and Assessments Center (/GLISA/). I would like to thank Frank Marsik, Marie S. O'Neill, Allison Steiner and Jeff Andresen for their helpful conversations. We also thank our reviewers and editor who provided superb critiques to our original submission to the Journal of Applied Meteorology and Climatology.

2.7. References

Alexander, L. V., X. Zhang, T. C. Peterson, J. Ceasar, B. Gleason, A. M. G. Klein Tank, M. Haylock, D. Collins, B. Trewin, F. Rahimzadeh and coauthors, 2006: Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res.,* **111**: D05109.

Alfaro, E. J., A. Gershunov, D. Cayan, 2006: Prediction of summer maximum and minimum temperature over the Central and Western United States: the roles of soil moisture and sea surface temperature. *Journal of Climate*, **19**: 1407-1421.

Anderson, B. G. and M. L. Bell, 2009: Weather-related mortality: how heat, cold and heat waves affect mortality in the United States. *Epidemiology*, **20**, 205-213.

Baccini, M., A. Biggeri, G. Accetta, T. Kosatsky, K. Katsouyanni, A. Analitis, H. R. Anderson, L. Bisanti, D. D'Ippoliti, J. Danova, B. Forsberg, S. Medina, A. Paldy, D. Rabczenko, C. Schindler, P. Michelozzi, 2008: Heat effects on mortality in 15 European cities. *Environmental Health*, **9**, 37, doi: 10.1186/1476-069X-9-37.

Basu, R., W. Feng, B. D. Ostro, 2008: Characterizing temperature and mortality in nine California counties. *Epidemiology,* **19**, 138-145.

Brown, D. G., K. M. Johnson, T. R. Loveland and D. M. Theobald, 2005: Rural land-use trends in the conterminous United States, 1950-2000. *Ecological Applications,* **15**: 1851-1863.

Brown, P. J., R. S. Bradley, F. T. Keimig, 2010: Changes in extreme climate indices for the Northeastern United States, 1870-2005. *J. Climate*, **23**: 6555-6572.

Brown, P. J. and A. T. DeGaetano, 2013: Trends in U.S. surface humidity, 1930-2010. *Journal of Applied Meteorology and Climatology*, **52**: 147-163.

DeGaetano, A. T. and R. J. Allen, 2002: Trends in Twentieth-Century Temperature Extremes across the United States. *Journal of Climate,* **15**: 3188-3205.

Della-Marta, P. M., M. R. Haylock, J. Luterbacher and H. Wanner, 2007: Doubled length of western European summer heat waves since 1880. *J. Geophys. Res.*, **112**: D15103.

Díaz, J., R. García, F. Velázquez de Castro, E. Hernández, C. López, A. Otero, 2002: Effects of extremely hot days on people older than 65 years of Seville (Spain) from 1986-1997. *International Journal of Biometeorology*, **46**: 145-149.

Díaz, J., R. García-Herrara, R. M. Trigo, C. Linares, M. A. Valente, J. M. De Miguel, E. Hernández, 2006: The impact of the summer 2003 heat wave in Iberia: how should we measure it? *International Journal of Biometeorology*, **50**, 159-166.

Di Luzio, M., G. L. Johnson, C. Daly, J. K. Eischeid, J. G. Arnold, 2008: Constructing retrospective gridded daily precipitation and temperature datasets for conterminous United States. *J. Appl. Meteor. and Climatology*, **47**, 475-497, doi: 10.1175/2007JAMC1356.1.

Durre, I., M. J. Menne and R. S. Vose, 2008: Strategies for evaluating quality assurance procedures. *J. Clim. Appl. Meteorol.*, **47,** 1785-1791. doi: 10.1175/2007JAMC1706.1.

Fouillet, A., G. Rey, E. Jougla, P. Frayssinet, P. Bessemoulin, D. Hémon, 2007: A predicative model relating daily fluctuations in summer temperatures and mortality rates. *BMC Public Health*, **7**: 114-125.

Fouillet, A., G. Rey, V. Wagner, K. Laaidi, P. Empereur-Bissonnet, A. Le Tertre, P. Frayssinet, P. Bessemoulin, F. Laurent, P. De Crouy-Chanel, E. Jougla, D. Hémon, 2008: Has the impact of heat waves on mortality changed in France since the European heat wave of summer 2003? A study of the 2006 heat wave. *International Journal of Epidemiology*, **37**, 309-317.

Fouillet, A., G. Rey, F. Laurent, G. Pavillon, S. Bellec, C. Guihenneuc-Jouyaux, J. Clavel, E. Jougla, D. Hémon, 2006: Excess mortality related to the August 2003 heat wave in France. *International Archives of Occupational Environmental Health*. **80**, 16-24. doi: 10.1007/s00420-006-0089-4.

Francis, J. A. and S. J. Vavrus, 2012: Evidence linking Arctic amplification to extreme weather in mid-latitudes. *Geophysical Research Letters*, **39**: L06801. doi: 10.1029/2012GL051000.

Frich, P., L. V. Alexander, P. Della-Marta, B. Gleason, M. Haylock, A. M. G. Klein Tank, T. Peterson, 2002: Observered coherent changes in climatic extremes during the second half of the twentieth century. *Climate Research*, **19**: 193-212.

Gaffen, D. J. and R. J. Ross, 1998: Increased summertime heat stress in the U.S. *Nature*, **396**: 529-530.

Gallo, K. P. D. R. Easterling, T. C. Peterson, 1996: The Influence of Land Use/Land Cover on Climatological Values of the Diurnal Temperature Range. *Journal of Climate*, **9**: 2941-2944. doi: 10.1175/1520-0442(1996)009<2941:TIOLUC>2.0.CO;2.

Gershunov, A., D. R. Cayan, S. F. Iacobellis, 2009: The heat wave over California and Nevada: Signal of an increasing trend. *J. Climate*, **22**: 6181-6203.

Gold, C. M., 1991: Problems with handling spatial data – the Voronoi approach. *Critical Incident Stress Management Journal*, **45:** 65-80.

Grize, L., A. Huss, O. Thommen, C. Schinder, C. Braun-Fahrländer, 2005: Heat wave 2003 and mortality in Switzerland. *Swiss Medical Weekly*, **135**: 200-205.

Griffiths, M. L. and R. S. Bradley, 2007: Variations of twentieth-century temperature and precipitation extreme indicators in the Northeast United States. *J. Climate*, **20**: 5401-5417.

Grumm, R. H., 2011: The Central European and Russian heat event of July-August 2010. *Bull. Amer. Meteor. Soc.*, **92**: 1285-1297.

Hajat, S., B. Armstrong, M. Baccini, A. Biggeri, L. Bisanti, A. Russo, A. Paldy, B. Menne, T. Kosatsky, 2006: Impact ofdaily maximumtemperatures on mortality: is there an added heat wave effect. *Epidemiology*, **6**: 632-638.

Hajat, S., R. S. Kovats, R. W. Atkinson, A. Haines, 2002: Impact of hot temperatures on death in London: a time series approach. *Journal of Epidemiology and Community Health*, **56**: 367-372.

Hamlet, A. F. and D. P. Lettenmaier, 2005: Production of temporally consistent gridded precipitation and temperature fields for the continental United States. *Journal of Hydrometeorology*, **6**: 330-336.

Hansen, J., R. Ruedy, M. Sato, K. Lo, 2010: Global surface temperature change, *Rev. Geophys.,* **48**: RG4004, doi: 10.1029/2010RG000345.

Hausfather, Z., M. J. Menne, C. N. Williams, T. Masters, R. Broberg, D. Jones, 2013: Quantifying the effect of urbanization on U.S. Historical Climatology Network temperature records. *Journal of Geophysical Research: Atmospheres*, **118:** 481-494, doi: 10.1029/2012JD018509.

Henschel, A., L. L. Burton, L. Margolies J. E. Smith, 1969: An analysis of the heat deaths in St. Louis during July, 1966. *American Journal of Public Health and the Nations Health.* **59**: 2232-2242.

Kalkstein, L. S. and R. E. Davis, 1989: Weather and human mortality: An evaluation of demographic and interregional responses in the United States. *Annals of the Association of American Geographers*, **79**: 44-64. doi: 10.1111/j.1467-8306.1989.tb00249.x.

Kalkstein, L. S., 1991: A new approach to evaluate the impact of climate on human mortality. *Environmental Health Perspectives*, **96**: 145-150.

Kalkstein, L. S. and K. E. Smoyer, 1993: The impact of climate change on human health: some international implications. *Cellular and Molecular Life Sciences*, **49**: 969-979.

Kaplan, E. L. and P. Meier, 1958: Nonparametric estimation for incomplete observations. *Journal of the American Statistical Association*. **53**: 457-481.

Klein Tank, A. M. G. and G. P. Können, 2003: Trends in indices of daily temperature and precipitation extremes in Europe, 1946-99. *J. Climate*, **16**: 3666-3680.

Klein Tank, A. M. G., T. C. Peterson, D. A. Quadir, S. Dorji, X. Zou, G. Tang, K. Santhosh, U. R. Joshi, A. K. Jaswal, R. K. Kolli, A. B. Sikder, N. R. Deshpande and coauthors, 2006: Changes in daily temperature and precipitation extremes in central and south Asia. *J. Geophys. Res.*, **111**: D16105.

Kuglitsch, F. G., A. Toreti, E. Xoplaki, P. M. Della-Mart, C. S. Zerefos, M. Türkes, J. Luterbacher, 2010: Heat wave change in the eastern Mediterranean since 1960. *Geophys. Res. Lett.*, **37**: L04802

Kunkel, K. E., X. Liang, J. Zhu, Y. Lin, 2006: Can CGCMs simulate the twentieth-century “warming hole” in the Central United States. *J. Climate*, **19**: 4137-4153.

Kyselý, J., 2010: Recent severe heat waves in central Europe: how to view them in a long-term prospect. *International Journal of Climatol*ogy, **30**: 89-109.

Meehl, C. A. and C. Tebaldi, 2004: More intense, more frequent, and longer lasting heat waves in the 21st century. *Science*, **13**: 994-997.

Menne, M. J. and C. N. Williams, J., 2005: Detection of undocumented changepoints using multiple test statistics and composite reference series. *Journal of Climate*, **18**: 4271-4286.

Menne, M. J. and C. N. Williams, Jr., 2009: Homogenization of temperature series via pairwise comparison. *Journal of Climate*, **22**: 1700-1717.

Menne, M. J., C. N. Williams, Jr., R. S. Vose, 2009: The United States Historical Climatology Network monthly temperature data – version 2. *Bull. Amer. Meteor. Soc.*, **90**: 993-1107.

Menne, M. J., C. N. Williams, Jr., R. S. Vose, 2011: United States Historical Climatology Network (USHCN) version 2 serial monthly dataset. *Carbon Dioxide Information Analysis Center*, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Menne, M. J., C. N. Williams, Jr., R. S. Vose, 2012: United States Historical Climatology Network daily temperature, precipitation and snow data. *Carbon Dioxide Information Analysis Center*, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

McCabe, G. J., M. A. Palecki, J. L. Betancourt, 2004: Pacific and Atlantic Ocean influences on multidecadal drought frequency in the United States. *P.N.A.S.,* **101:** 4136-4141. Doi: 10.1073/pnas.0306738101.

Misra, V., J. P. Michael, R. Boyles, E. P. Chassingnet, M. Griffin, J. J. O'Brien, 2012: Reconciling the spatial distribution of the surface temperature trends in the Southeastern U.S. *J. Climate*, **25**: 3610-3618.

Moberg, A. and P. D. Jones, 2005: Trends in indices for extremes in daily temperature and precipitation in central and western Europe, 1901-99. *International Journal of Climatology*, **25**: 1149-1171.

New, M., B. Hewitson, D. B. Stephenson, A. Tsiga, Andries Kruger, A. Manhique, B. Gomez, C. A. S. Coelho and coauthors, 2006: Evidence of trends in daily climate extremes over southern and west Africa. *J. Geophys. Res.*, **111**: D14102.

Ostro, B. D., L. A. Roth, R. S. Green, R. Basu, 2009: Estimating the mortality effect of the July 2006 California heat wave. *Environmental Research*, **109**: 614-619.

Páldy A, J. Bobvos, A. Vámos, R. S. Kovats, S. Hajat, 2005: The effect of temperature and heat waves on daily mortality in Budapest, Hungary, 1970–2000. In: Kirch W, Menne B, Bertollini R (eds) Extreme weather events and public health responses. Springer, New York, pp 99–107.

Pan, Z., R. W. Arritt, E. S. Takle, W. J. Gutowski Jr., C. J. Anderson, M. Segal, 2004: Altered hydrologic feedback in a warming climate introduces a "warming hole". *Geophys. Res. Lett.*, **31**: L17109. doi: 10.1029/2004GL020528.

Perkins, S. E. and L. V. Alexander, 2013: On the measurement of heat waves. *J. Climate.* **26**: 4500-4517. doi: 10.1175/JCLI-D-12-00383.1.

Pielke Sr., R. A., A. Pitman, D. Niyogi, R. Mahmood, C. McAlpine, F. Hossain, K. K. Goldewijk, U. Nair, R. Betts, S. Fall, M. Reichstein, P. Kabat, N. de Noblet, 2011: Land use/land cover changes and climate: modeling analysis and observational evidence. *WIREs Climate Change*, **2**: 828–850. doi: 10.1002/wcc.144.

Robinson, W. A., R. Reudy, J. E. Hansen, 2002: General circulation model simulations of recent cooling in the east-central United States, *J. Geophys. Res.*, **107**:4748, doi: 10.1029/2001JD001577.

Rooney, C., A. J. McMichael, R. S. Kovats, M. P. Coleman, 1998: Excess mortality in England and Wales, and in Greater London, during the 1995 heatwave. *Journal of Epidemiology and Community Health*, **52**: 482-486.

Schwartz, J., 2005: Who is sensitive to extremes of temperature? A case-only analysis. *Epidemiology*, **1**: 67-72.

Segal, M,. Z. Pan., R. W. Turner, E. S. Takle, 1998: On the potential impact of irrigated areas in North America on summer rainfall caused by large-scale systems. *J. Appl. Meteor.*, **37**: 325-331.

Shen, S. S. P., A. Gurung, H. Oh, T. Shu, D. R. Easterling, 2011: The twentieth century contiguous US temperature changes indicated by daily data and higher statistical moments. *Climate Change*, **109**: 287-317. doi: 10.1007/s10584-011-0033-9

Steadman, R. G., 1984: A universal scale of apparent temperature. *J. Climate and Applied Meteorology*. **23**: 1674-1687.

Tamrazian, A., S. LaDochy, J. Willis, W. C. Patzert, 2008: Heat waves in Southern California: Are they becoming more frequent and longer lasting? *Yearbook of the Association of Pacific Coast Geographers,* **70**: 59-69.

Tan, J., Y. Zheng, G. Song, L. S. Kalkstein, A. J. Kalkstein, X. Tang, 2007: Heat wave impacts on mortality in Shanghai, 1998 and 2003. *International Journal of Biometeorology*, **51**: 193-200.

Tomozeiu, R., V. Pavan, C. Cacciamani, M. Amici, 2006: Observed temperature changes in Emilia-Romagna mean values and extremes. *Climate Research*, **31**: 217-225.

USGS, 2012: “Trends in water use in the U.S., 1950 to 2005, the USGS Water Science School. Accessed 25 September, 2012. [http://ga.water.usgs.gov/edu/wateruse-trends.html]

Vaneckova, P., G. Neville, V. Tippett, P. Aitken, G. FitzGerald, S. Tong, 2011: Do biometeorological indices improve modeling outcomes of heat-related mortality? *J. Climate Appl. Meteor.,* **50**: 1165-1176. doi: 10.1175/2011JAMC2632.1.

Vose, R. S. and M. J. Menne, 2004: A method to determine station density requirements for climate observing networks. *J. Climate*, **17**: 2961-2971.

Williams, C. N., M. J. Menne, P. W. Thorne, 2012: Benchmarking the performance of pairwise homogenization of surface temperatures in the United States. *J. Geophys. Res.*, **117**: D05116. doi: 10.1029/2011/JD016761.

Woollings, T. and M. Blackburn, 2012: The North Atlantic Jet Stream under climate change and its relation to the NAO and EA patterns. *Journal of Climate*, **25**: 886-902. Doi: 10.1175/JCLI-D-11-00087.1.

Wu, Z., H. Lin, J. Li, Z. Jiang, T. Ma, 2012: Heat wave frequency variability over North America: two distinct leading modes. *J. Geophys. Res.*, **117**: D02102, doi: 10.1029/2011JD016908

Yin, J. H., 2005: A consistent poleward shift of the storm track in simulations of 21st century climate. *Geophsical Research Letters*, **32**: L18701, doi: 10.1029/2005GL023684.

Zhang, K., R. B. Rood, G. Michailidis, E. M. Oswald, J. D. Schwartz, A. Zanobetti, K. L. Ebi, M. S. O'Neill, 2012: Comparing exposure metrics for classifying 'dangerous heat' in heat wave and health warning systems. *Environmental International,* **46**: 23-29. doi: 10.1016/j.envint.2012.05.001.

Zhang, X., G. Hegerl, F. W. Zwiers., J. Kenyon, 2005: Avoiding inhomogeneity in percentile-based indices of temperature extremes. *Journal of Climate*, **18**: 1641-1651.

2.8. Figures

Figure 2.1. Two schematics. The top illustrates how EHEs are diagnosed. Light (dark) colored bars indicate Tmax (Tmin) percentile exceedence on that date. Thus the Tmin EHE dates are July 8, 9, 14, 15, 25, 26, 27, 28, 29 and 30; Tmax EHE dates are July 8, 9, 15, 16, 24, 25, 28 and 29; two Tmnx EHEs exist and are diagnosed as two days in length (July 8, 9), two days (July 24, 25, 26) and 2.5 days (July 28, 29, 30) in length. The bottom, of the calendar date-dynamic base climate sample used to calculate the percentiles at each station. Thirty years worth of the dates marked with a dark grey (light grey, mild grey) line through them are used to determine the percentiles on the date marked with a dark grey (light grey, mild grey) X on it.

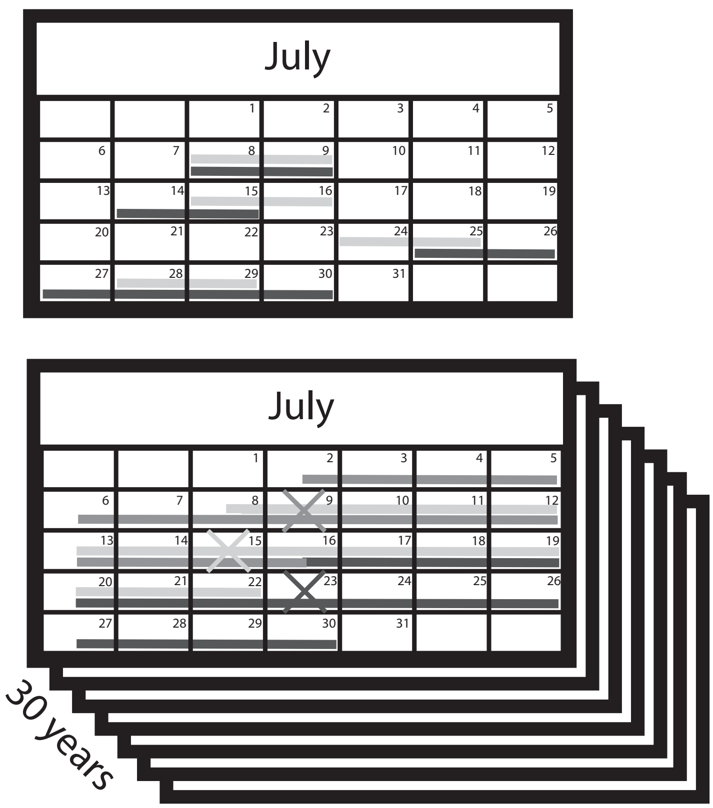

Figure 2.2. Time series of regional and continental average EHE-day values per summer. The legend assigning colors to specific regions provides the number of stations used to create the regional averages and the spatial weight of the region in the continental average, respectively.

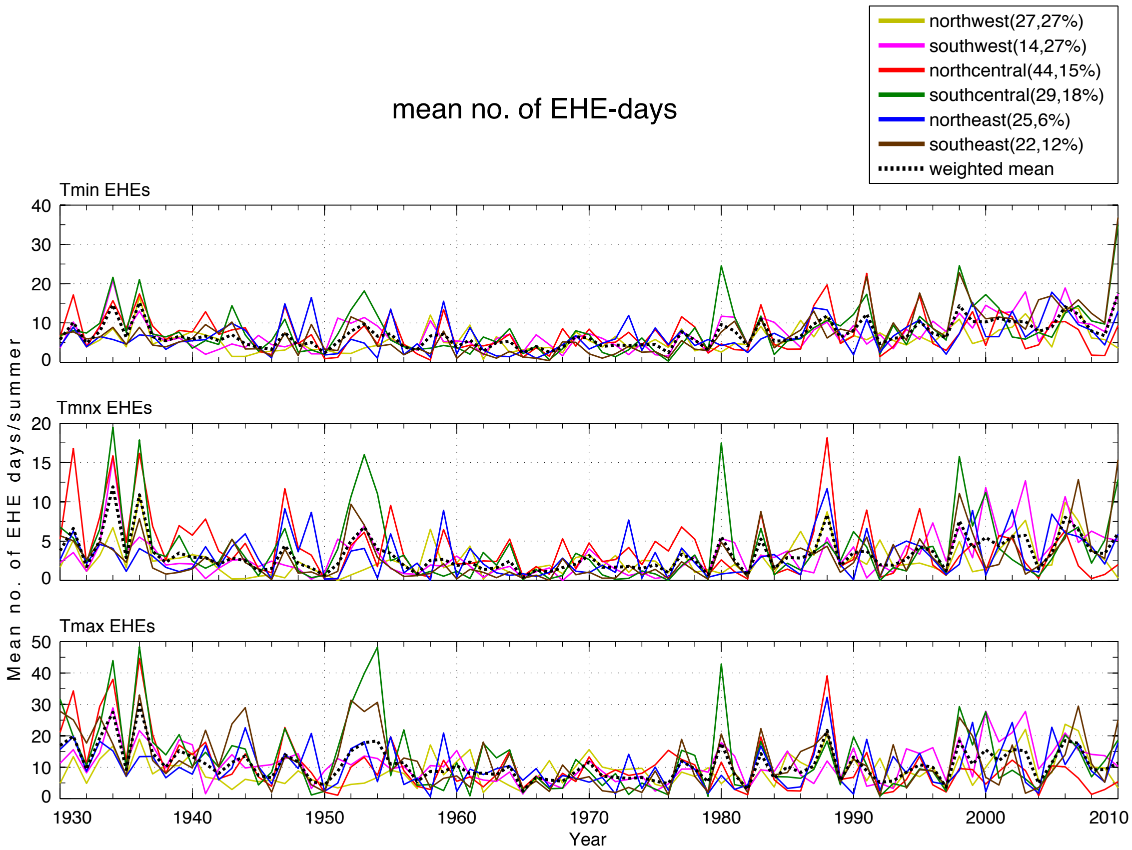

Figure 2.3. Visual display of the CONUS average EHE characteristics trends. Trends are represented by bars and arranged by EHE characteristics vertically and EHE type and time period horizontally. Increases are in red and decreases in blue.

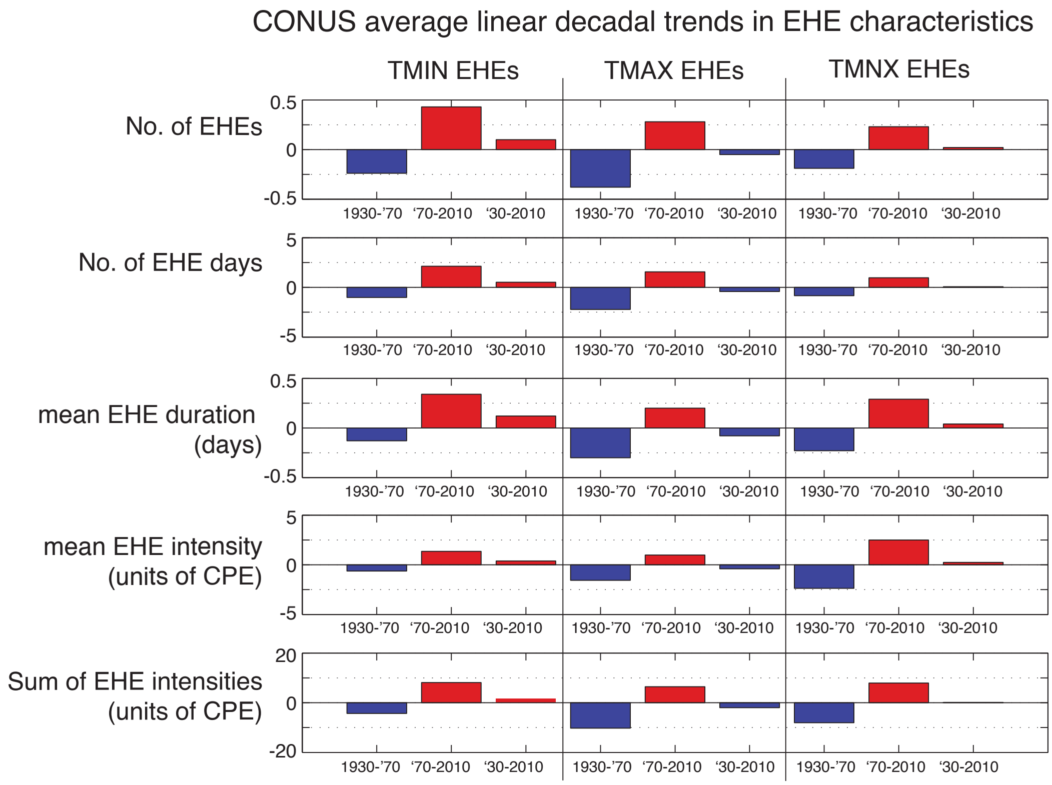

Figure 2.4. The decadal trends of EHE characteristics in Tmnx type EHEs at each station during the 1970-2010 period. The trend significance (alpha=0.10) is indicated by symbol shading. The graduated symbol groupings are based on standard deviations away from the zero value, and are different for each map. The trends in the number of EHEs, the number of EHE days, the mean EHE duration, the mean EHE intensity and the sum of all EHEs intensities per year are listed respectively from *a* to *e*.

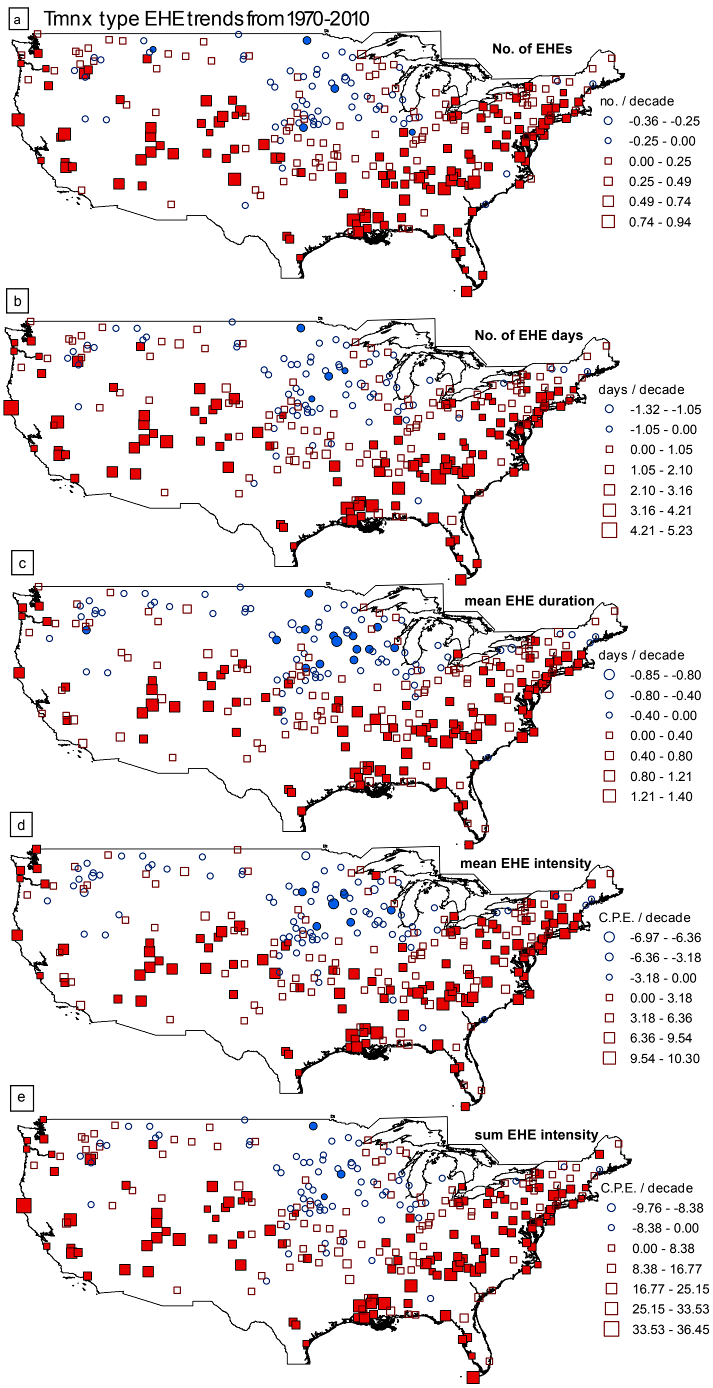

Figure 2.5. The decadal trends of EHE characteristics in Tmnx type EHEs at each station during the 1930-1970 period. All else is the same as figure 2.4.

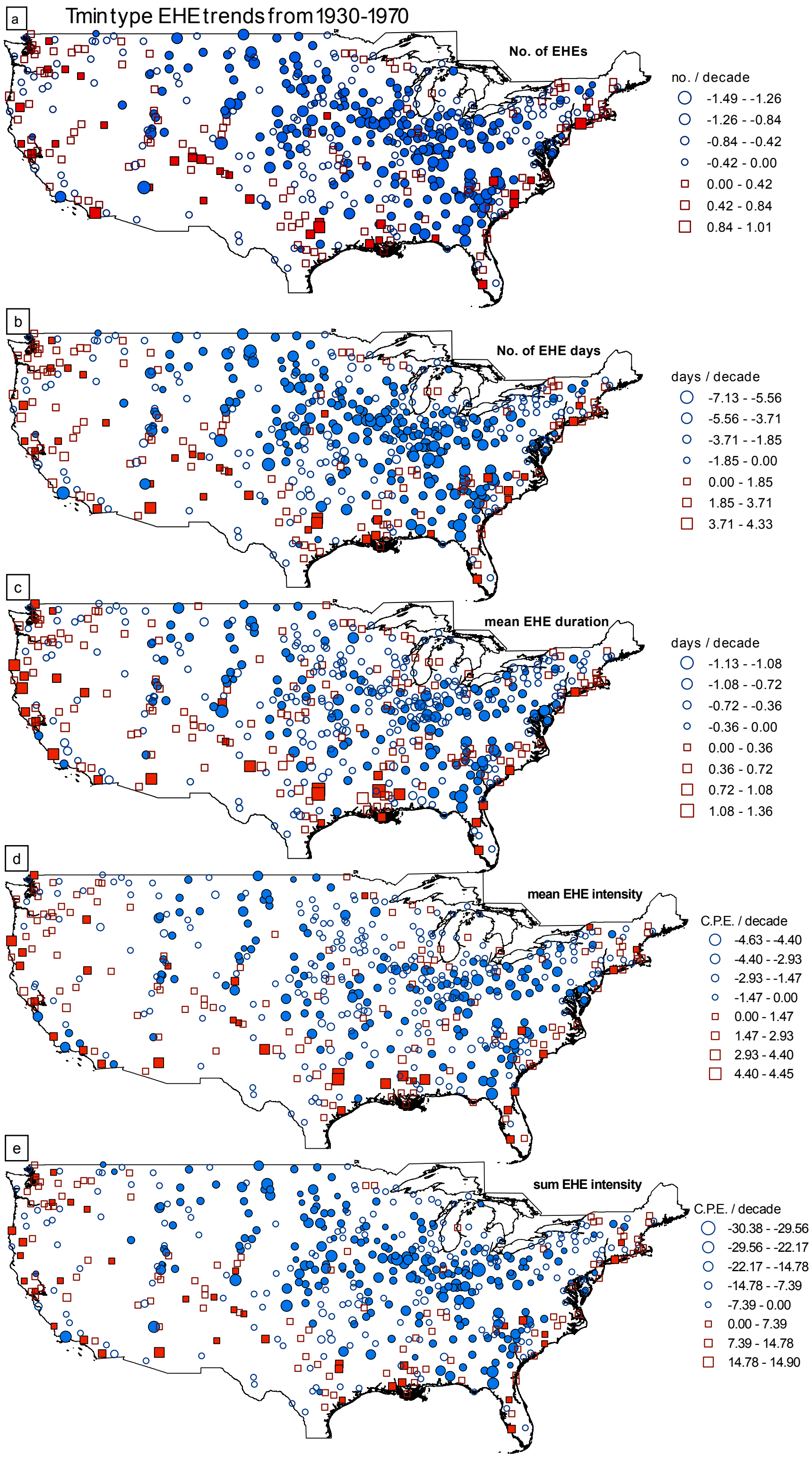

Figure 2.6. The decadal trends of EHE characteristics in Tmax type EHEs at each station during the 1930-1970 period. All else is the same as figure 2.4.

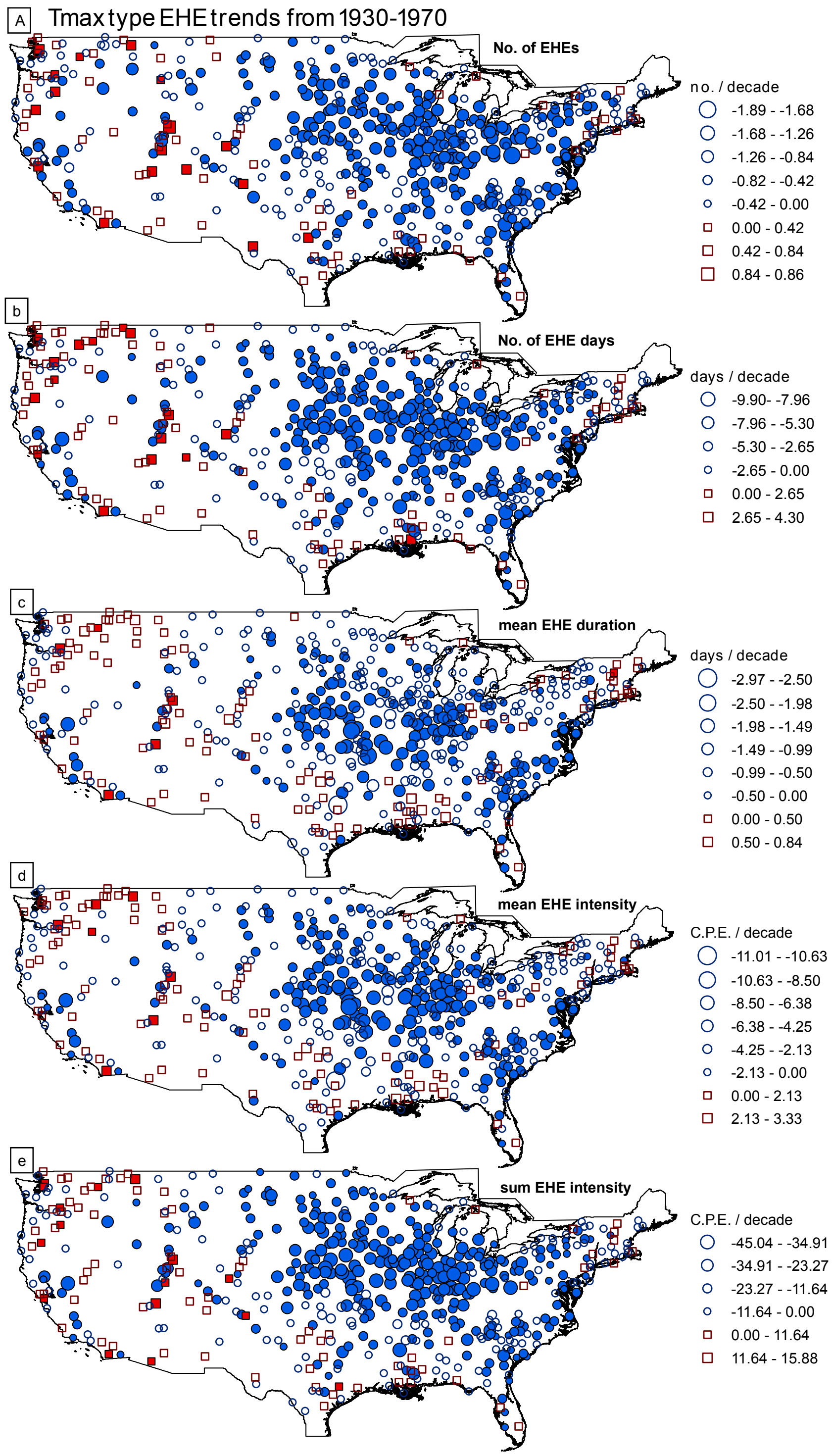

Figure 2.7. The decadal trends of EHE characteristics in Tmnx type EHEs at each station during the 1930-1970 period. All else is the same as figure 2.4.

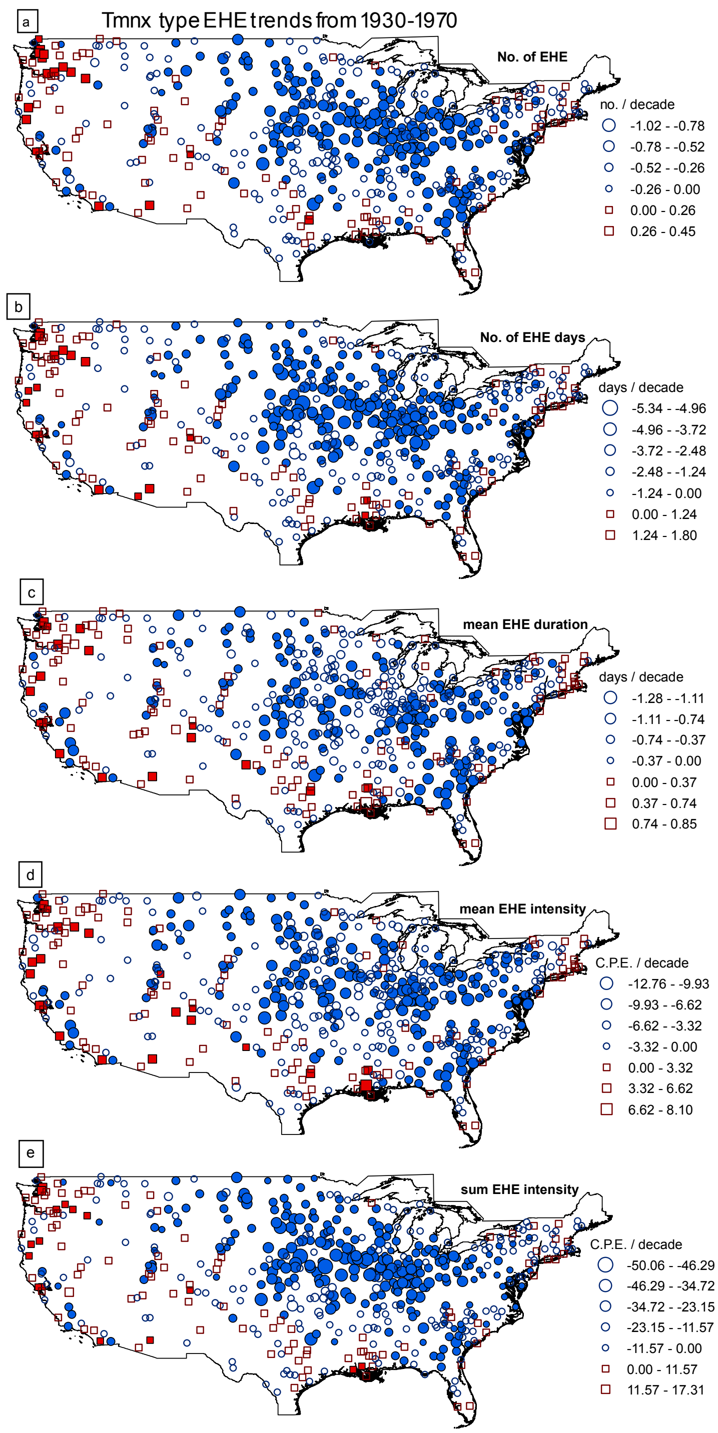

Figure 2.8. The decadal trends of EHE characteristics in Tmin type EHEs at each station during the 1970-2010 period. All else is the same as figure 2.4.

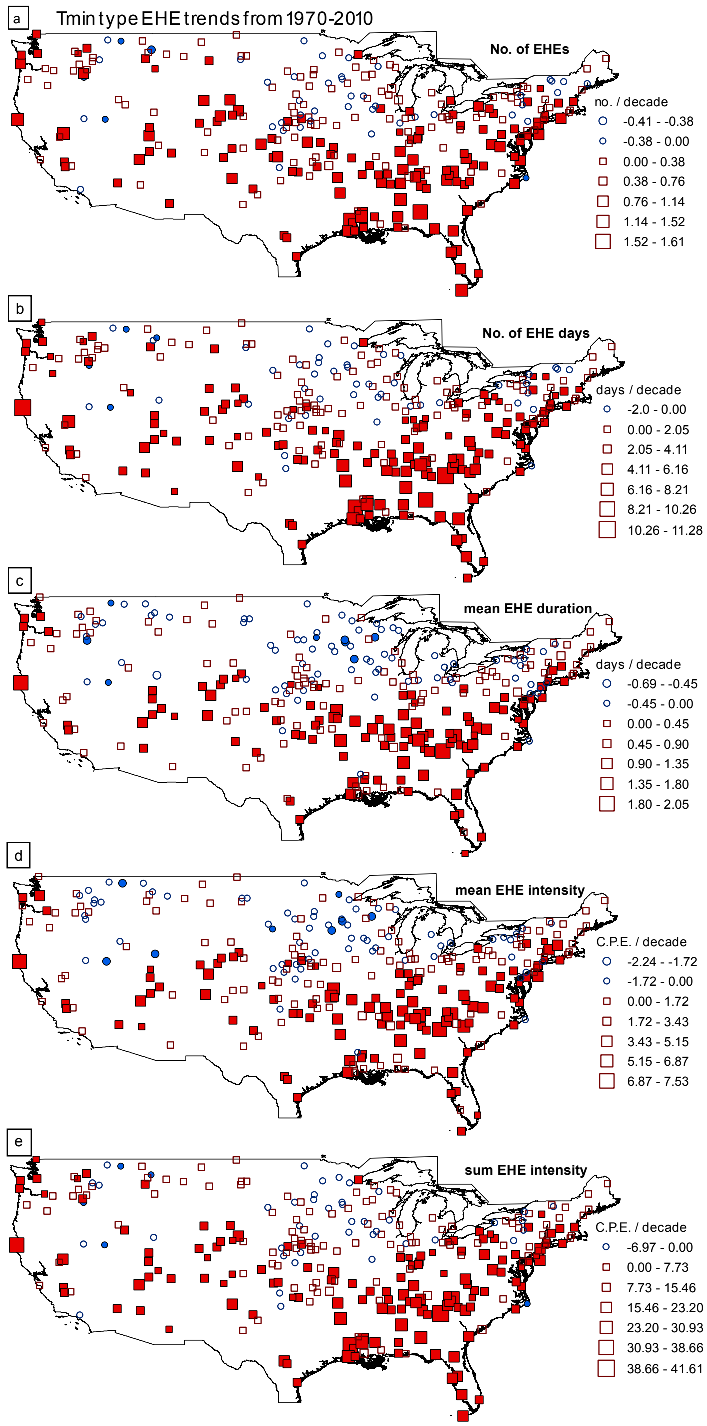

Figure 2.9. The decadal trends of EHE characteristics in Tmax type EHEs at each station during the 1970-2010 period. All else is the same as figure 2.4.

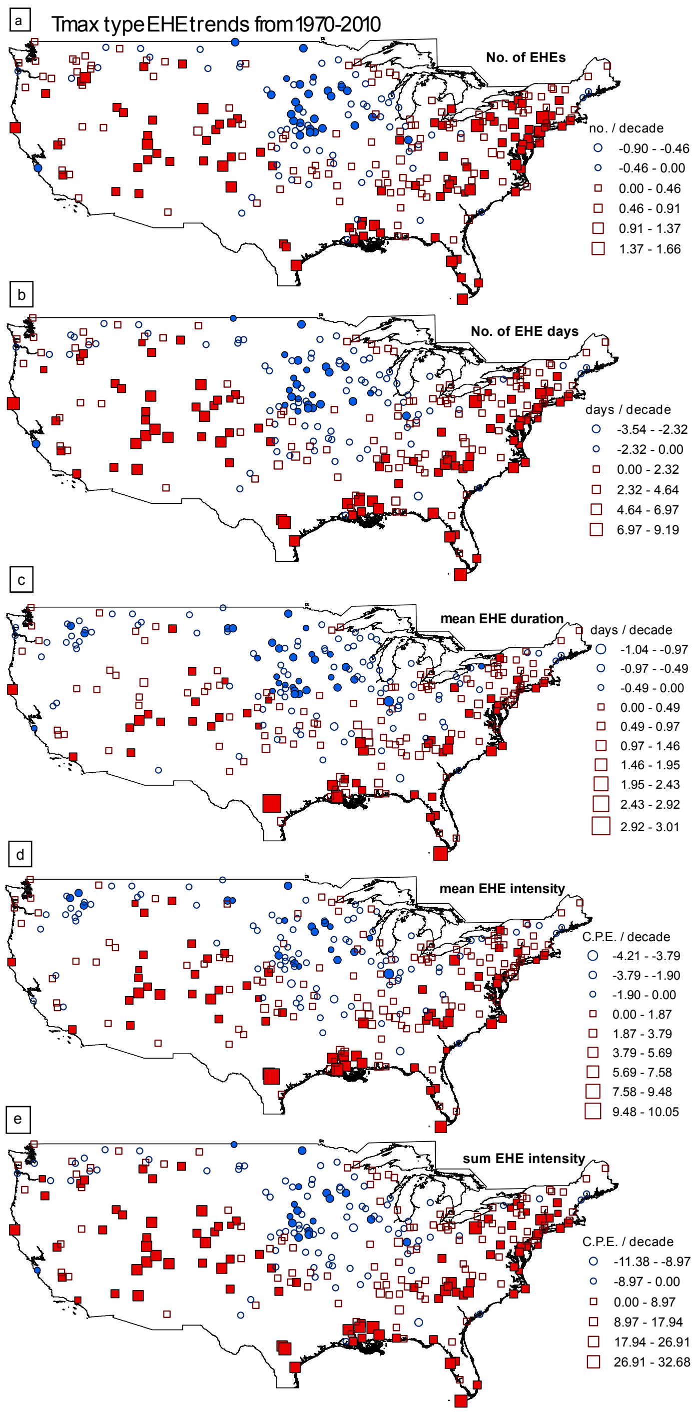

Figure 2.10. The decadal trends of EHE characteristics in Tmin type EHEs at each station during the 1930-2010 period. All else is the same as figure 2.4.

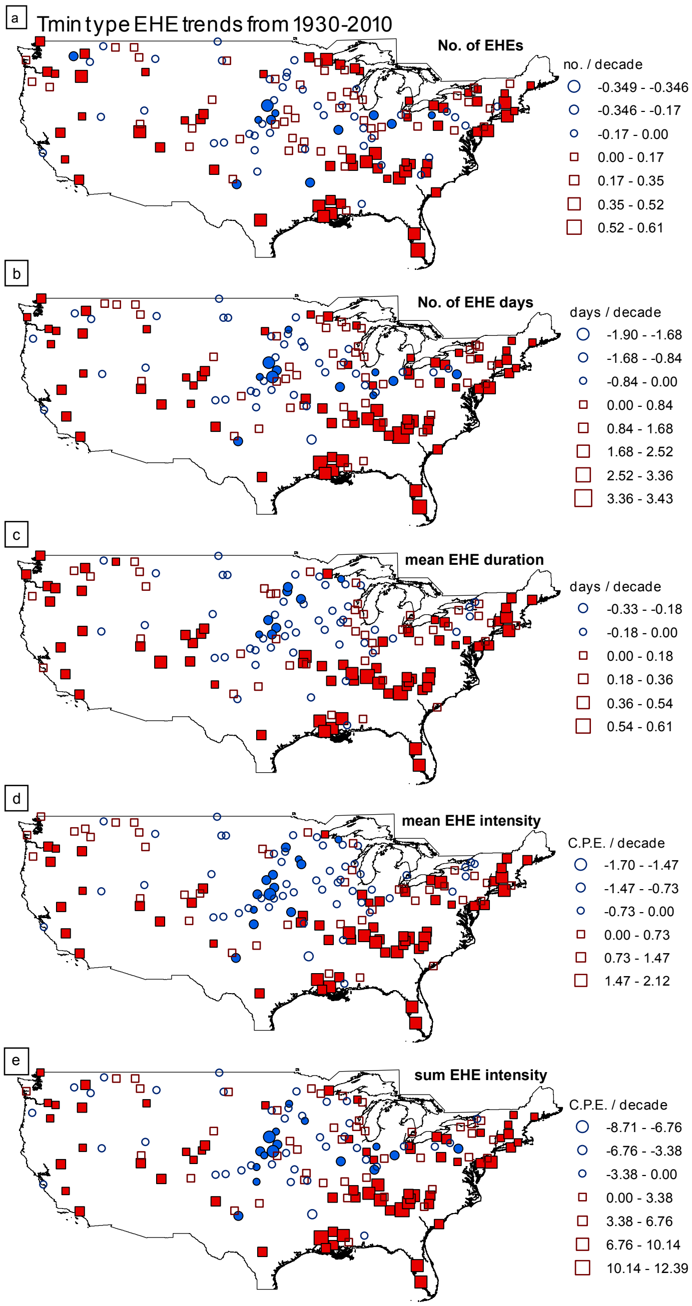

Figure 2.11. The decadal trends of EHE characteristics in Tmax type EHEs at each station during the 1930-2010 period. All else is the same as figure 2.4.

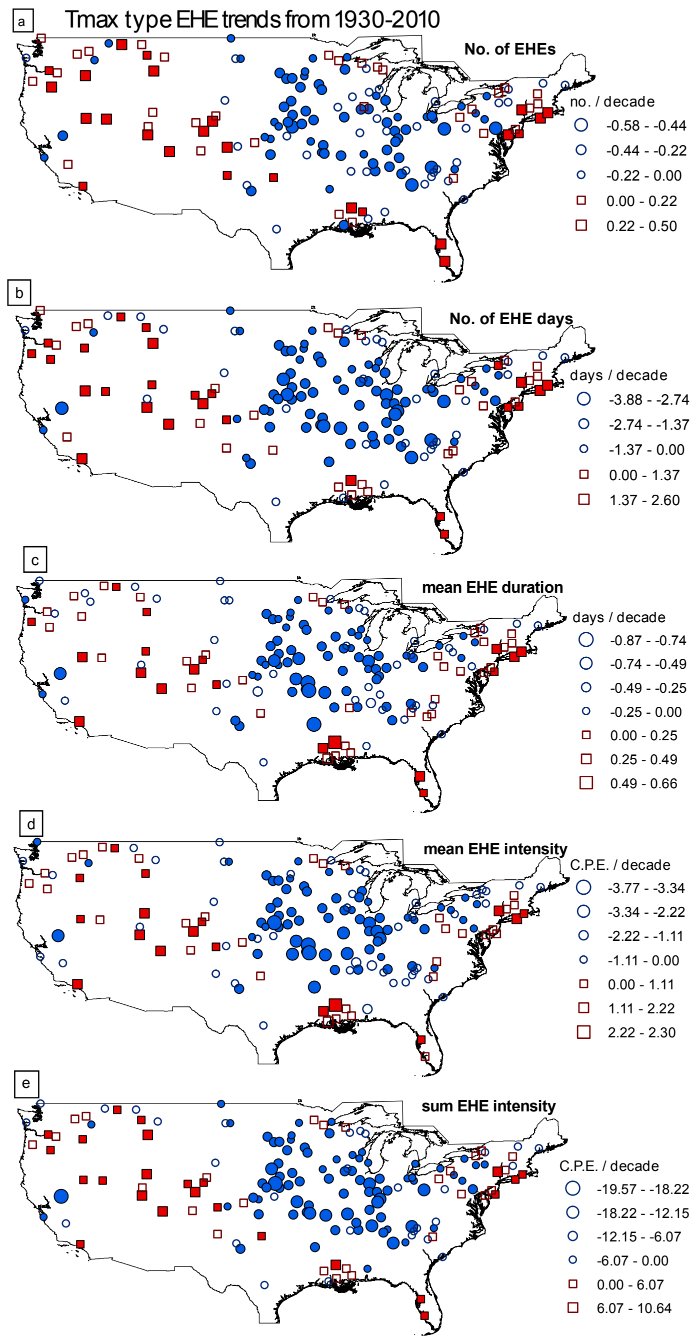

Figure 2.12. The decadal trends of EHE characteristics in Tmnx type EHEs at each station during the 1930-2010 period. All else is the same as figure 2.4.

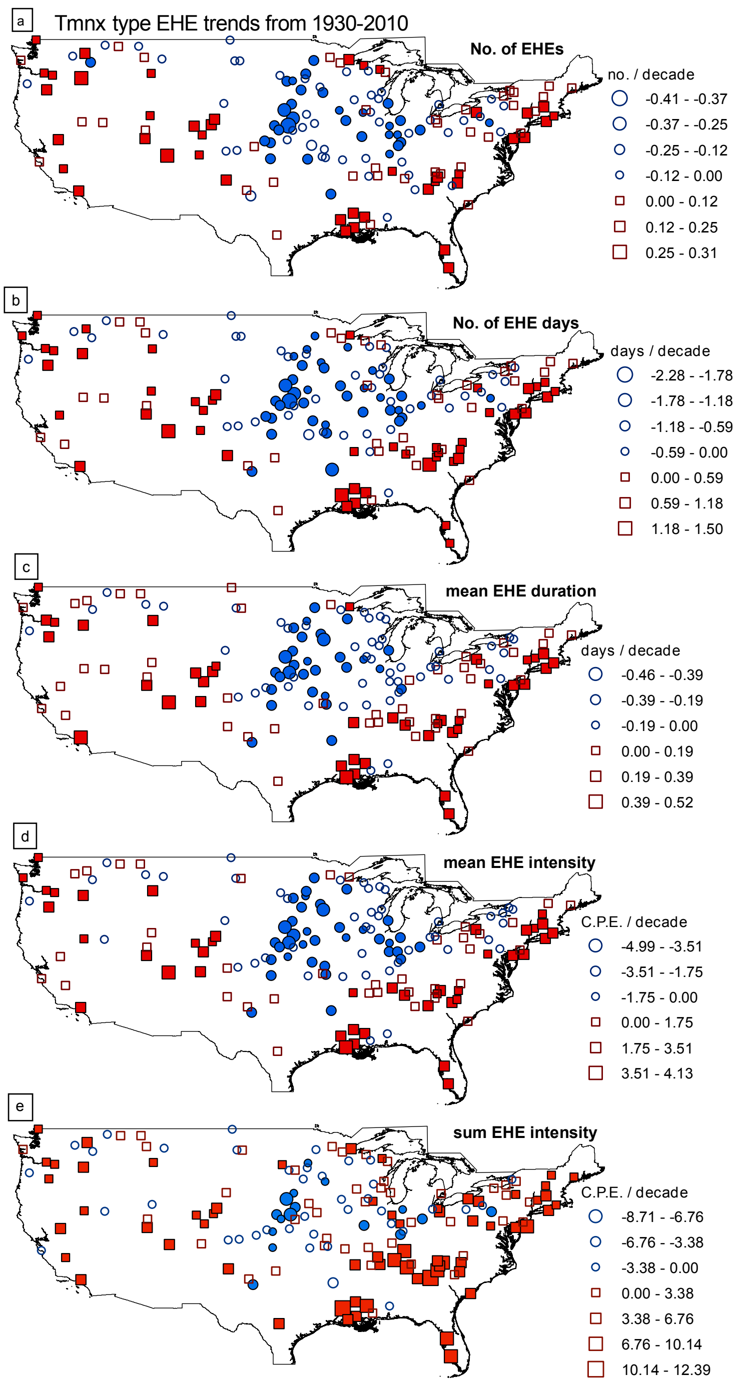

Figure 2.13. The decadal trends in the number of EHEs per summer at each station during the 1930-1970 period. All three types of EHEs are given, and the symbology is the same as figure 2.4.

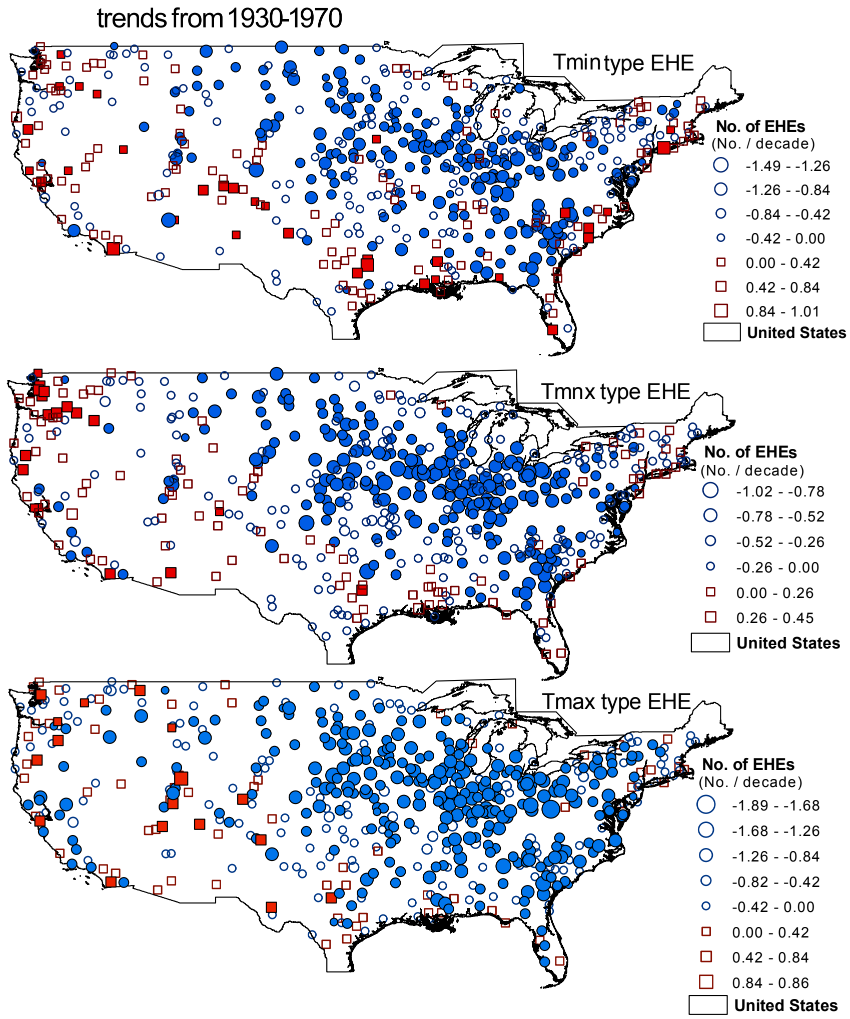

Figure 2.14. The decadal trends in the mean EHE duration at each station during the 1930-1970 period. All three types of EHEs are given, and the symbology is the same as figure 2.4.

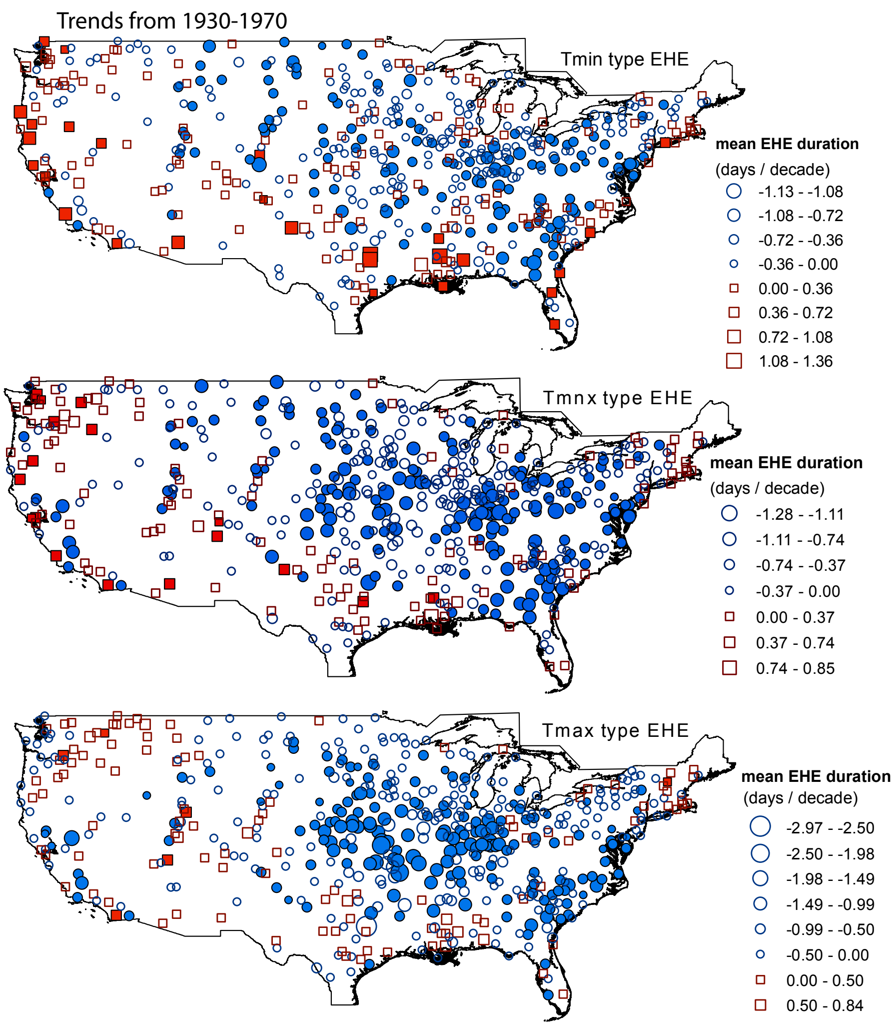

Figure 2.15. The decadal trends in the mean EHE duration at each station during the 1970-2010 period. All three types of EHEs are given, and the symbology is the same as figure 2.4.

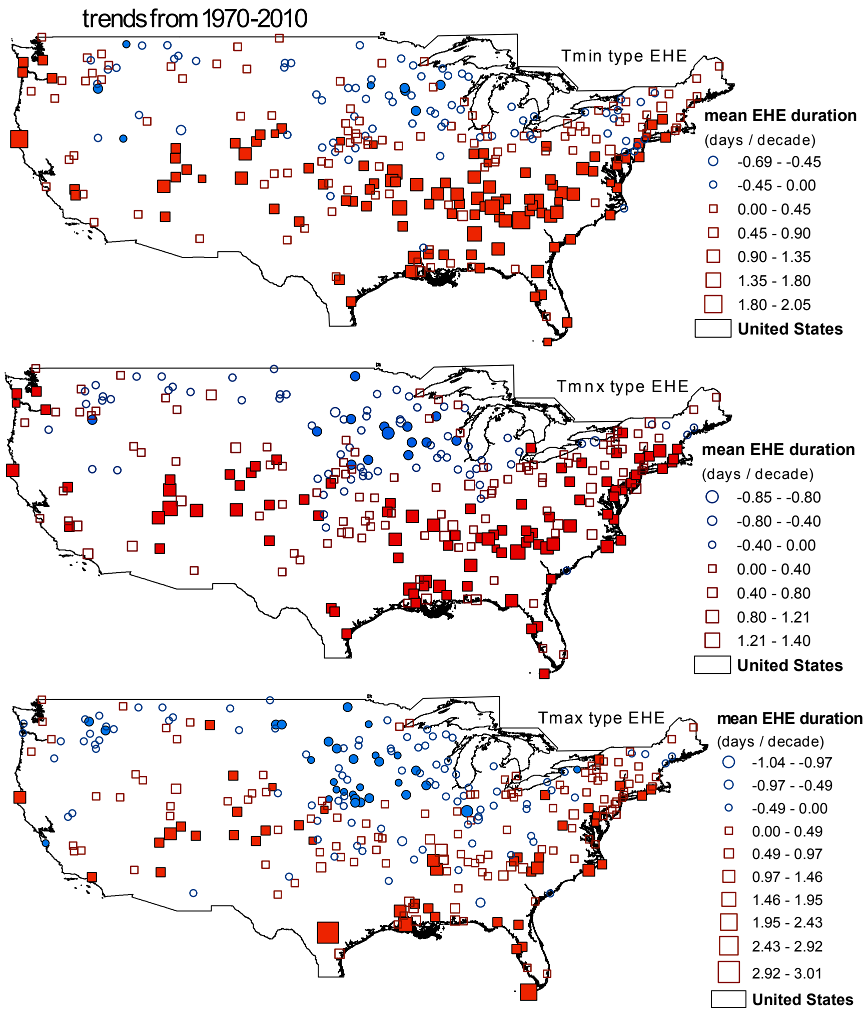

Figure 2.16. The decadal trends in the mean EHE duration at each station during the 1930-2010 period. All three types of EHEs are given, and the symbology is the same as figure 2.4.

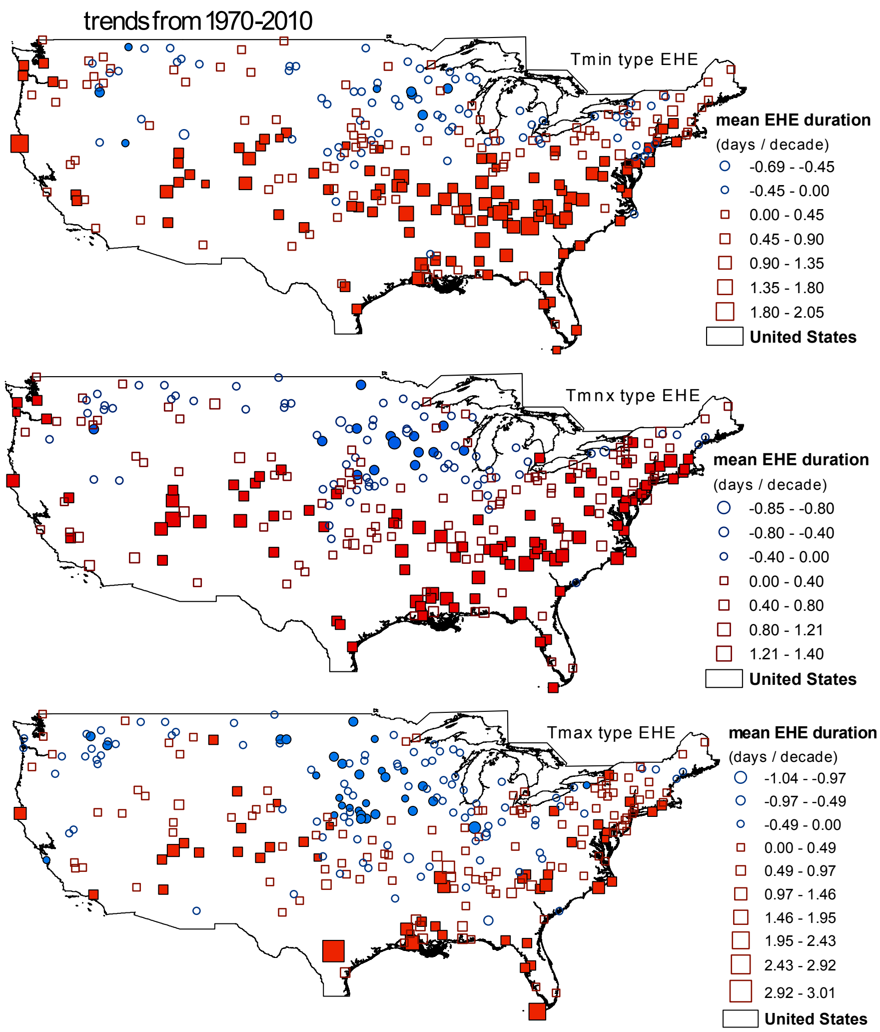

Figure 2.17. The decadal trends in the mean EHE intensity at each station during the 1930-2010 period. All three types of EHEs are given, and the symbology is the same as figure 2.4.

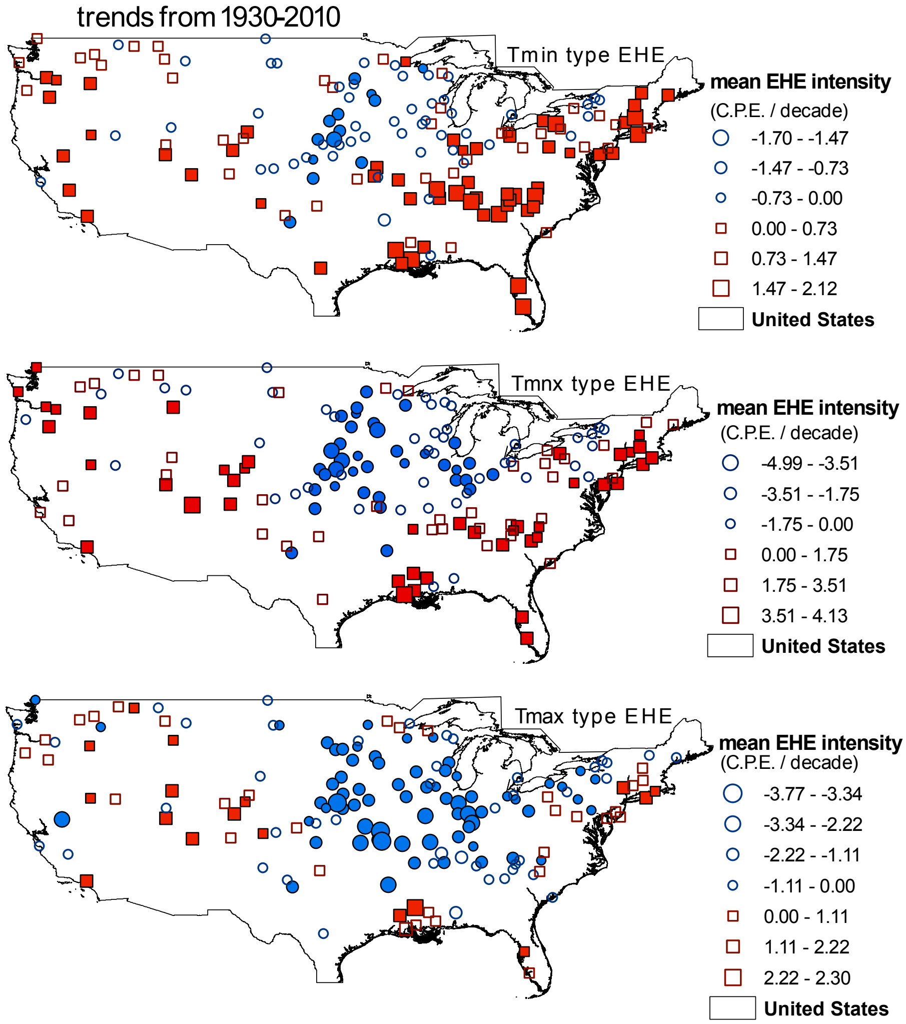

Figure 2.18. Trends in the number of EHE-days during the 1930-1970 period for different EHE types and intra-seasonal focus. All three types of EHEs are given, with three seasonal focuses and the symbology is the same as figure 2.4.