Distribution Agreement

Definition of Extreme Heat Events: Evaluation and Comparison

By

Shengnan Han
Master of Public Health

Global Environmental Health

Jeremy Hess, MD, MPH
Committee Chair

Paige Tolbert, PhD
Committee Member

Definition of Extreme Heat Events: Evaluation and Comparison

By

Shengnan Han

B.S.
Tianjin University
2009

Thesis Committee Chair: Jeremy Hess, MD, MPH

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Global Environmental Health
2015

Abstract

Definition of Extreme Heat Events: Evaluation and Comparison
By Shengnan Han

Purpose: Since there are diversity of communities engaged in heat events monitoring and research, numerous definitions are currently being used to identify extreme heat events from a variety of perspectives. However, the various definitions can also lead to some miscommunication and, from a public health perspective, differences in exposure characterization. So the difference between different definitions should be evaluated.
Methods: in order to illustrate how different the definitions can be, the geographic patterns and trends for 18 cities in the nine U.S. climate regions are studied, and three definitions are selected to be compared and evaluated: the U.S. National Weather Service Heat Index, the Australia Bureau of Meteorology Excess Heat Factor, and Spatial Synoptic Classification Weather Type. Two famous heat waves, 1995 Chicago heat wave and 2013 Phoenix heat wave, are selected as cases to compare the definitions in detail.
Results: the results show that although all the three definitions can recognize extreme heat events, there are still a lot of differences between them. In this paper, the measures' variances, spatial and temporal differences, and their ability to identify the severity of extreme heat event are discussed. Each definition is evaluated by the same metrics.
Conclusions: the differences between the three definitions are significant. Every definition has its flaws, so researcher should pay extra attention when they measure the relationship between extreme heat events and health outcomes.
Keywords: heatwave, extreme heat, definition

Definition of Extreme Heat Events: Evaluation and Comparison

By

Shengnan Han

B.S.
Tianjin University
2009

Thesis Committee Chair: Jeremy Hess, MD, MPH

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Global Environmental Health
2015

Table of Contents

1. Background

An extreme heat event, also known as a heat wave, as a concept is not synonymous with high temperature. While there is no generally agreed-upon definition of extreme heat events, most of the definitions describe an extreme heat event as a period of time with extremely hot weather that can potentially harm human health (Hayhoe, K., et al., 2004). Extreme heat events are regarded by both the National Weather Service and the Centers for Disease Control and Prevention as a major cause of weather-related fatalities annually in the United States, more than hurricanes, lightning, tornadoes, floods and earthquakes (Luber & McGeehin, 2008). With climate change, it is very likely the extreme heat events will increase in severity, frequency, scope, and duration (Robinson, 2001). As the mean global temperature keeps rising, more attention is being paid to the social and environmental effects of heat extremes. Temperature change has both direct and indirect impacts on human health. The direct impacts on human health can arise from increased frequency and severity of extreme weather such as heat waves, while the indirect impacts on human health can arise from the secondary events caused by environmental and social change, such as air pollution, vector-borne diseases, and food supplies (Xun, et al., 2010).

There have been detailed analyses of individual extreme heat events and their impacts. For example, according to the Pubmed, there are more than 60 papers studying the 1995 heat wave in Chicago and more than 300 papers studying the 2003 heat wave in Europe. After that, the projection of future extreme heat events caused by climate change has become a popular topic. In the last 20 years, more studies have focused on identifying health impact, developing early warning systems, and reviewing response plans, but few studies address what an extreme heat event actually is. Trends in maximum temperature and minimum temperature of a certain day are related to, but not necessarily synonymous with, the trends in extreme heat events (Smith, 2012). Work still needs to be done to accurately identify extreme heat events.

There are a lot of definitions for extreme heat events. For example, the spatial synoptic climatological index system was developed by Davis and Kalkstein to define different weather types (Davis & Kalkstein, 1990), where the moist tropical class and the dry tropical class are considered heat waves. The National Weather Service has created a heat index chart. The chart recognizes four classes of likelihood of heat disorders with prolonged exposure or strenuous activity by Rothfusz's equation (Rothfusz & Headquarters, 1990). The Australian Bureau of Meteorology implemented a practical indicator system of heat waves (Nairn, 2009), which incorporates the weather prior to the events and have desirable ability to identify the peak of the extreme heat events. Since there are diversity of communities engaged in heat events monitoring and research, numerous definitions are currently being used to identify extreme heat events from a diversity of perspectives. Health researchers are interested in the aspects of heat waves that are most relevant to human well-being. In contrast, climatologists, who are primarily interested in the evolution of weather statistics and evaluating evidence on climate change, prefer to use definitions that include the probability of exceedance in certain relatively straightforward metrics, usually defined relative to a long term mean.

Because of the diversity of stake holders involved in the study of heat waves, researchers from different disciplinary backgrounds are able to collaborate to evaluate health risks posed by climate variability and change. However, the various definitions can also lead to some miscommunication. Many of these definitions are not universally applicable across the United States; more study is needed to decide how effective these different definitions are when used to assess public health during extreme heat events. It is valuable to compare and evaluate the definitions. In this paper, the geographic patterns and trends for 18 cities in the nine U.S. climate regions (Thomas & Walter, 1984) are analyzed. Three definitions are chosen: one from the U.S. National Weather Service, one from the Australia Bureau of Meteorology, and another one called Spatial Synoptic Classification. The reason why these three definitions are chosen is that they are commonly used and distinct from each other. The objective is not only to compare how the selection of definitions will influence the conclusions regarding the observed frequency of extreme heat events but also to determine their strengths and weaknesses.

2. Method

2.1 Data

The National Climatic Data Center identifies nine climatically consistent regions within the contiguous United States through climate analysis: Central, East North Central, Northeast, Northwest, South, Southeast, Southwest, West, and West North Central. The classified regions are shown in figure 1. Two cities from each region were selected, thus there are 18 cities in the sample: Charlotte, Chicago, Denver, Detroit, Houston, Indianapolis, Jacksonville, Lincoln, Los Angeles, Milwaukee, New York, Omaha, Philadelphia, Phoenix, Portland, San Antonio, San Diego, Seattle.

John Nairn's definition of extreme heat events requires at least 30 years to calculate a long term climate reference value, so data from 1/1/1984 to 12/31/2014 was selected. The daily maximum and minimum temperatures were gathered from the Global Historical Climatology Network (GHCN) database, and hourly computed relative humidity was gathered from the NCDC Climate Data Online database. SSC data was provided by Kent State University, Department of

Geography.

2.2 JN definition

The Centre for Australian Weather and Climate Research and the Australian Government's Bureau of Meteorology developed a practical system for indicating heat waves. The index combines both the long-term temperature anomaly characterized by each location's own climatology of heat and the short-term temperature anomaly measuring recent thermal acclimatization.

In this definition, the daily mean temperature T is defined as:

$$\mathrm{T} = (\mathrm{T}_{max} + \mathrm{T}_{min})/2 \tag{1}$$

T_{max} is the maximum temperature of a certain day, and T_{min} is the minimum temperature of a certain day.

A concept called excess heat is used to describe an unusually high daytime temperature that is not sufficiently released in the night since the overnight temperature is also high. The excess heat index EHI_{sig} is defined as:

$$\mathrm{EHI}_{sig} = \frac{T_i + T_{i+1} + T_{i+2}}{3} - T_{95} \tag{2}$$

T_i is the mean temperature of a certain day, i. T_{i+1} is the mean temperature of the day after day i. T_{95} is the 95th percentile of daily mean temperature (T_i) in the study period, 1984-2014.

A concept called heat stress is used to describe a period if its temperature is warmer than the recent past. The index EHI_{accl} is defined as:

$$\mathrm{EHI}_{accl} = \frac{T_i + T_{i+1} + T_{i+2}}{3} - \frac{T_{i-1} + \cdots + T_{i-30}}{30} \tag{3}$$

i has the same meaning as it does in equation (2).

Then, the third concept is excess heat factor, which combines the effect of excess heat and heat stress together. A heat event exists when EHF is positive.

$$\mathrm{EHF} = \mathrm{EHI}_{sig} \times \max(1, EHI_{accl}) \tag{4}$$

The unit of EHF is ℃^{2}.

A threshold of EHF should be set to identify extreme heat events, using the probability of peaks over threshold theory (Katz, 2010). Extreme heat events can be characterized as a heavy-tail distribution, which is well described by the Pareto distribution:

$$\Pr(X > x) = 1 - (1 - kx/b)^{1/k} \quad (5)$$

In a more practical way, the 80/20 rule is used. The rule generally means "the last 20% cumulative distribution contributes to the 80% of the distribution function". In Nairn's study, 85% of the cumulative distribution is set as the threshold in order to reduce the effect of sampling variability (Nairn, 2013). Thus, any day whose EHF is higher than EHF_{85} should be classified as an extreme heat event.

2.3 SSC definition

Kalkstein, Greene and other climatologists developed a weather type classification scheme named Spatial Synoptic Classification system, or SSC (Kalkstein, et al., 1996). At first, the system can only classify winter and summer weather type across the contiguous United States. But the system was redeveloped by different scientists. The study area is expend to US, Canada, and the western Europe (Bower, et al., 2007). The system is also able to classify days year-round (Sheridan, 2002). The version of SSC used in this paper is based on surface observations at individual stations. Temperature, dew point, wind, pressure and cloud cover are incorporated into the model.

There are six categories in the model: dry polar (DP), dry moderate (DM), dry tropical (DT), moist polar (MP), moist moderate (MM), moist tropical (MT, include two subsets: MT+ and MT++). Transitional days (TR) are defined as days in which one weather type yields to another. The weather-type characteristics vary from station to station and day to day.

Among those weather types, DT and MT, including MT+ and MT++, are chosen as extreme heat events. DT represents the hottest and driest days at a certain location. MT indicates the day is warm and very humid. MT+, a subset of MT, accounts for the days where both morning and afternoon temperatures are above seed day means. MT++, a subset of MT+, indicates the morning and afternoon apparent temperatures are at least one standard deviation higher than seed day means. The seed day mean is the mean value of typical days in each type, and the typical days were selected by each ground station from different times of the year.

2.4 NWS definition

The National Weather Service has created a heat index chart, which is the figure 2. The heat index shows how it actually feels when relative humidity is factored in with the air temperatures.

The chart is derived from Rothfusz's equation (Rothfusz, 1990). The regression equation of Rothfusz is:

$$\mathrm{HI} = -42.379 + 2.04901523 \times \mathrm{T} + 10.14333127 \times \mathrm{RH} - 0.22475541 \times \mathrm{T} \times \mathrm{RH} - 0.00683783 \times \mathrm{T}^2 - 0.05481717 \times \mathrm{RH}^2 + 0.00122874 \times \mathrm{T}^2 \times \mathrm{RH} + 0.00085282 \times \mathrm{T} \times \mathrm{RH}^2 - 0.0000199 \times \mathrm{T}^2 \times \mathrm{RH}^2 \tag{6}$$

A simplified formula of this is:

$$\mathrm{HI} = 0.5 \times \{\mathrm{T} + 61.0 + [(\mathrm{T} - 68.0) \times 1.2] + (\mathrm{RH} \times 0.094)\} \tag{7}$$

HI is heat index, T is Fahrenheit degree, and RH is relative humidity.

There are four levels of the likelihood of heat disorders with prolonged exposure or strenuous activity: caution, extreme caution, danger, and extreme danger. In this study, the average temperature and relative humidity are used to decide the class of a day. Extreme danger is considered as an extreme heat event.

The heat index values were devised for shady, light wind conditions, thus exposure to full sunshine can increase heat index values by up to 15°F. At the same time, strong winds, particularly with very hot, dry air, can be extremely hazardous. Due to the lack of available data, this study does not include the effect of sunshine and wind.

2.5 Comparison and Evaluation

All the days between 1/1/1984 and 12/31/2014 will be given 3 binary variables in each of three extreme heat indicator categories: JN, SSC, and NWS for the cities being analyzed. A 1 means day i is classified as an extreme heat event by the definition, while a 0 means day i is not. The proportion of extreme heat events among all the days is standardized, and then given a variance analysis. The equation of standardization is:

$$\text{New} = \frac{\text{Old} - \text{Old}_{min}}{(\text{Old}_{max} - \text{Old}_{min})} \tag{8}$$

Even though the total numbers are similar by some definitions, the specific date may not be the same. To illustrate the different temporal distribution of extreme heat events, a case study of specific location is essential. Since NWS-HI and JN-EHF are continuous, time-series plots were generated to illustrate the difference.

The three definitions will be evaluated from following aspects: 1. Is it easy to understand? 2. Is it measured with readily available data? 3. Is it easily visualized? 4. Is it predictable with reasonable accuracy? 5. Is it useful as an indicator of impact? 6. And is there any other flaws for that definition?

3. Result

The total numbers of extreme heat events for each city by different definitions are summarized by table 1. The table 1 shows the total numbers of extreme heat events for each city by different definitions. For example, Charlotte experienced 72 days classified as extreme heat events out of 11,323 days by the JN definition, 3,616 days classified as extreme heat events out of 11,323 days by the SSC definition, and 62 days classified as extreme heat events out of 11,323 days by the NWS definition.

After standardization, the standard deviation for the JN definition is 0.24, the standard deviation for the SSC definition is 0.28, and the standard deviation for the NWS definition is 0.28.

3.1 July 1995 heatwave in Chicago

The 1995 Chicago heat wave led to at least 700 deaths from July 12th to July 17th (Semenza, 1996). Figure 3 shows that, based on both the JN and NWS definitions, the highest index value occurred on July 13th. According to the table 2, JN definition recognizes that the extreme heat event lasted four days. The NWS definition recognizes that that particular event lasted two days. The SSC definition recognizes a seven day period as an extreme event.

The correlation between EHF-JN and HI-NWS is 0.72336, and the p value is less than 0.0001.

3.2 Summer 2013 heatwave in Phoenix

The 2013 heat wave in the southwestern United States occurred in an area that spanned California to Arizona, lasting approximately four days from late June to early July. Figure 4 shows that both the JN and the NWS definitions identify June 29th as the peak of this extreme heat event. According to Table 3, JN recognizes a two-day extreme event, NWS recognizes a five day event, and SSC recognizes a period around seven days.

The correlation between EHF-JN and HI-NWS is 0.77742, the p value is less than 0.0001.

4. Discussion

From the tables and figures above, the main characteristics for each definition can be summarized.

4.1 Variance

For Table 1, the variances of the proportions show that the JN definition is smaller than the other two definitions. This characteristic is due to the different equations, especially how the long term climate reference value is defined. The equations used in the JN definition reflects the hottest days within a certain period, for a certain location. It identifies a proportion of days with the highest temperature within that period as extreme heat events. The proportion is steady because the threshold is locally set. The other two definitions have a uniform threshold for all the locations. For example, no matter where the city is, the SSC definition will classify DT and MT, including MT+ and MT++, as extreme heat events.

4.2 Spatial Difference

The value of the excess heat factor for the JN definition indicates how hot a day is compared to all the days in the study period. The value of the heat index for the NWS definition indicates how hot that day is compared to the uniformly set threshold.

The SSC and NWS definitions reflect spatial difference, which may be useful for studying a relatively large area. In contrast, the JN definition is highly local, which may be more useful for developing a local warning or surveillance system.

4.3 Temporal Difference

From Figures 3 and 4, the trends of EHF-JN and HI-NWS are visually similar. The correlation between these two indexes is strong, which also supports that the trends are similar.

But there is a large difference between the SSC definition and the other two definitions. SSC classifies a significantly high number of days as extreme heat events, which is not comparable to the others.

4.4 Severity

The values of each of the two indicators show different judgment of how serious an extreme heat event is. For example, the peak EHF value of the Chicago heat wave is above 50, and the peak EHF value of the Phoenix heat wave is around 15. Thus, by the JN definition, the severity of the 1995 extreme heat event was worse for Chicago residents than the severity of the 2013 extreme heat event was for Phoenix residents. However, the peak HI value of the 1995 Chicago extreme heat event is below 110, while the peak EHF value of the 2013 Phoenix extreme heat event is above 120. Thus, based on these values, the severity of the 2013 Phoenix heat event is worse than the severity of 1995 Chicago extreme heat event. In this case, the judgments by the two indicators are contradictory.

4.5 Evaluation

Evaluation is based on six evaluation terms, which is shown in table 4.

The three definitions all have binary indicators to show if there is an extreme heat event on a certain day. But considering the equation to obtain the value of the indicators, the NWS definition is the simplest to apply.

The JN definition only needs the daily temperatures for all the outcomes to be computed, the NWS definition needs both temperature and humidity, and the SSC definition needs temperature, dew point, wind, pressure and cloud cover. Thus, the data for the JN definition is the easiest to obtain.

For the JN and NWS definitions, there are multiple ways to visualize the results. For example, using the JN definition, we can show the binary variable by a table, the excess heat factor by time series plot, or an interpolated spatial map for the excess heat factor.

The accuracy of recognition is always a concern. Almost all the three definitions successfully recognize the well-known heat waves that happened in the past. However, there are also some days that are classified as extreme heat events by one or two definitions but not the other. If the projection of temperature and other data are available, all the definitions can identify extreme heat events.

One of the major roles of the three definitions is to serve as an indicator of impact. The SSC definition does not perform well here. Since there are too many days classified as extreme heat events, it will overpredict and thus underperform as an indicator. If we apply SSC for health warnings, it could also lead to message fatigue, wherein people begin to ignore warnings when they do not see significant risk associated with them.

Every definition has its flaws. For example, the JN definition only considers air temperatures, which is not the same as apparent temperature. At the same time, the NWS definition does not combine the temporal effects of the days before the tested day. The SSC definition, as mentioned before, classifies too many days as extreme heat effects, which will weaken the effectiveness of indicating the impact of a heat wave.

4.6 Limitation

One main limitation of this study is the selection criteria of extreme heat events definitions. The three selected definitions are commonly used, but there are other commonly used definitions. Future studies should consider compare and evaluate more definitions, which might reveal some patterns of the difference.

The sample size is sufficient to support this study, but if more cities and longer period data are included, more patterns can be revealed. For example, if more cities are sampled, a local indicators of spatial autocorrelation model can be applied to determine if the cities with more extreme heat events are clustered together, and if the cities with less extreme heat events are next to each other (Anselin, 1995), which could be useful to explain why the definitions are different.

The questions in the table 4 are subjective. If an objective measure of impacts can be used in this study, the results and implications will be more persuasive. For example,

if the correlation between emergency room visits/excess mortality and heat index /excess heat factor can be obtained, it should be a more persuasive evidence to evaluate the performances of different extreme heat events definitions.

5. Conclusion

Taken together, all the three definitions have strengths and weaknesses. Most of the inferences from the three definitions are similar, but some of the judgments are inconsistent. There is no doubt that these inconsistencies will affect how researchers interpret the causal relationship between extreme heat events and health outcomes. Thus, future studies should pay extra attention when considering heat wave definitions. And future works can improve the study by evaluating more definitions, increasing the sample size, applying spatial statistic models, and using objective measure of impacts to evaluate the definitions.

References

Hayhoe, K., Cayan, D., Field, C. B., Frumhoff, P. C., Maurer, E. P., Miller, N. L., ... & Verville, J. H. (2004). Emissions pathways, climate change, and impacts on California. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(34), 12422-12427.

Luber, G., & McGeehin, M. (2008). Climate change and extreme heat events. *American journal of preventive medicine*, *35*(5), 429-435.

Robinson, P. J. (2001). On the definition of a heat wave. *Journal of Applied Meteorology*, *40*(4), 762-775.

Xun, W. W., Khan, A. E., Michael, E., & Vineis, P. (2010). Climate change epidemiology: methodological challenges. *International journal of public health*, *55*(2), 85-96.

Smith, T. T., Zaitchik, B. F., & Gohlke, J. M. (2013). Heat waves in the United States: definitions, patterns and trends. *Climatic change*, *118*(3-4), 811-825.

Davis, R. E., & Kalkstein, L. S. (1990). Development of an automated spatial synoptic climatological classification. *International Journal of Climatology*,*10*(8), 769-794.

Rothfusz, L. P., & Headquarters, N. S. R. (1990). The heat index equation (or, more than you ever wanted to know about heat index). *Fort Worth, Texas: National Oceanic and Atmospheric Administration, National Weather Service, Office of Meteorology*, 90-23.

Nairn, J., Fawcett, R., & Ray, D. (2009, November). Defining and predicting Excessive Heat events, a National system. In *Proceedings of the CAWCR modelling workshop: understanding high impact weather* (Vol. 30, pp. 83-86).

Karl, T., & Koss, W. J. (1984). *Regional and National Monthly, Seasonal, and Annual Temperature Weighted by Area, 1895-1983*. National Climatic Data Center.

Katz, R. W. (2010). Statistics of extremes in climate change. *Climatic Change*, *100*(1), 71-76.

Nairn, J. R., & Fawcett, R. G. (2013). *Defining heatwaves: heatwave defined as a heat-impact event servicing all community and business sectors in Australia*. K. A. Day (Ed.). Centre for Australian Weather and Climate Research.

Kalkstein, L. S., Nichols, M. C., Barthel, C. D., & Greene, J. S. (1996). A new spatial synoptic classification: application to air mass analysis. *International Journal of Climatology*, *16*(9), 983-1004.

Bower, D., McGregor, G. R., Hannah, D. M., & Sheridan, S. C. (2007). Development of a spatial synoptic classification scheme for western Europe.*International Journal of Climatology*, *27*(15), 2017-2040.

Sheridan, S. C. (2002). The redevelopment of a weather-type classification scheme for North America. *International Journal of Climatology*, *22*(1), 51-68.

Rothfusz, L. P. (1990). The heat index equation. *National Weather Service Technical Attachment (SR 90–23)*.

Semenza, J. C., Rubin, C. H., Falter, K. H., Selanikio, J. D., Flanders, W. D., Howe, H. L., & Wilhelm, J. L. (1996). Heat-related deaths during the July 1995 heat wave in Chicago. *New England journal of medicine*, *335*(2), 84-90.

Anselin, L. (1995). Local indicators of spatial association—LISA.*Geographical analysis*, *27*(2), 93-115.

Tables & Figures

Figure 1 US Climate Regions

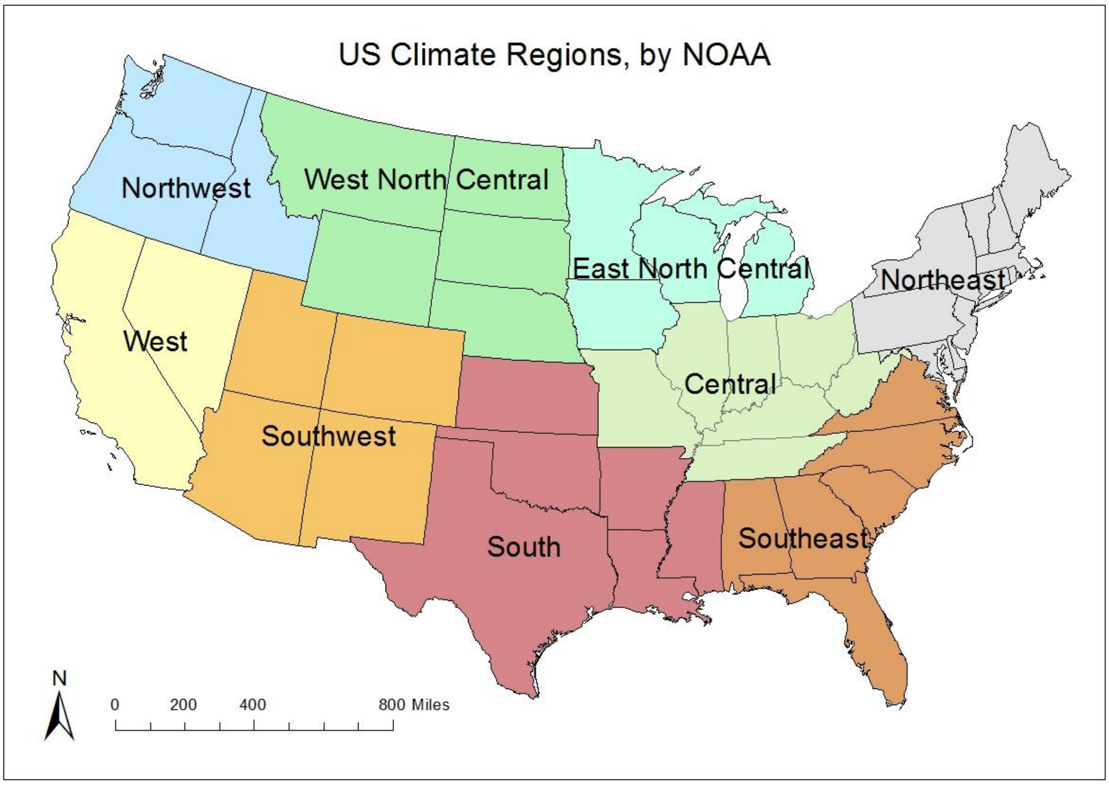
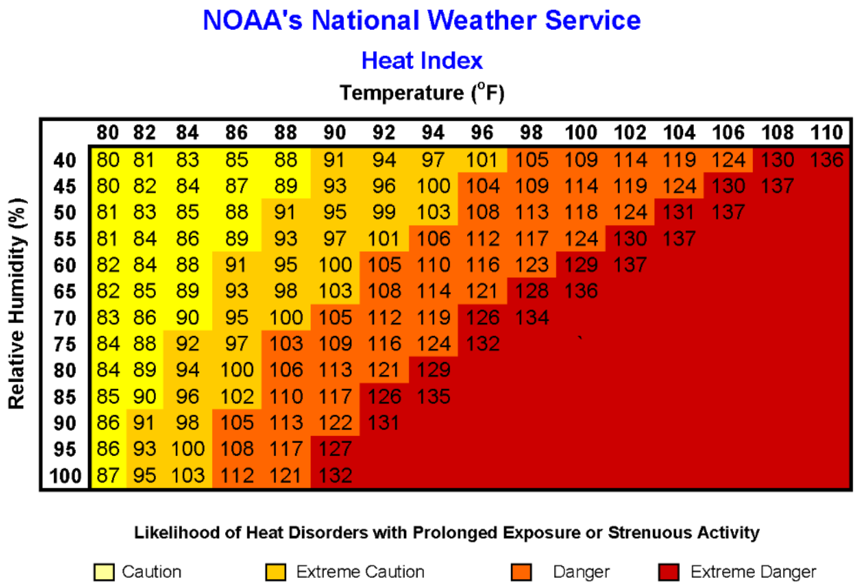


Figure 2 NWS Heat Index Chart



Relative Humidity (%)	80	82	84	86	88	90	92	94	96	98	100	102	104	106	108	110
40	80	81	83	85	88	91	94	97	101	105	109	114	119	124	130	136
45	80	82	84	87	89	93	96	100	104	109	114	119	124	130	137	
50	81	83	85	88	91	95	99	103	108	113	118	124	131	137		
55	81	84	86	89	93	97	101	106	112	117	124	130	137			
60	82	84	88	91	95	100	105	110	116	123	129	137				
65	82	85	89	93	98	103	108	114	121	128	136					
70	83	86	90	95	100	105	112	119	126	134						
75	84	88	92	97	103	109	116	124	132							
80	84	89	94	100	106	113	121	129								
85	85	90	96	102	110	117	126	135								
90	86	91	98	105	113	122	131									
95	86	93	100	108	117	127										
100	87	95	103	112	121	132										

Figure 3July 1995 heatwave in Chicago

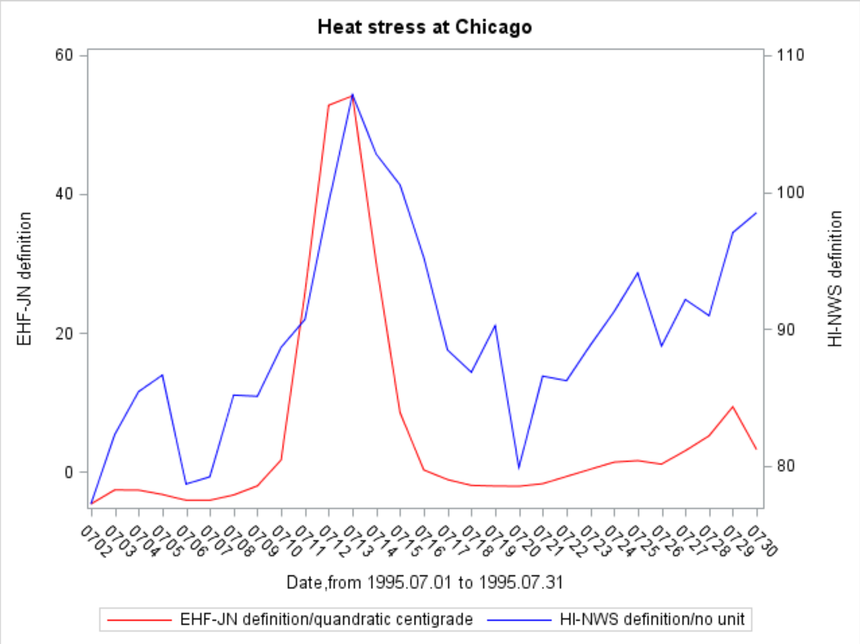


Figure 4Summer 2013 heatwave in Phoenix

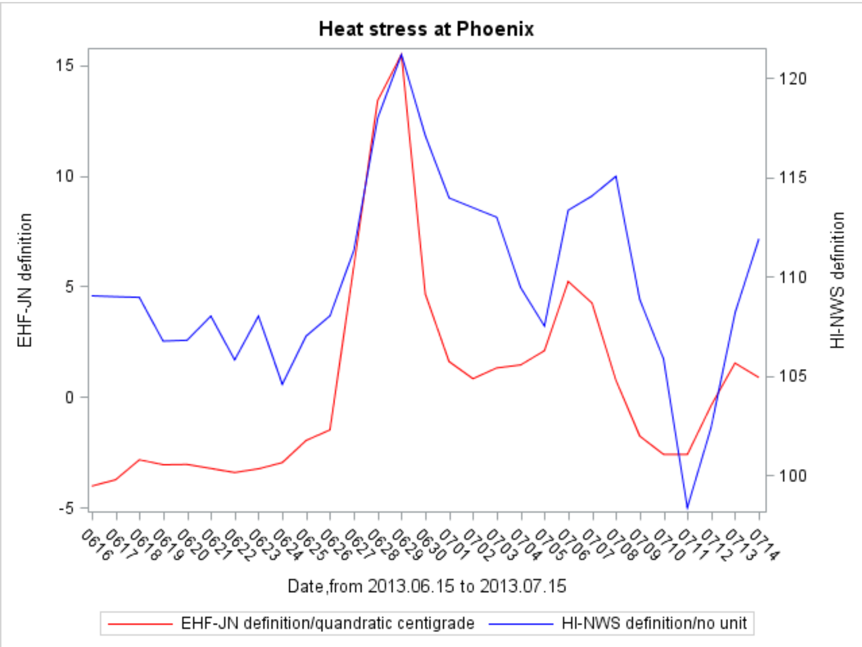


Table 1Events Counts for Three Definitions

	JN	SSC	NWS
Charlotte	72(11323)	3616(11323)	62(11323)
Chicago	64(11323)	1685(11323)	24(11323)
Denver	74(10470)	1581(11323)	85(10204)
Detroit	68(11323)	1952(11323)	20(11323)
Houston	72(11323)	6009(11323)	73(9764)
Indianapolis	63(11323)	2334(11323)	27(11323)
Jacksonville	67(11323)	4907(11323)	315(11323)

Lincoln	68(11323)	2445(11323)	136(11323)
Los Angeles	77(11323)	2458(11323)	1(11323)
Milwaukee	65(11323)	1627(11323)	16(11323)
New York	68(11323)	2413(11323)	78(11323)
Omaha	64(11323)	2033(11323)	86(11323)
Philadelphia	64(11323)	2654(11323)	41(11323)
Phoenix	72(11323)	7205(11323)	883(11323)
Portland	72(11323)	614(11323)	5(11323)
San Antonio	67(11323)	5804(11323)	681(11323)
San Diego	81(11323)	1675(11323)	6(11323)
Seattle	67(11323)	536(11323)	0(11323)

Note: a(b)means a days in the total number of b days are classified as extreme heat days

Table 2 July 1995 Heatwave in Chicago

Date	JN	Ssc	nws
19950702	0	0	0
19950703	0	0	0
19950704	0	0	0
19950705	0	1	0
19950706	0	1	0
19950707	0	0	0
19950708	0	0	0
19950709	0	0	0
19950710	0	0	0
19950711	1	1	0
19950712	1	1	0
19950713	1	1	1
19950714	1	1	1
19950715	0	1	0
19950716	0	1	0
19950717	0	1	0
19950718	0	0	0
19950719	0	0	0
19950720	0	0	0
19950721	0	0	0
19950722	0	1	0
19950723	0	1	0
19950724	0	1	0
19950725	0	1	0
19950726	0	1	0
19950727	0	1	0
19950728	0	0	0
19950729	0	1	0
19950730	0	1	0

Note: 1: extreme heat 0: not extreme heat

Table 3 Summer 2013 Heatwave in Phoenix

date	JN	ssc	nws
20130616	0	1	0
20130617	0	0	0
20130618	0	0	0
20130619	0	1	0
20130620	0	1	0
20130621	0	1	0
20130622	0	1	0
20130623	0	0	0
20130624	0	1	0
20130625	0	1	0
20130626	0	1	0
20130627	0	0	1
20130628	1	1	1
20130629	1	1	1
20130630	0	1	1
20130701	0	0	1
20130702	0	0	0
20130703	0	0	1
20130704	0	1	0
20130705	0	1	0
20130706	0	1	1
20130707	0	1	1
20130708	0	1	1
20130709	0	0	0
20130710	0	1	0
20130711	0	1	0
20130712	0	1	0
20130713	0	1	0
20130714	0	1	1

Note: 1: extreme heat 0: not extreme heat

Table 4 Evaluation of the Three Definitions

Evaluation terms	JN Definition	SSC Definition	NWS Definition
Is it easy to understand?	4	4	5
Is it measured with readily available data?	5	4	4
Is it easily visualized?	5	4	5
Is it predictable with reasonable accuracy?	5	5	5
Is it useful as an indicator of impact?	5	3	5
Is there any flaws for that definition?	4	4	4

Scores range from one to five.

Appendices

SAS code (e.g. New York)

```
*city: New York;
*import the dataset;
proc import out=work.a1 datafile='H:\Thesis\data\New York\newyork_ghcn.csv' dbms=csv
replace;
 getnames=yes;
run;
*checking if there is any missing value;
proc means data=a1 n nmiss;
run;
proc print data=a1 (firstobs=1 obs=10);
run;
*only use one station data;
data a2;
 set a1;
 where station='GHCND:USW00014732' and date<20150101;
run;
proc means data=a2 n nmiss;
run;
*seperate year and monthday from date for future calculation;
data a3;
 set a2;
 year=substr(date,5,4);
 monthday=date-year*10000;
run;
proc print data=a3 (obs=10);
run;
*calculate the daily average data;
data work.b;
 set a3;
 tmax1=tmax/10;
 tmin1=tmin/10;
 t=(tmax1+tmin1)/2;
 label tmax1='daily max temperature/centigrade' tmin1='daily min temperature/centigrade'
t='daily mean temperature/centigrade';
run;
proc contents data=b;
run;
proc print data=b (firstobs=1 obs=10);
run;
*calculate 95% of t for the past 30 years;
proc means data=b p95;
 var t;
 class monthday;
 output out=c p5=t_p5 p95=t_p95;
run;
proc print data=c (firstobs=1 obs=61);
run;
proc means data=b p95;
 var t;
 output out=c1 p5=t_p5 p95=t_p95;
run;
*perform many-to-one match merge;
proc sort data = b;
 by monthday;
run;
proc sort data = c;
 by monthday;
run;
data d;
 merge b c;
```

```
  by monthday;
run;
*sort by year;
proc sort data=d;
  by year;
run;
proc print data=d (firstobs=1 obs=60);
run;
*data cleaning to delete missing value;
data dd;
  set d;
  ind=1;
  if year =. then ind =0;
run;
data d1;
  set dd;
  where ind=1;
  drop ind _type_ _freq_ tmax tmin;
run;
proc print data=d1 (firstobs=1 obs=60);
run;
*Positive contiguous three-day-average daily temperature;
proc sql;
  create table e as select t from d1;
quit;
proc print data=e (firstobs=1 obs=60);
run;
proc transpose data=e out=e1;
run;
data e2;
  set e1;
  drop _NAME_;
run;
proc print data=e2;
var col1-col20;
run;
data e3;
  set e2;
  array a(11324) col1-col11324;
  array b(11324) b1-b11324;
  do i=1 to 11322 by 1;
  b[i]=(a[i+2]+a[i+1]+a[i])/3;
  end;
run;
proc print data=e3;
run;
data e4;
  set e3;
  drop col1-col11324;
run;
proc transpose data=e4 out=e5;
run;
data e6;
  merge d1 e5;
run;
data e7;
  set e6;
  if _NAME_='i' then delete;
run;
data e8;
  set e7;
```

```
 drop _NAME_;
run;
data e9;
 set e8;
 tda=col1;
 label tda='three days average';
run;
data e10;
 set e9;
 drop col1;
run;
proc print data=e10 (firstobs=1 obs=60);
run;
*calculate the Excess Heat;
data f;
 set e10;
 ehisig=tda-27.25;
 label ehisig='excess heat index';
run;
proc print data=f (firstobs=1 obs=60);
run;
*acclimatisation: negative contiguous thirty-day-average daily temperature;
proc sql;
 create table g as select t from f;
quit;
proc print data=g (firstobs=1 obs=60);
run;
proc transpose data=g out=g1;
run;
data g2;
 set g1;
 drop _NAME_;
run;
proc print data=g2;
var col1-col20;
run;
data g3;
 set g2;
 array a(11324) col1-col11324;
 array b(11324) b1-b11324;
 do i=31 to 11322 by 1;
 b[i]=mean(a[i-30],a[i-29],a[i-28],a[i-27],a[i-26],a[i-25],a[i-24],a[i-23],a[i-22],a[i-21],a[i-20],a[i-
19],a[i-18],a[i-17],a[i-16],a[i-15],a[i-14],a[i-13],a[i-12],a[i-11],a[i-10],a[i-9],a[i-8],a[i-7],a[i-6],a[i-
5],a[i-4],a[i-3],a[i-2],a[i-1]);
 end;
run;
proc print data=g3;
var col1-col60;
run;
data g4;
 set g3;
 drop col1-col11324;
run;
proc transpose data=g4 out=g5;
run;
data g6;
 merge f g5;
run;
data g7;
 set g6;
 if _NAME_='i' then delete;
```

```
run;
data g8;
 set g7;
 drop _NAME_;
run;
data g9;
 set g8;
 accl=col1;
 label accl='negative contiguous thirty-day-average daily temperature';
run;
data g10;
 set g9;
 drop col1;
run;
proc print data=g10 (firstobs=1 obs=60);
run;
*calculate the Heat Stress;
data h;
 set g10;
 ehiaccl=tda-accl;
 label ehiaccl='heat stress index';
run;
proc print data=h (firstobs=1 obs=60);
run;
*calculate the Excess Heat Factor;
data i;
 set h;
 a=max(1,ehiaccl);
 ehf=ehisig*a;
 label ehf='excess heat factor/quandratic centigrade';
run;
data i1;
 set i;
 drop a;
run;
proc print data=i1 (firstobs=1 obs=60);
run;
*Coldwave Calculation;
data j;
 set i1;
 ecisig=tda-t_p5;
 eciaccl=ehiaccl;
 a=min(-1,eciaccl);
 ecf=-ecisig*a;
 label ecisig='excess cold index' eciaccl='cold stress index' ecf='excess cold factor';
run;
data j1;
 set j;
 drop a;
run;
proc print data=j1 (firstobs=1 obs=60);
run;
*Threshold of the Metropolitan statistical area;
proc univariate data=i1;
 var ehf;
run;
data k;
 set j1;
 where ehf>=0;
run;
proc univariate data=k noprint;
```

```
  var ehf;
  output out=k1 pctlpts=85 pctlpre=P;
run;
proc print data=k1;
run;
data k2;
  set k;
  where ehf>13.2018;
run;
proc print data=k2;
run;
data k3;
  set j1;
  if ehf>13.2018 then JN=1;
  else JN=0;
run;
data k4;
  set k3;
  keep date awnd year monthday tmax1 tmin1 t ehf JN;
run;
proc print data=k4 (obs=60);
run;
proc freq data=k4;
  tables JN;
run;
*import ssc data;
data l1;
  infile "H:\Thesis\data\New York\newyork_ssc.txt";
  input station$ datetime weathertype;
run;
data l2;
  set l1;
  where datetime >= 19840101;
run;
proc freq data=l2;
  tables weathertype;
run;
data l3;
  set l2;
  if weathertype=3 then ssc=1;
  else if weathertype=6 then ssc=1;
  else if weathertype=66 then ssc=1;
  else if weathertype=67 then ssc=1;
  else ssc=0;
run;
proc print data=l3 (obs=10);
run;
proc freq data=l3;
  tables ssc;
run;
*merge the JN and SSC;
data l4;
  set l3;
  date=datetime;
run;
data l5;
  set l4;
  drop datetime;
run;
proc print data=l5 (obs=10);
run;
```

```
proc sort data=l5;
 by date;
run;
proc sort data=k4;
 by date;
run;
data m1;
 merge l5 k4;
 by date;
run;
proc print data=m1 (obs=10);
run;
*data cleaning;
data m2;
 set m1;
 ind=1;
 if date=. then ind=0;
run;
data m3;
 set m2;
 where ind=1;
 drop station ind;
run;
proc print data=m3 (obs=10);
run;
*import the humidity data;
data n1;
 infile "H:\Thesis\data\New York\newyork_crh.txt" dsd;
 input USAF NCDC Date HrMn I$ Type$ QCP$ Temp Q1$ Dewpt Q2$ RHx ;
run;
proc print data=n1 (obs=10);
run;
data n2;
set n1;
keep Date HrMn RHx;
if RHx=999 then RHx=50;
run;
proc means data=n2 n nmiss;
run;
proc univariate data=n2;
var rhx;
run;
proc print data=n2;
where rhx=107;
run;
*calculate the average humidity during the 24 hours;
proc sql;
 create table n3 as select rhx from n2;
quit;
proc print data=n3 (firstobs=1 obs=60);
run;
proc transpose data=n3 out=n4;
run;
data n5;
 set n4;
 drop _NAME_;
run;
proc print data=n5;
var col1-col20;
run;
data n5;
```

```
 set n4;
 array a(310160) col1-col310160;
 array b(310160) b1-b310160;
 do i=13 to 310148 by 1;
 b[i]=(a[i+11]+a[i+10]+a[i+9]+a[i+8]+a[i+7]+a[i+6]+a[i+5]+a[i+4]+a[i+3]+a[i+2]+a[i+1]+a[i]+a[i-
1]+a[i-2]+a[i-3]+a[i-4]+a[i-5]+a[i-6]+a[i-7]+a[i-8]+a[i-9]+a[i-10]+a[i-11]+a[i-12])/24;
 end;
run;
data n6;
 set n5;
 drop col1-col310160;
run;
proc transpose data=n6 out=n7;
run;
data n8;
set n7;
meanrhx=RHx;
drop RHx;
run;
proc print data=n8 (obs=20);
run;
data n9;
 merge n1 n8;
run;
proc print data=n9 (obs=20);
run;
data n10;
 set n9;
 if _NAME_='i' then delete;
run;
data n11;
 set n10;
 drop _NAME_;
run;
proc print data=n11 (firstobs=1 obs=60);
run;
data n12;
set n11;
where hrmn=1200;
run;
data n13;
set n12;
keep date meanrhx;
run;
proc print data=n13 (obs=10);
run;
*merge JN SSC crh;
proc sort data=m3;
by date;
run;
proc sort data=n13;
by date;
run;
data o1;
merge m3 n13;
by date;
run;
proc print data=o1 (obs=10);
run;
*national weather services;
data p1;
```

```
set o1;
F=tmax1*9/5+32;
run;
data p2;
set p1;
Rothfusz1=(F-68.0)*1.2;
Rothfusz2=F+61.0+Rothfusz1+(meanrhx*0.094);
Rothfusz=0.5*Rothfusz2;
run;
proc print data=p2 (obs=20);
run;
proc univariate data=p2;
var F;
run;
proc univariate data=p2;
var meanrhx;
run;
data p3;
set p2;
if F>=108 then nws=1;
else if F>=106 and meanrhx>=45 then nws=1;
else if F>=104 and meanrhx>=50 then nws=1;
else if F>=102 and meanrhx>=55 then nws=1;
else if F>=100 and meanrhx>=60 then nws=1;
else if F>=98 and meanrhx>=65 then nws=1;
else if F>=96 and meanrhx>=70 then nws=1;
else if F>=94 and meanrhx>=80 then nws=1;
else if F>=92 and meanrhx>=85 then nws=1;
else if F>=90 and meanrhx>=95 then nws=1;
else nws=0;
run;
proc freq data=p3;
tables nws;
run;
*export the final dataset;
data p4;
set p3;
keep date jn ssc nws;
run;
proc export data=p4 dbms=xlsx outfile="H:\Thesis\data\New York\final.xlsz" replace;
run;
proc freq data=p4;
table jn ssc nws;
run;
*import the dataset;
proc import out=work.a1 datafile='H:\Thesis\data\overalltable.csv' dbms=csv replace;
 getnames=yes;
run;
proc print data=a1;
run;
data a2;
set a1;
JN1=(JN-58)/23;
SSC1=(SSC-536)/6669;
NWS1=NWS/883;
*variance analysis;
proc univariate data=work.a2;
var JN1;
run;
proc univariate data=work.a2;
var SSC1;
```

```
run;
proc univariate data=work.a2;
var NWS1;
run;
```