



# Comparing exposure metrics for classifying ‘dangerous heat’ in heat wave and health warning systems

Kai Zhang <sup>a,b,\*</sup>, Richard B. Rood <sup>c</sup>, George Michailidis <sup>d</sup>, Evan M. Oswald <sup>c</sup>, Joel D. Schwartz <sup>e</sup>, Antonella Zanobetti <sup>e</sup>, Kristie L. Ebi <sup>f</sup>, Marie S. O'Neill <sup>a,b</sup>

<sup>a</sup> Department of Environmental Health Sciences, University of Michigan, Ann Arbor, MI, USA

<sup>b</sup> Department of Epidemiology, University of Michigan, Ann Arbor, MI, USA

<sup>c</sup> Department of Atmospheric, Oceanic and Space Sciences, University of Michigan, Ann Arbor, MI, USA

<sup>d</sup> Department of Statistics, University of Michigan, Ann Arbor, MI, USA

<sup>e</sup> Department of Environmental Health, Harvard University, MA, USA

<sup>f</sup> Department of Environmental Health, Stanford University, Palo Alto, CA, USA

## ARTICLE INFO

### Article history:

Received 11 October 2011

Accepted 7 May 2012

Available online 5 June 2012

### Keywords:

Air mass

Heat wave

Heat health warning system

Model-based clustering

Temperature

## ABSTRACT

Heat waves have been linked to excess mortality and morbidity, and are projected to increase in frequency and intensity with a warming climate. This study compares exposure metrics to trigger heat wave and health warning systems (HHWS), and introduces a novel multi-level hybrid clustering method to identify potential dangerously hot days. Two-level and three-level hybrid clustering analysis as well as common indices used to trigger HHWS, including spatial synoptic classification (SSC), and the 90th, 95th, and 99th percentiles of minimum and relative minimum temperature (using a 10 day reference period), were calculated using a summertime weather dataset in Detroit from 1976 to 2006. The days classified as ‘hot’ with hybrid clustering analysis, SSC, minimum and relative minimum temperature methods differed by method type. SSC tended to include the days with, on average, 2.5 °C lower daily minimum temperature and 5.3 °C lower dew point than days identified by other methods. These metrics were evaluated by comparing their performance in predicting excess daily mortality. The 99th percentile of minimum temperature was generally the most predictive, followed by the three-level hybrid clustering method, the 95th percentile of minimum temperature, SSC and others. Our proposed clustering framework has more flexibility and requires less substantial meteorological prior information than the synoptic classification methods. Comparison of these metrics in predicting excess daily mortality suggests that metrics thought to better characterize physiological heat stress by considering several weather conditions simultaneously may not be the same metrics that are better at predicting heat-related mortality, which has significant implications in HHWSs.

© 2012 Elsevier Ltd. All rights reserved.

## 1. Introduction

Heat waves have been receiving increasing attention recently due to the risks they pose to human health. The Russian heat wave of 2010, for example, is estimated to have caused around 55,000 heat-related deaths (Barriopedro et al., 2011). The frequency and intensity of heat waves as well as other extreme weather events are expected to increase as a consequence of climate change (Meehl and Tebaldi, 2004). Epidemiological studies have shown that hot weather is associated with mortality, hospital admissions, heat stroke, heat exhaustion, cardiovascular and respiratory diseases (Kovats and Hajat, 2008). A variety of heat wave and health plans have been proposed and implemented to prevent such health consequences. One key component of these plans is a

timely and accurate heat alert system, also commonly called a heat wave and health warning system (HHWS) (Matthies and Menne, 2009). A HHWS is a ‘system that uses meteorological forecasts to initiate acute public health interventions designed to reduce heat-related impacts on human health during atypically hot weather’ (Kovats and Ebi, 2006). HHWSs are intended to help local government and residents prepare for potential heat waves.

Several approaches have been used to identify potentially dangerous hot days based on weather forecasts, and define a metric that triggers an alert for a HHWS. These metrics include: absolute or percentile-based temperature threshold, heat index (HI), physiologically-based discomfort classifications, the temperature–mortality relationship derived from epidemiologic analysis, and spatial synoptic classification (SSC) (Hajat et al., 2010; National Weather Service, 2005). The SSC method was first proposed for application in Eastern and Central U.S. by Kalkstein and Greene (1997). It was then refined by Sheridan and Kalkstein (2004) to take Western U.S. and Canada into consideration, and was recently modified for Western Europe by Bower et al. (2007).

\* Corresponding author at: 1420 Washington Heights, SPH-II Room 5039, Ann Arbor, MI 48109-2029, USA. Tel.: +1 734 763 2663; fax: +1 734 936 7283.

E-mail address: [zhangkai@umich.edu](mailto:zhangkai@umich.edu) (K. Zhang).

Detailed descriptions and examples of these metrics are described in supporting information. SSC and its variations have been used in many HHWSs in the U.S. and elsewhere (Sheridan and Kalkstein, 2004). SSC itself does not account for timing in season. However, when Kalkstein et al. (2011) established a HHWS using SSC, they modeled mortality as a function of SSC categories, timing in season, duration of a heat wave and other factors within a linear regression framework.

Two recent studies compared different heat–health warning trigger metrics in estimating heat–mortality relationships, and their findings were substantially different. Metzger et al. (2010) compared five metrics (maximum HI, maximum, minimum, and average temperature, and SSC) in estimating mortality risk during hot weather for New York City, and reported that the maximum HI performed similarly to other metrics. Hajat et al. (2010) compared four HHWS trigger approaches (SSC, epidemiologic assessment of the temperature–mortality associations, temperature–humidity index, and physiologic classification) in predicting dangerously hot days using data from Chicago, U.S.; London, U.K.; Madrid, Spain; and Montreal, Canada, and found that these four approaches did not consistently identify the same days as being ‘dangerous’ for health. Although these evaluations of trigger metrics in combination with health data are informative, their focus was not on the methods for creation of these metrics, and they did not consider the quality of weather forecast and other factors. Our goal is to evaluate the exposure metrics that are used, or could be used, as HHWS triggers to provide more in-depth insights into how these triggers are created and how they might be used in designing a HHWS.

Conceptually, the air mass approach seems to hold advantages over the other methods because it uses a variety of meteorological parameters to determine potentially oppressive weather conditions. However, SSC methods rely on a sophisticated calculation which is neither easily replicated nor transparent to non-technical audiences. For example, SSC depends on the subjective selection of seeding days (days that are considered most representative of the particular air mass type for that area, such as moist tropical (MT) category), and the ability to select these days is limited to a population of trained and experienced experts. Moreover, SSC-based HHWSs are even more complicated and difficult to interpret. They generally select variables from a large number of correlated weather parameters, resulting in considerable differences in variable selection for nearby cities with similar weather. For example, Kalkstein et al. (2011) show that maximum temperature, Julian day number in summer and day number in sequence were the most important explanatory variables for excess mortality within an offensive air mass for Baltimore, MD, but maximum temperature, maximum and minimum dew point temperature, and Julian day number were the key predictors of mortality in nearby Washington D.C. This suggests overfitting is an issue. From the public health perspective, a data-driven and flexible method for identifying extreme hot days, easily understood and interpreted for public health intervention applications, may have potential applications in heat health warning systems. In addition, discussions with meteorologists involved in issuing heat warnings have confirmed that further evaluation of triggers used for HHWSs would be a welcome contribution (O'Neill et al., 2011).

This study aims to develop a flexible data-driven method to identify “dangerously” hot weather patterns from an exposure assessment perspective, and then to compare it to some selected exposure metrics commonly used in HHWSs for classifying hot days. This method takes advantage of recent advances in statistical methods and decreases the demand for prior meteorological background information. We present a novel multi-level hybrid clustering method that combines the model-based clustering method the partition around medoids (PAM) and hierarchical methods. This proposed method was applied to a time series of observed meteorological data from Detroit, Michigan, USA. Results from the new clustering method were compared to some common exposure metrics in HHWSs, including SSC triggers, in terms of their concordance in identifying potentially dangerously hot days.

## 2. Methods for identifying dangerously hot days

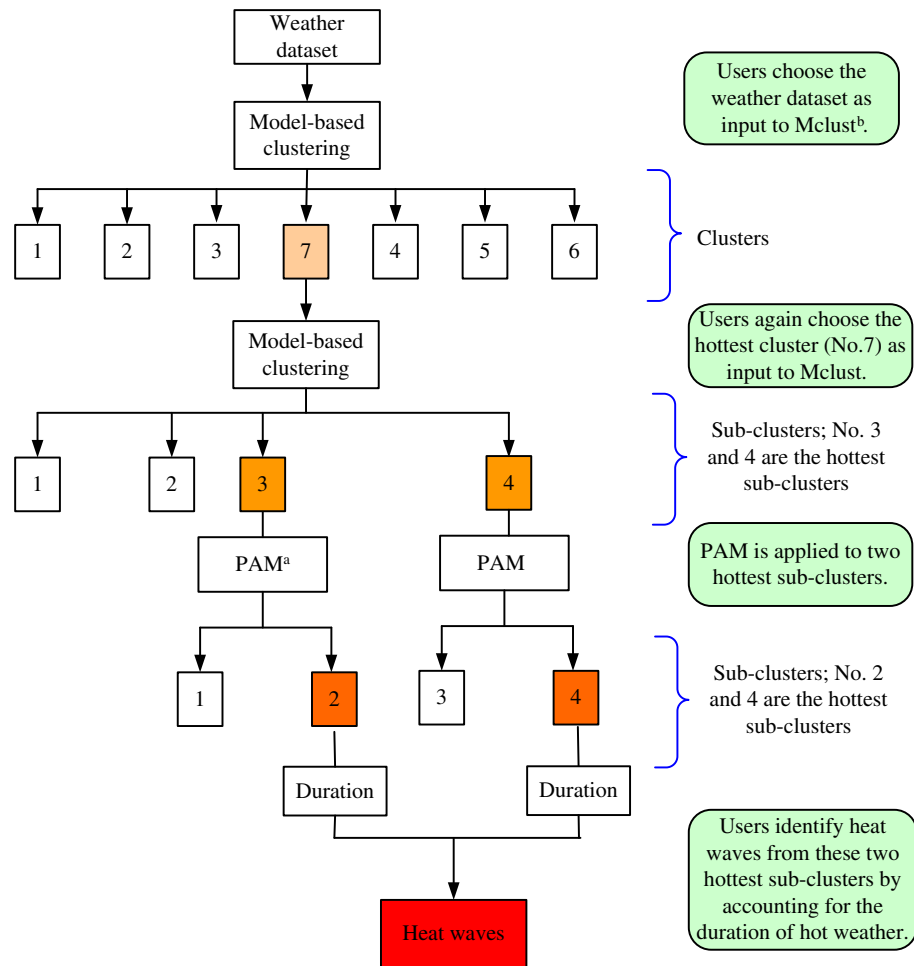
### 2.1. Introduction to clustering methods

Cluster analysis methods include partitioning methods, hierarchical cluster methods and model-based clustering methods, among others. Partitioning and hierarchical methods do not require underlying statistical models, while model-based clustering methods are established on probability models. K-means and the PAM algorithm are two popular partitioning methods. K-means and PAM methods divide all observations into several clusters at once with the number of clusters determined *a priori*. PAM is more robust than the K-means clustering method to outliers and missing values because it minimizes a sum of un-squared dissimilarities, rather than a sum of squared Euclidean distances, and avoids initial guesses for the cluster centers which is what is done with K-means (Struyf et al., 1997). Hierarchical cluster methods produce different levels of clustering of the dataset, and usually produce a graphical tree displaying this hierarchical clustering structure. Both partitioning and hierarchical methods are distance-based approaches, and lack underlying statistical models (Izenman, 2008; Struyf et al., 1997). In contrast, model-based clustering methods are based on probability models, and view data as generated from a finite mixture model (Fraley and Raftery, 2007). Each component (distribution) of a mixture model corresponds to a cluster, and is usually modeled as a normal or Gaussian distribution.

Model parameters are estimated using a maximum likelihood approach, and the number of clusters can be optimally determined by use of statistical criteria such as the Bayesian Information Criterion (BIC), rather than subjective choice in partitioning methods. Model-based approaches provide the flexibility to identify clusters by accounting for their geometric features (e.g., shape, volume and orientation), which are critical to the performance of clustering methods. Distance-based approaches cannot take into consideration the clusters' shape and structure since they depend on the proximity between data objects. For example, supporting Information Figures S1–2 illustrate a few ellipse-shaped clusters with different volumes and orientations. The popular K-means method usually works well only for spherical clusters with similar size while the model-based cluster method can deal with clusters with different shapes and varying sizes. We provided an extended version in introducing these clustering methods in supporting information.

### 2.2. New clustering approach

In the current paper, we propose a multi-level hybrid clustering method that includes advantages of the hierarchical, partitioning and model-based approaches described above. We apply this method to meteorological data to evaluate its potential for use in triggering a HHWS. Fig. 1 describes the method's framework, which incorporates model-based clustering and PAM approaches and also takes into account the duration of hot weather. The basic idea of this method is sequential clustering: model-based clustering techniques are used to iteratively divide observations into clusters as many times as possible depending on user's demand, and some of the clusters identified by the previous clustering step are used as input for the next clustering step. PAM is introduced when identified small clusters cannot be divided further by the model-based clustering method. Users can determine which clusters are the hottest at each step by examining the descriptive statistics of the temperature parameter within each cluster (e.g., mean, median, minimum and maximum values). The duration of hot weather is taken into consideration in the last step, and applied to the hottest clusters identified by the sequential clustering. The three stages shown in this figure are an arbitrary number selected for illustrative purposes. The number of levels in practice can range from two to a large number, depending on the user's needs. This strategy provides flexibility for local governments to customize their own HHWSs.



**Fig. 1.** Multi-level hybrid clustering framework as applied to daily temperature and dew point data measured in Detroit, Michigan, 1976–2006. (Note: a. PAM, a partitioning clustering method that divides data into several clusters around ‘medoids’; b. Mclust is an R package for model-based clustering.)

The SSC is a classification procedure with a fixed number of clusters based on meteorological identification of air mass types, e.g., the seven air mass types described in Kalkstein and Greene (1997). However, air mass characteristics may vary with cities and their relation to human health has not been established for many cities. The linear discriminant analysis (Kalkstein and Greene, 1997), equal weighting (Sheridan, 2002), and hierarchical K-means (Bower et al., 2007) methods used in SSC perform well for spherical cluster shapes. However, we explored meteorological parameters in high dimensions using GGobi (an open source visualization program, Buja et al. (2003)), and found the meteorological data used to define air masses do not always have sphere-like shapes. As Supporting Information Figs. 1 and 2 show, daily maximum temperature and dew point during summers in Detroit have ellipse-like shapes. As mentioned earlier, model-based clustering methods provide more flexibility to model different shapes of clusters, and determine the number of clusters automatically using a data-driven method. An R package for model-based clustering, “mclust”, was selected to implement model-based clustering (Fraley and Raftery, 2007). In some cases, the hottest sub-cluster identified by a model-based clustering method might be relatively large. Certain model-based methods cannot divide sub-clusters further since they utilize the Bayesian Information Criterion (BIC), which penalizes larger numbers of clusters. Thus, we introduce the PAM method to allow the user the flexibility to create additional ‘hot’ sub-clusters in such cases. An R package, “cluster”, was chosen to implement PAM (Struyf et al., 1997). The number of clusters is determined by looking at the tree produced by applying a hierarchical cluster method.

We chose an agglomerative hierarchical algorithm in the “cluster” package, “agnes”, for this study (Struyf et al., 1997).

In the final stage of our proposed method, duration of high temperatures is taken into account because a period of duration of high temperature is likely to be more hazardous to health than a one-day hot spell followed by a rapid drop in temperature and/or humidity (Rocklov et al., 2011). The number of heat wave days is calculated by counting days where at least two consecutive days occurred in the hottest sub-clusters identified by the two level hybrid clustering method (HCM2) or the three level hybrid clustering method (HCM3). The more consecutive days that are counted, the more potentially dangerous hot weather is for health. The proposed cluster analysis was implemented in the R statistical software (version 2.13.1) (R Development Core Team, 2006).

### 2.3. Meteorological data

We obtained hourly weather data during the summertime (May 1st to September 30th) from 1976 to 2006 from the National Climate Data Center, derived from the Detroit Metropolitan Airport monitoring station in Romulus, Michigan, USA (Station name: Detroit/Metropolitan). We abbreviate this dataset as ‘DTW’ to correspond to that airport’s code. Daily measures of maximum and minimum temperatures and dew point temperatures were created from these hourly observations. These four meteorological variables were used in the data analysis. Results including the characteristics of clusters and number of heat wave days identified from different exposure metrics in HHWS triggers were compared.

## 2.4. Comparison of the exposure metrics in HHWS triggers

The purpose of the comparison among different exposure metrics in HHWS triggers is to assess the level of agreement across the different methods in terms of classifying particular days or periods as being ‘dangerously’ hot. In particular, we compared the ‘dangerous’ heat wave days lasting at least two consecutive days identified by each method, which are more dangerous than those occurring in one day. We chose a two-day threshold to achieve a reasonable sample size of identified heat wave days for our comparative analysis, although we recognize that longer durations are likely more health-relevant. Use of a one-day threshold resulted in too many identified heat wave days over the study period, while the three-day threshold reduced the number of identified days dramatically. Several exposure metrics in HHWS triggers (SSC,  $\geq 90$ th, 95th, 99th minimum temperature/relative temperature and lasting for at least two days) were calculated using the Detroit weather data discussed above. SSC categories for each day during the study period (May 1st to September 30th from 1976 to 2006) were obtained for Detroit from a website maintained by Dr. Scott Sheridan (Sheridan, 2009). Heat waves identified by percentiles of temperature in this study are defined as days which minimum temperature is above a percentile based on this measured daily temperature metric over all summertime dates from 1976 to 2006 in Detroit. Three percentiles (90th, 95th and 99th) of relative minimum temperature were estimated with reference to the average temperatures for every fixed 10-day interval (e.g., May 1–10, 11–20, etc.) from May to September over the 31 years of the weather dataset. For example, the daily minimum temperatures for May 1st, 2nd, 3rd through 10th are always compared to the distribution of daily minimum temperatures of the first ten days in May over 31 years; and the minimum temperatures of May 11th and 12th use the same reference period of May 11–20. Compared to the minimum temperature, the relative minimum temperature can account for the temporal component of ‘dangerous’ heat. In other words, an equivalent daily temperature could be more dangerous in May than in August because it is unusual that early in the season. The duration of high temperature is also considered separately for these selected exposure metrics in HHWS triggers by counting at least two consecutive days occurred in the combined dataset of dry tropical (DT), MT+ and MT++ clusters for SSC, or at least two consecutive days occurred in which the minimum temperatures exceeded given thresholds.

## 2.5. Health effect analysis

A health effect analysis was conducted using daily mortality records from Detroit Metropolitan area obtained from the National Center for Health Statistics. International Classification of Disease (ICD) codes Eighth revision (ICD-8), Ninth revision (ICD-9) and Tenth revision (ICD-10) were used for the mortality data from 1976–1978, 1979–1997, and 1998–2006, respectively. Daily total mortality included nonaccidental causes (ICD-8 Codes 0–999; ICD-9 Codes 0–999; ICD-10 Codes beginning with S through Z were excluded). We modeled total mortality counts as a smooth function of date where we used an integer value of the day of the time series (degrees of freedom = 5) while adjusting for day of week and year, over the time period of our study (1976–2006). This single smooth function was created to represent the annual ‘expected’ pattern of daily mortality averaged over the entire 31 years of data. Then, using the daily deaths predicted by this smooth function, we calculated the difference between the observed daily and the ‘expected’ for all-cause mortality. This variable can take on negative or positive values and we refer to it as deviation from typical daily mortality counts. We compared the median values of deviation from typical daily mortality counts across all “extremely hot days” identified by our method to those calculated based on other selected exposure metrics. The smooth function was implemented using the penalized splines of

“mgcv” R package (version 1.7–6) in the R statistical software (version 2.13.1) (R Development Core Team, 2006; Wood, 2008).

## 3. Results

### 3.1. Descriptive statistics

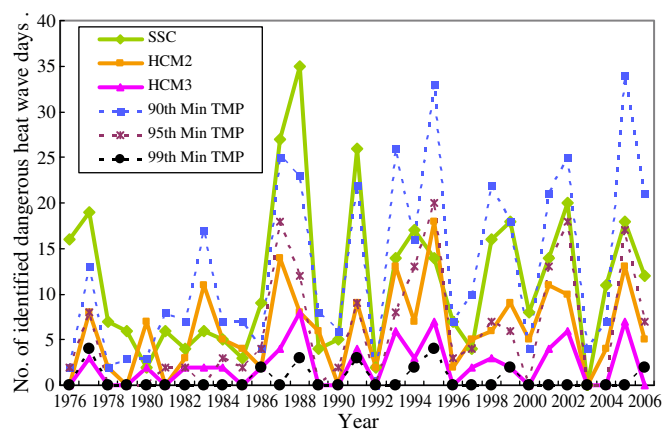
Supporting information Table S1 presents descriptive statistics of the four meteorological variables of interest during the summertime for the 31-year period in Detroit. The distributions of the four meteorological parameters were slightly left skewed, and some extreme values could be seen in the two tails (Supporting Information Table S1 and Figures S3–4). During the study period, the average maximum temperature and dew point were around 10 and 6 °C higher than the average minimum temperature and dew point, respectively. Strong significant correlations (all Pearson correlation coefficients ( $r$ )  $\geq 0.7$ ; all  $p$ -values  $< 0.05$ ) have been shown between the different meteorological metrics. The strongest correlation ( $r = 0.89$ ) occurred between daily minimum temperatures and daily maximum dew points, followed by the correlation ( $r = 0.88$ ) between daily minimum temperatures and daily minimum dew points, the correlation ( $r = 0.88$ ) between the two daily dew point metrics, and the correlation ( $r = 0.76$ ) between the two temperature metrics (shown in Supporting information Table S2).

### 3.2. Hybrid clustering analysis results

Cluster analysis was applied at three levels: the model-based clustering method was applied at the first two levels; and PAM was used at the last level. Seven clusters were identified at the first level, and the hottest cluster (7th) was further divided into four sub-clusters at the second level. Two sub-clusters were relatively hotter than the remaining clusters, and both had higher mean temperature and dew point temperature. A slight difference was observed for these two sub-clusters in that one was inclined to having more moisture and the other tended to have higher temperatures. The third sub-cluster at the second level was further classified into two clusters (1st and 2nd at the third level), and the fourth sub-cluster was divided into another two clusters (3rd and 4th at the third level). The 1st and 4th cluster at the third level had both higher temperature and dew point temperature. Like the second level, one cluster (1st) was associated with higher temperature, while another (4th) was related to higher dew point values. Supporting information Table S3 presents the results obtained by applying our proposed hybrid clustering method to the observed DTW dataset.

### 3.3. Comparison of the exposure metrics in HHWS triggers

Fig. 2 presents the number of annual heat wave days identified by selected exposure metrics in HHWS triggers over a 31 year period. For the percentile-related minimum temperature metrics, the number of extremely hot days was mainly determined by the level of percentile. With increased percentiles from 90th, 95th to 99th, the number of identified hot days decreased from up to 34 to less than 4, depending on the year. In general, these metrics differed moderately or substantially in terms of the number of identified dangerous heat wave days. For example, the 90th percentile minimum temperature tended to define more potentially dangerous heat wave days (405 days or 8.5% of summer days



**Fig. 2.** Number of annual dangerous heat wave days identified by selected exposure metrics in the Detroit, Michigan area, 1976–2006. (Note: The number of heat wave days was calculated by counting days where at least two consecutive days occurred in the hottest sub-clusters identified by the two level hybrid clustering method (HCM2), the three level hybrid clustering method (HCM3), spatial synoptic classification (SSC), and the 90th, 95th, and 99th percentiles of minimum temperature (TMP) during the study period. For SSC, the number of days was calculated by counting days where at least two consecutive days occurred in the combined dataset of dry tropical, moist tropical+ and moist tropical++ clusters. For HCM2 and HCM3, the number of days was calculated by counting days where at least two consecutive days occurred in the combined dataset of two hottest subclusters.)



**Table 1**

Descriptive statistics of the number of heat wave days (defined as at least two consecutive days of hot days identified by each method) identified by SSC, but not by other methods (hybrid clustering method, HCM; temperature, TMP). Meteorological characteristics of those days also are shown (mean  $\pm$  standard deviation in units of  $^{\circ}\text{C}$ ) (Min temp, daily minimum temperature; Max temp, daily maximum temperature; Min DP, daily minimum dew point; and Max DP, daily maximum dew point).

| Methods          | N   | Min temp       | Max temp       | Min DP         | Max DP         |
|------------------|-----|----------------|----------------|----------------|----------------|
| SSC-HCM2         | 294 | 18.3 $\pm$ 4.0 | 31.6 $\pm$ 2.9 | 13.4 $\pm$ 5.2 | 19.2 $\pm$ 4.6 |
| SSC-HCM3         | 311 | 18.5 $\pm$ 4.0 | 31.5 $\pm$ 2.8 | 13.7 $\pm$ 5.2 | 19.4 $\pm$ 4.5 |
| SSC-90th Min TMP | 204 | 16.5 $\pm$ 3.5 | 30.7 $\pm$ 2.7 | 11.6 $\pm$ 4.9 | 17.6 $\pm$ 4.4 |
| SSC-95th Min TMP | 246 | 17.3 $\pm$ 3.7 | 31.0 $\pm$ 2.7 | 12.7 $\pm$ 5.1 | 18.4 $\pm$ 4.5 |
| SSC-99th Min TMP | 334 | 18.7 $\pm$ 3.9 | 31.5 $\pm$ 2.7 | 14.2 $\pm$ 5.3 | 19.6 $\pm$ 4.5 |

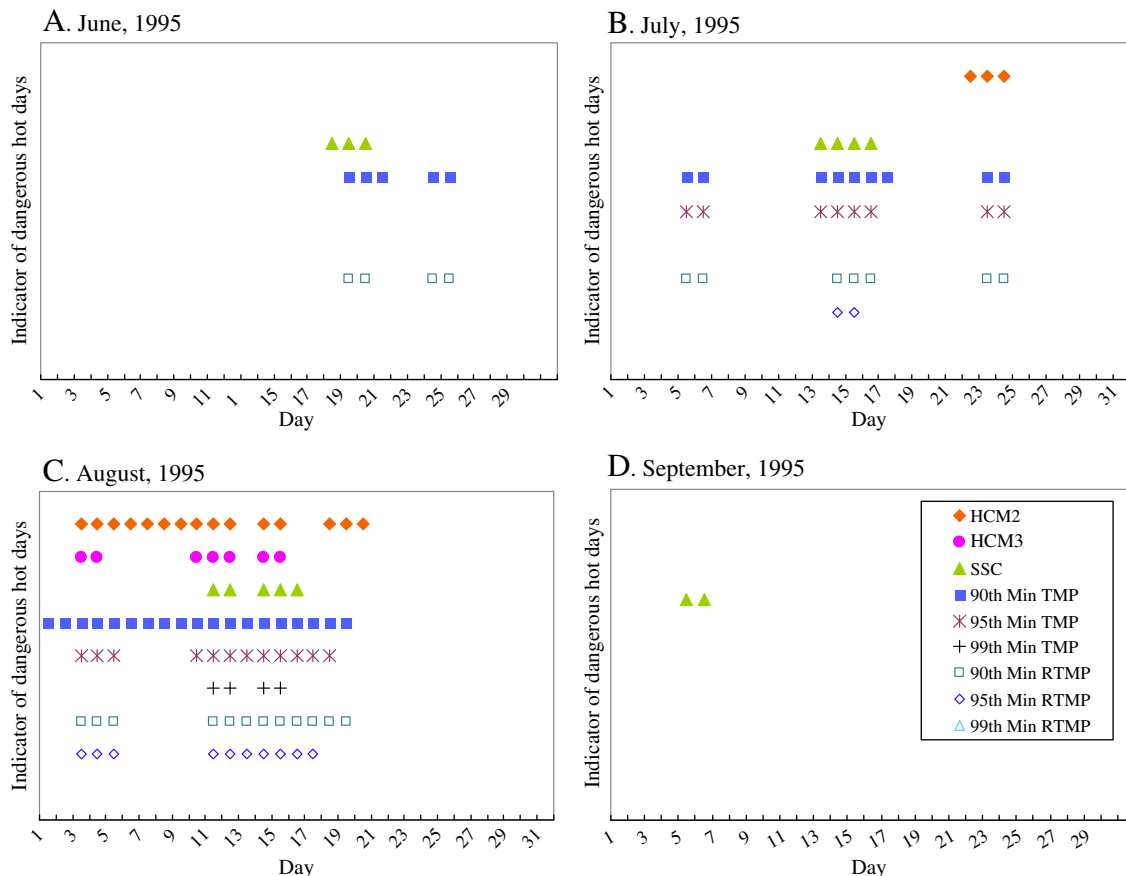
over 31 years) than SSC (355 days or 7.5%), two-level hybrid clustering analysis (189 days or 4.0%), three-level hybrid clustering analysis (69 days or 1.5%), the 95th and 99th percentile minimum temperatures (180 and 22 days or 3.8% or 0.5%). The number of dangerous heat wave days identified by the SSC was more than those classified by the 90th percentile minimum temperature in the first few years, and then was less than the latter for most of the rest of the years; the highest number of heat wave days classified by SSC occurred in 1988, but all other methods identified the most heat wave days in 1995. However, the temporal patterns shown in Fig. 2 indicate the identified heat wave days are generally highly correlated across those metrics, which is consistent with the correlations of the number of annual heat wave days identified by each method (Supporting Information Table S4). After excluding the 99th percentile minimum temperature, other methods showed strong correlations (Pearson correlation coefficient ( $r$ )  $> 0.7$ ,  $p$  values  $< 0.05$ ) except that the two-level hybrid clustering analysis was moderately correlated with SSC ( $r = 0.55$ ,  $p < 0.05$ ). The heat days identified by the 99th percentile minimum temperature were moderately correlated with those classified by other methods ( $r$  between 0.37 and 0.54). Supporting Information Table S5 shows characteristics of the seven SSC air masses calculated by Sheridan and colleagues for Detroit for the entire study period. Supporting information Figure S5 shows the number of annual dangerous heat wave days identified by SSC and the relative minimum temperature, and suggests the overall trends of minimum temperature and relative temperature were similar. Supporting information Figure

S6 further showed the associations among SSC, temperature and relative temperature methods. SSC, the 90th percentile minimum temperature and relative temperature were highly correlated.

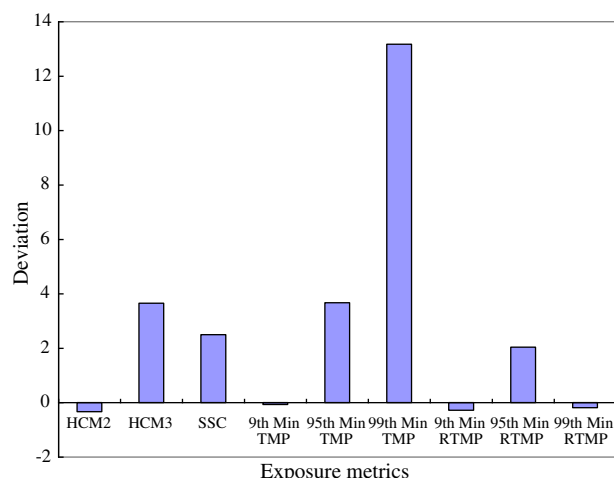
SSC tended to include the days with lower daily minimum temperature and dew point (shown in supporting information Table S6). The average minimum temperature and dew point of SSC-identified heat wave days were 2.5 and 5.3  $^{\circ}\text{C}$  lower than the same parameters on those days identified by other methods. As expected, the averages of four meteorological metrics increased with either the increased number of levels for hybrid clustering analysis or the increased percentile for minimum temperature.

Table 1 shows the days identified differently between SSC and other methods. Considering both Table 1 and Supporting Information Table S6, the number of commonly-identified 'hot' days between SSC and other metrics ranged from 21 to 151 days (0.5 to 3.2% of summer days over 31 years). SSC defined more than 200 heat wave days that were not classified as 'hot' by other methods. Compared to other metrics, both SSC and the 90th percentile minimum temperature include more heat wave days (355 and 405 days or 7.5% and 8.5% of summer days, respectively; Supporting Information Table S6). However, Table 1 shows these two methods shared relatively fewer heat wave days (151 days or 3.2%) in common. This is also true for two-level hybrid clustering analysis, which shared 61 days in common with SSC. Moreover, by comparing Table 1 and Supporting Information Table S6, the heat wave days identified only by SSC were associated with lower temperature/dew points than those days shared by SSC and other methods, e.g., daily minimum temperatures were 2 to 6  $^{\circ}\text{C}$  lower.

Fig. 3 shows the comparison of the exact dates of dangerous days identified by selected exposure metrics in HHWS triggers for the hottest summer (1995) during our study period. This figure shows these methods performed differently from another perspective. There were no identified dangerous hot days by all methods in May, 1995. In June, only SSC, the 90th percentile minimum temperature and relative temperature identified several dangerous hot days (3, 5 and 4 days, respectively), and they shared two days in common. Likewise, in July, all methods except for the three-level hybrid clustering analysis and the 99th percentile relative temperature classified a few extremely hot days, but there was no overlap among all methods. In August, all methods except for the 99th percentile relative temperature shared four days in common and differed for other identified days. In September, only SSC identified two dangerous days. In addition, the 90th percentile minimum temperature and relative temperature performed similarly and this was also true for the 95th percentile minimum temperature and relative temperature. Figures of such comparison are similar over other years. Figure S7 shows the comparison in 1988.



**Fig. 3.** Comparison of the exact dates of dangerous heat wave days identified by selected HHWS exposure metrics during summer of 1995 in Detroit, MI. (minimum temperature: Min TMP; minimum relative temperature: Min RTMP).



**Fig. 4.** Median values of daily deviation from expected all-cause mortality counts among days categorized as dangerously hot by different exposure metrics. (Expected mortality counts were derived by modeling daily all-cause mortality counts as a smooth function of date in Detroit Metropolitan area, Michigan (Detroit metropolitan area) from 1976 to 2006.)

Fig. 4 shows the median deviations for all-cause mortality in the extremely hot days identified by each method. The highest median positive deviation among these identified days is with the 99th percentile of minimum temperature (13 death counts per day). HCM3 and the 95th percentile of minimum temperature are both the second highest (3.7). SSC (2.5) is slightly lower than HCM3. HCM2 is the least (−0.3). Interestingly, the three exposure metrics based on percentiles of relative minimum temperature (accounting for time in season) had generally lower deviations than the other metrics.

#### 4. Discussion

This study proposed and demonstrated a flexible hybrid cluster analysis method for classifying potentially dangerous hot weather using recent advances in clustering methods. The flexibility provided by this hybrid clustering method allows users the freedom to establish local-specific HHWS triggers, using a somewhat more sophisticated approach than choosing an absolute heat index or percentile threshold calculated with reference to historical weather conditions. Also, compared with SSC, this method does not require expert meteorological knowledge and removes some subjectivity. In other words, others could apply the methods described here to replicate the results we found either using Detroit-area weather data or data from other localities.

The comparison of annual dangerous heat wave days identified by selected exposure metrics in HHWS triggers suggests these methods behaved quite differently in terms of the number of annual heat days identified, and which specific days were identified, although the temporal patterns of annual heat wave days were moderately or strongly correlated. This comparison suggests that there is enough of a difference that should matter for health, and some methods might work better than others in predicting days with higher deaths. This idea is corroborated by Hajat et al. (2010) who showed little agreement among several metrics in terms of their association with mortality. The level of disagreement reflects the difference in data inputs and analysis of these methods. SSC requires the most diverse range of weather information (cloud cover, pressure and other parameters), followed by hybrid clustering method and temperature. SSC and temperature are based on both meteorological and statistical principles, while hybrid clustering method purely relies on statistical models.

Based on our analysis using data from Detroit, Michigan, SSC tended to include days with lower daily minimum temperature and dew point and might be not sensitive to extreme heat waves. The identified heat days by SSC were usually associated with lower daily minimum temperature and dew point. SSC behaved differently from other methods in identifying heat wave days in the hottest month

(August, 1995) and the hottest summer (May to September, 1995) over the entire study period.

Hybrid clustering method has limitations including a lack of accounting for early season heat and the necessity of separately accounting for heat wave duration at the current stage. Heat waves occurring early in the summer have been shown to be more dangerous than those occurring in later summer, and risks also increase with longer heat duration (Anderson and Bell, 2010). The hybrid clustering method takes the time dimension into consideration separately because it is not easy to incorporate time scale automatically in the clustering procedure in a simple way. However, we could incorporate time-of-year and heat duration automatically in evaluating heat–mortality associations, similar to what is done when using the SSC exposure metric to determine when a heat warning or advisory should be issued.

The number of levels chosen for the clusters is subjective and depends on the user's purposes and experiences. We chose three levels for DTW because when we examined the temperatures of days in the two hottest sub-clusters (sample sizes are 169 and 158 days, respectively; Table S3) at the second level, they were not extremely hot. The model-based clustering method could not be used to further split these two sub-clusters because the number of clusters in the model-based clustering was determined by optimizing BIC automatically. Thus we applied PAM, which does not have this constraint, to further divide these two sub-clusters. Having this flexibility in picking the number of levels would allow users responsible for issuing heat advisories or warnings to tailor their decision according to their local conditions.

The minimum relative temperature method performed similarly to the minimum temperature method mainly because the relative temperature was calculated based on the minimum temperature relative to the 10-day reference period. Because of this stratification, relative temperature usually had lower thresholds in earlier or later months (e.g., May and September) and had higher thresholds in mid-summer months (e.g., July and August). These features of relative temperature can explain to a large extent why relative temperature identified more days in May and September and fewer days in July and August compared to temperature. In general, relative temperature seems more promising than temperature because it takes into account temporal variability of temperature and dew point. However, further evaluations using health outcome data in different geographical areas would be necessary to validate this supposition.

The health analysis points out that there is a difference between metrics that provide information on the extent of physiological stress and metrics that accurately predict when mortality will increase. It suggests that our proposed hybrid clustering method and SSC might better characterize physiological stress, which is necessary but not sufficient for predicting mortality. In addition, the 99th percentile of minimum temperature is associated with the highest mortality deviation mainly because of small sample size (22 days over 31 years). However, the 95th percentile of minimum temperature having similar deviation with HCM3 and SSC indicates a simple, easily understandable system could also be a good option for a HHWS. This finding needs to be further evaluated in the cities other than Detroit.

Our objective in this paper is not to design a new warning system, but to propose a new exposure metric which has potential to be incorporated into a HHWS design, with additional consideration of the quality of weather forecasts and associations with health outcomes. Although our new method has some promising features, it is not an alternative to SSC at the current stage. Further evaluation with data from more cities is needed, as well as testing of how well it works with forecast data and how it predicts health outcomes such as mortality. We plan to evaluate the performance of this proposed hybrid clustering method and other triggers within a HHWS context in a subsequent paper, which addresses the issues of sensitivity, specificity, generalization to other cities, and the influence of weather forecast quality (comparing triggers calculated using archived forecast *versus* observed data) on the alert system. In addition, we intend to evaluate performance of

these HHWS triggers in terms of ability to predict heat-related health outcomes using epidemiologic models and how their performance varies with type of health outcomes (e.g., total mortality and cause-specific mortality).

## 5. Conclusions

A novel hybrid multi-level clustering method based on four meteorological variables (daily minimum and maximum temperature, and daily minimum and maximum dew point) was proposed to identify extremely hot days, and was demonstrated using historical meteorological data from Detroit, Michigan. This method consists of a model-based clustering method and a partitioning clustering method (PAM). The number of levels and combination of the two types of clustering method are flexible and subject to users' demand. This multi-level method is further enhanced by some advantages of model-based clustering methods such as determining the number of cluster by an objective statistical criterion (BIC) and accounting for a more diverse range of cluster shapes. Comparison of identified extremely hot days among selected HHWS triggers suggests that these methods differ moderately or largely depending on method type; the 90th and 95th percentile minimum temperature and relative temperature performed similarly; and SSC tended to define more heat wave days with lower daily minimum temperature and dew point. Compared with SSC, our proposed hybrid clustering method reduces the possibility of including days with low temperature and dew points, and avoids including too many days. Results justified the usability of hybrid clustering method as a sequential application of clustering methods within the context of heat waves. It also provides the flexibility and transparency to define dangerous heat. The health analysis indicates the metrics that are potentially better in quantifying physiological heat stress do not necessarily result in better prediction of excess daily mortality. Thus, a simple metric (e.g., the 95th percentile daily minimum temperature) might be a good metric for HHWSs if we use the deviation from typical daily mortality counts as an evaluation criterion for HHWSs. This raises a question for future research: how to evaluate HHWSs? Our proposed approach and findings are relevant to heat wave prevention, heat-related epidemiological studies and risk assessment, and potentially pollutant mixture research.

## Acknowledgments

The research described in this paper was funded through support of the Graham Environmental Sustainability Institute at the University of Michigan; the U.S. Environmental Protection Agency Science to Achieve Results (STAR) grant R832752010; the U.S. Centers for Disease Control and Prevention grant R18 EH 000348 and National Institute for Environmental Health Sciences grant ES-016932.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.envint.2012.05.001>.

## References

- Anderson GB, Bell ML. Heatwaves in the United States, mortality risk during heatwaves and effect modification by heatwave characteristics in 43 US communities. *Environ Health Perspect* 2010;119:210–8.
- Barriopedro D, Fischer EM, Luterbacher J, Trigo RM, García-Herrera R. The hot summer of 2010, redrawing the temperature record map of Europe. *Science* 2011;332:220–4.
- Bower D, McGregor GR, Hannah D, Sheridan SC. Development of a spatial synoptic classification scheme for western Europe. *Int J Climatol* 2007;27:2017–40.
- BuJa A, Lang DT, Swayne DF. GGobi, evolving from XGobi into an extensible framework for interactive data visualization. *Comput Stat Data Analysis* 2003;43(4):423–44.
- Fraley C, Raftery A. Model-based methods of classification, using the mclust software in chemometrics. *J Stat Softw* 2007;18(6):1–13.
- Hajat S, Sheridan SC, Allen MJ, Pascal M, Laaidi K, Yagouti A, et al. Heat–health warning systems: a comparison of the predictive capacity of different approaches to identifying dangerously hot days. *Am J Public Health* 2010;100(6):1137–44.
- Izenman AJ. Modern multivariate statistical techniques, regression, classification, and manifold learning. Springer Science + Business Media. New York: LLC; 2008.
- Kalkstein LS, Greene JS. An evaluation of climate/mortality relationships in large U.S. cities and the possible impacts of a climate change. *Environ Health Perspect* 1997;105:84–93.
- Kalkstein L, Greene S, David M, Samenow J. An evaluation of the progress in reducing heat-related human mortality in major U.S. cities. *Nat Hazards* 2011;56(1):113–29.
- Kovats RS, Ebi KL. Heat waves and public health in Europe. *Eur J Public Health* 2006;16(6):592–9.
- Kovats RS, Hajat S. Heat stress and public health, a critical review. *Annu Rev Public Health* 2008;29:41–55.
- Matthies F, Menne B. Prevention and management of health hazards related to heat waves. *Int J Circumpolar Health* 2009;68(1):8–22.
- Meehl GA, Tebaldi C. More intense, more frequent, and longer lasting heat waves in the 21st century. *Science* 2004;305:994–7.
- Metzger KB, Ito K, Matte TD. Summer heat and mortality in New York city, how hot is too hot? *Environ Health Perspect* 2010;118(1):80–6.
- National Weather Service. Heat wave: a major summer killer. [http://www.nws.noaa.gov/om/brochures/heat\\_wave.shtml](http://www.nws.noaa.gov/om/brochures/heat_wave.shtml)2005. [accessed 25th May 2010].
- O'Neill M, Sampson N, McCormick S, Rood RB, Buxton M, Ebi KL, et al. The heat is on: decision-maker perspectives on when and how to issue a heat warning. American Geophysical Union 2011 Fall Meeting, December 5–9, San Francisco; 2011.
- R Development Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing3-900051-07-0; 2006. See <http://www.R-project.org> [accessed 2 June 2011].
- Rocklöv J, Ebi K, Forsberg B. Mortality related to temperature and persistent extreme temperatures—a study of cause-specific and age stratified mortality. *Occup Environ Med* 2011;68(7):531–6.
- Sheridan S. The redevelopment of a weather-type classification scheme for North America. *Int J Climatol* 2002;22:51–68.
- Sheridan S. Spatial synoptic classification data. <http://sheridan.geog.kent.edu/ssc.html>2009. [accessed May 27th 2009].
- Sheridan SC, Kalkstein LS. Progress in heat watch/warning system technology. *B Am Meteorol Soc* 2004;85(12):1931–41.
- Struyf A, Hubert M, Rousseeuw P. Clustering in an object-oriented environment. *J Stat Softw* 1997;1(4):1–30.
- Wood SN. Fast stable direct fitting and smoothness selection for generalized additive models. *J R Stat Soc Series B Stat Methodol* 2008;70:495–518.