Development of Objective Air-Mass Classifications for Studies of Heat-Related Mortality in Central Europe and East Asia

Kysely, J. [1], J. Kim [2], R. Huth [1] and B.-C. Choi [2]

[1] Institute of Atmospheric Physics AS CR, Prague, Czech Republic
[2] Korea Meteorological Administration, Seoul, South Korea
Email: kysely@ufa.cas.cz

Keywords: *Biometeorology, human mortality, air mass classifications, central Europe, east Asia*

EXTENDED ABSTRACT

The most direct effects of weather on the human health in mid-latitude areas are observed during and after summer heat waves. The present study aims at developing objective classifications of weather types (usually referred to as 'air masses') utilized for evaluating and predicting excess mortality due to heat stress. The air-mass-based synoptic approach takes into account the entire weather situation rather than single elements, and it has been recognized as a particularly useful tool for the identification of weather conditions most strongly associated with impacts on mortality in many parts of the world.

The applicability of the method is compared for populations in central Europe (the Czech Republic) and east Asia (South Korea). The determination of air masses is based on common meteorological variables measured several times a day at a number of sites across each region; principal component analysis and cluster analysis are used to define the air masses. In the present paper, specific attention is paid to the selection of meteorological input data and the clustering techniques (hierarchical vs. non-hierarchical approaches). 'Oppressive' types are identified among the objectively determined air masses as those associated with elevated mortality. As the variance in mortality within the oppressive air masses is large, a regression analysis is needed to find meteorological conditions responsible for mortality increases in a consequent step of the study.

Results demonstrate that the air-mass-based approach is a useful alternative to traditional methods in assessing impacts of heat stress on mortality in both areas, and may be applicable also for predicting health outcomes and increased mortality in heat-watch-warning systems. A comparison of the mortality impacts reveals varied responses to oppressive weather between the populations living in different environments (but under comparable socio-economic conditions) and helps to better understand the roles of various societal, cultural and habitual factors in heat-related risks.

1. INTRODUCTION

Evaluation of meteorological causes of extreme heat waves and assessments of their impacts on various fields of human activities, including human health, are topical issues in scientific literature. The interest is at least partly related to the recently observed increases in the frequency and severity of warm temperature extremes over many parts of the world (Moberg *et al.* 2006, Nicholls and Alexander 2007) as well as a growing concern of their future changes related to global warming (IPCC 2007) and associated adverse effects.

Increasing trends in temperature extremes over recent decades, with record-breaking heat waves in the 1990s and 2000s, have been observed in both areas under study: the Czech Republic (central Europe) and South Korea (east Asia) (Figure 1). Both countries lie in mild climate zone, with a mixture of continental and maritime influences leading to relatively cold winters and hot summers. Variability of weather on synoptic time scales is large in both regions, and hot weather conditions in summer occur almost every year.

In terms of mortality impacts, heat has been recognized as the most deadly among all atmospheric hazards (e.g. Sheridan and Kalkstein 2004). This is particularly true for extratropical areas where large seasonal temperature changes and high within-season variability of weather cause that the populations are prone to heat-related mortality. However, improved understanding of ‘oppressive’ weather conditions, namely development of methods by which it is determined whether a day will be oppressive, and a subsequent implementation of heat-watch-warning systems that take action when an oppressive day is forecasted, have decreased the numbers of victims due to heat stress in many mid-latitudinal locations (e.g. Kalkstein *et al.* 1996a, Ebi *et al.* 2004).

Figure 1. Areas under study and locations of meteorological stations.

The present study focuses on a development of *objective classifications of weather types* (usually referred to as ‘*air masses*’, AMs) in human mortality studies in the two regions. The ‘air-mass-based’ (termed also ‘synoptic’) approach takes into account the entire weather situation rather than single elements and links mortality to objectively determined AMs; it is an alternative to traditional methods for the evaluation of weather-mortality links in which mortality impacts are assessed using single variables or their combinations in biometeorological indices. The synoptic approach has been recognized as a particularly useful tool for the identification of weather conditions most strongly associated with historical increases in mortality in the US (Kalkstein *et al.* 1996a, Smoyer *et al.* 2000) but its applications involve also Australia (Guest *et al.* 1999), China (Tan *et al.* 2004), central Europe (Kyselý and Huth 2004) and southern Europe (Cegnar and Kalkstein 2000). The AM classification methodology is described in detail in Section 3.

2. MORTALITY DATA

Daily data on all-cause (total) mortality in the Czech Republic (population of about 10 million inhabitants) are available over the period 1986-2005; in South Korea (population of about 47 million inhabitants) over 1991-2005. To account for long-term changes in mortality (related to demographic, medical-technological and life-style changes) as well as the seasonal and weekly cycles, the daily death counts have been standardized; series of excess daily total mortality are calculated as deviations of the observed number of deaths from the expected number of deaths for each day of the examined period, using the methodology described in Kyselý (2004).

3. OBJECTIVE CLASSIFICATIONS OF AIR-MASSES

3.1. Input meteorological data

Air temperature, dew-point deficit, zonal wind, meridional wind and total cloud amount are used as input variables in the classification of weather types (‘air masses’, AMs). This suite of weather elements is similar to other studies (e.g. Kalkstein *et al.* 1996a, Smoyer *et al.* 2000). Datasets available consist of measurements at 7 stations in the Czech Republic and 10 stations in South Korea (Figure 1), carried out 3 times a day in the Czech Republic (at 7, 14 and 21 LT) and 4 times a day in South Korea (at 3, 9, 15 and 21 LT). The period covered is from mid-May to mid-September (i.e. the 4-month period when most heat waves occur).

Three kinds of the AM classification that differ in the way how stations’ datasets are taken into account are performed: The input variables originate from (i) the *average* series for the area under study (both areas are comparable in size and relatively small, which justifies the use of average series); (ii) the *pooled* series at all 7 (10) Czech (Korean) stations considered together; and (iii) the series for a *single station* (Prague/Seoul). We examine whether there are systematic differences among the resulting AM classifications and their relationships to excess mortality.

The AM classification that leads to the daily catalogue of AM types termed ‘Temporal Synoptic Index’ (TSI) is applied since the areas under study are small (Figure 1), and one index value is usually representative for the whole countries. We point to the fact that over large (continental and sub-continental) areas, ‘Spatial Synoptic Classification’ (SSC; Kalkstein *et al.* 1996b) may be more useful, although its drawback is that representative days (seed days) must be identified manually, involving a large degree of subjectivity (e.g. Bower *et al.* 2007).

3.2. Classification methodology

Since the input variables are mutually dependent, unrotated principal component analysis (PCA) is performed first to reduce their number and to form a set of new orthogonal variables (PCs). In all cases (averaged, pooled and single-station input data), time series of 4 leading PCs enter the cluster analysis. The number of PCs conforms with the criteria recommended in literature (e.g., Richman 1986); in particular, there is a large gap in explained variance between the 4th and 5th PC.

Two basic approaches to the cluster analysis, *hierarchical* (‘between groups average linkage’) and *non-hierarchical* (‘k-means’), are employed and compared. The average linkage method (used e.g. in Kalkstein *et al.* 1996a and Smoyer *et al.* 2000) suffers from the undesirable ‘snowball effect’; to remove this effect, we utilize the approach of terminating the clustering procedure at different dissimilarity levels in various parts of the dendrogram, similarly to Huth *et al.* (1993). Hierarchical cluster analysis was applied for averaged input data only; individual ‘reasonable’ classifications, yielding 19 and 13 (16 and 11) AMs for the Czech Republic (South Korea), are formed by merging clusters from the ‘basic’ classification, defined by termination at a single dissimilarity level in the whole dataset. Two (one) out of the clusters of the above-mentioned classifications appear only on 1 or 2 days over the examined period in the Czech Republic (South Korea) and are disregarded from a further analysis, so the actual number of AMs is 17 and 11 (15 and 10) in the Czech Republic (South Korea).

Unlike average linkage, the k-means method yields clusters of a comparable size; its disadvantage is that they are less compact (i.e. their within-cluster variability is relatively high). The preliminary number of clusters was set a priori to 6, 10 and 15 so that it spans a range around values appearing in the literature. The non-hierarchical cluster analysis was applied to averaged, pooled, and single-station input data.

In the present application, *oppressive AMs* are defined as those associated with a mean mortality increase of at least 3% relative to the baseline mortality. If more than one AM is found oppressive, the one with the largest mean mortality increase is termed ‘main’ while the other oppressive AMs are ‘secondary’ (cf. Kyselý and Huth 2004). We decided not to establish the delimitation of AMs that are considered oppressive on rigorous statistical testing of mean excess mortality (using e.g. t-test or Mann-Whitney test) since its results are overly influenced by (highly variable in case of the hierarchical cluster analysis) sample sizes of individual AMs, and may lead to counter-intuitive and inadequate conclusions. Lags of mortality effects after weather of 0 to 2 days are considered and the results are compared.

4. RESULTS

4.1. Czech Republic

Non-hierarchical cluster analysis (NHCA) with 6 and 10 clusters yields one oppressive AM with enhanced mortality (Table 1a, Figure 2); the NHCA with 15 clusters yields two oppressive AMs (the main one and the secondary one, cf. Kyselý and Huth 2004). The mean excess mortality in the oppressive AMs is more pronounced when pooled input meteorological data are used (compared to averaged and particularly single-station data, the latter being the least suitable option). It should be noted that mean excess mortality is negative or close to zero in all other non-oppressive AMs (Figure 2), which indicates a good skill of the classification procedure in identifying weather conditions associated with mortality impacts.

Hierarchical cluster analysis (HCA) appears to be somewhat less successful in determining oppressive weather conditions. The HCA with 11 clusters yields one oppressive AM while the HCA with 17 clusters leads to three oppressive AMs (the main one and two secondary ones), but they are associated with less pronounced or at best comparable mean mortality increases relative to the NHCA, despite much smaller relative frequencies of the oppressive AMs (Table 1a). (The smaller frequencies of the oppressive AMs obtained by the HCA mean that a relatively large percentage of days with enhanced mortality are not captured by the oppressive AMs.)

In all classifications, the (main) oppressive AM is the warmest one among all AMs; it is associated with a below-average cloud amount (but usually not the smallest one among AMs). Also the secondary oppressive AMs are associated to above-average temperatures, but their other properties differ in individual classifications: In NHCA with 15 clusters, the secondary oppressive AM is characterized by prevailing southeasterly flow (while the main one by the southwesterly flow) and smaller cloud amount; in HCA with 17 clusters, both secondary oppressive AMs are moister and more cloudy compared to the main one. The main oppressive AM in all classifications may be termed ‘dry tropical’ as the dew-point deficit is large (on average reaching or exceeding 15ºC at 14 LT in all classifications) and the average dew-point temperature (a measure of absolute humidity) is comparable to other AMs. ‘Moist tropical’ AM is separated only in the HCA with 17 clusters and appears to be split into two secondary oppressive AMs.

Table 1. Basic characteristics of the oppressive AMs of individual classifications. NHCA-X (HCA-X) denotes non-hierarchical (hierarchical) cluster analysis with X clusters, with pooled (averaged) input meteorological data.

a. Czech Republic, unlagged relationship

	Main oppressive AM		Secondary oppressive AM	
	Mean excess mortality	Relative freq. [%]	Mean excess mortality	Relative freq. [%]
NHCA-6	17.9	16.6	---	---
NHCA-10	24.2	7.2	---	---
NHCA-15	26.4	5.7	12.3	6.5
HCA-11	22.2	2.6	---	---
HCA-17	26.8	2.1	14.9	9.5
			12.2	4.8

b. South Korea, unlagged relationship

	Main oppressive AM		Secondary oppressive AM	
	Mean excess mortality	Relative freq. [%]	Mean excess mortality	Relative freq. [%]
NHCA-6	---	---	---	---
NHCA-10	20.1	13.6	---	---
NHCA-15	26.7	7.4	---	---
HCA-10	33.3	0.5	---	---
HCA-15	33.3	0.5	28.6	6.6

Table 2a demonstrates that days with mortality exceeding by at least 5% and 10% the estimated baseline mortality are much better covered by the oppressive AMs from the NHCA, which is a particularly strong indicator of its superiority; a large number of days that are in fact associated to elevated mortality are left out from the main oppressive AMs of the HCA.

Table 2. Percentage of days with enhanced mortality in the period May 16-September 15 (exceeding by at least 5% and 10% the estimated baseline mortality; EM05, EM10) covered by the oppressive AMs of individual classifications.

a. Czech Republic

	Main oppressive AM		Secondary oppressive AM	
	EM05 [%]	EM10 [%]	EM05 [%]	EM10 [%]
NHCA-6	35.9	42.8	---	---
NHCA-10	17.8	24.8	---	---
NHCA-15	15.1	21.4	11.2	13.0
HCA-11	5.6	7.6	---	---
HCA-17	5.3	7.6	18.4	22.9
			8.0	10.7

b. South Korea

	Main oppressive AM		Secondary oppressive AM	
	EM05 [%]	EM10 [%]	EM05 [%]	EM10 [%]
NHCA-6	---	---	---	---
NHCA-10	21.7	25.9	---	---
NHCA-15	14.1	15.5	---	---
HCA-10	1.1	1.7	---	---
HCA-15	1.1	1.7	12.0	13.2

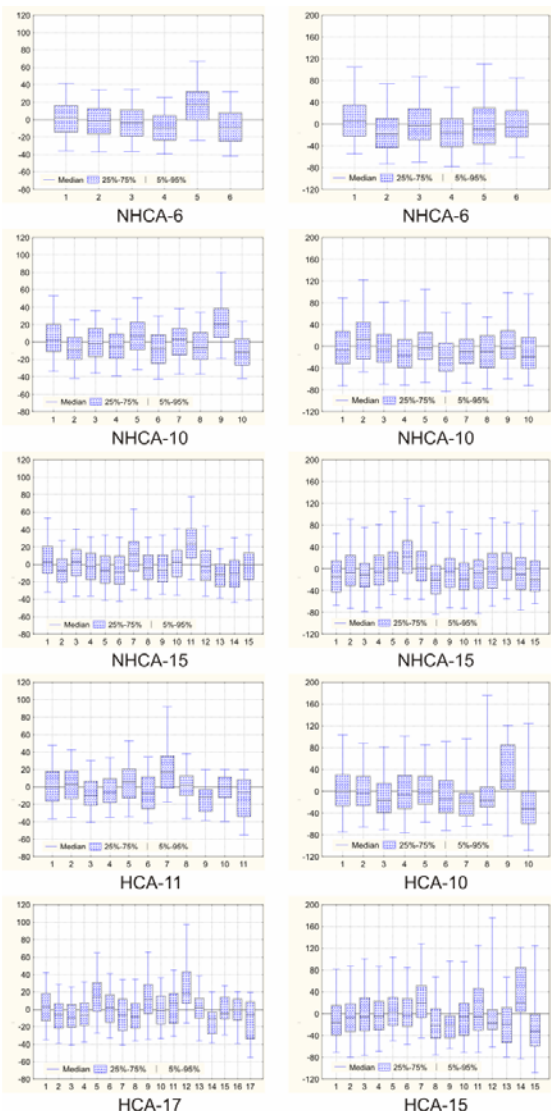

Figure 2. Box-plots of excess total mortality in individual AMs of the examined classifications. AMs are ranked according to their frequency of occurrence. NHCA-X (HCA-X) denotes non-hierarchical (hierarchical) cluster analysis with X clusters. Left: Czech Republic, right: South Korea.

If 1-day lag of mortality impacts after weather is considered the mortality response in oppressive

AMs is weaker. This is particularly the case of the classifications that make use of HCA; the mean mortality increase in the main oppressive AM is 10-14 deaths daily compared to 22-27 deaths daily if unlagged relationships are examined. The fact that the number of secondary oppressive AMs increases in some classifications is another unfavourable feature as they are not associated to specific stressful weather conditions (measured by temperature and humidity). The association between oppressive AMs and mortality becomes even weaker for 2-day lag of mortality after weather.

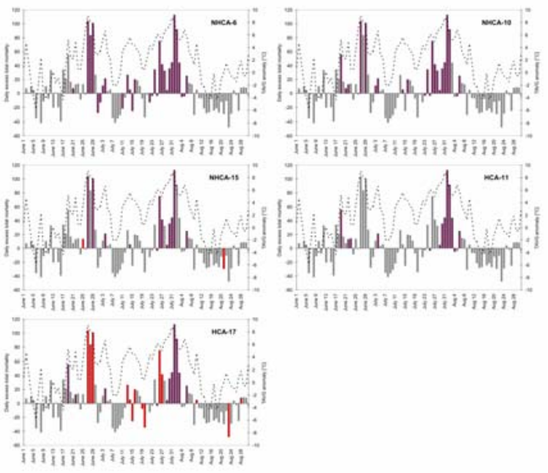

Figure 3. Daily excess total mortality (columns, left axis) and anomalies of average daily air temperature (TAVG, dashed curve, right axis) during summer 1994 in the Czech Republic. Days with oppressive AMs are marked by red columns with black margins (dark red: main oppressive AM, light red: secondary oppressive AMs). NHCA-X (HCA-X) stands for non-hierarchical (hierarchical) cluster analysis with X clusters.

An example of results of the objective AM classifications is shown in Figure 3 for the hot summer of 1994 when the most severe heat waves ever recorded in the area of the Czech Republic occurred (e.g. Kyselý 2002). A consecutive period of above-normal average daily air temperatures was observed from mid-June till mid-August (dashed curve in Figure 3), with two severe heat waves in late June and late July-early August. Both were associated with pronounced mortality impacts, reaching more than 100 excess death on the peak days (more than 30% relative increase), and altogether 481 excess deaths for the period June 17-30 and 558 excess deaths for July 24-August 8 (corresponding to 10-12% relative increases in both heat waves). All classifications yield oppressive AMs during the heat waves. Among NHCA, the classification which recognizes 6 AMs predicts the oppressive AM most frequently, covering almost all days with pronounced mortality impacts; the NHCA-10 classification appears to be a possible suitable alternative as it very efficiently reduces the number of days with the oppressive AM that are associated with below-average mortality (4 in NHCA-10 compared to 10 in NHCA-6). Some days with strong mortality response lie outside oppressive AMs in all other classifications. Particularly unsuitable appears to be the HCA with 11 AMs, in which the late June heat wave is classified with non-oppressive AMs; the HCA with 17 AMs yields better results.

4.2. South Korea

Non-hierarchical cluster analysis (NHCA) with 6 clusters does not yield an oppressive AM; the NHCA with 10 and 15 clusters lead to one oppressive AM with enhanced mortality if averaged input meteorological data are used (Table 1b, Figure 2). However, the mean relative mortality increase in the oppressive AMs (3 to 4%) is smaller compared to the oppressive AMs of analogous classifications in the Czech Republic (8 to 9%). Oppressive AMs entirely disappear if single-station (Seoul) input data are used, and they become associated with less pronounced mortality impacts for pooled input data. The fact that the mean excess mortality is negative or close to zero in all other non-oppressive AMs (Figure 2) is, nevertheless, an indication of some skill of the classification procedure in identifying weather conditions associated with mortality impacts.

Hierarchical cluster analysis (HCA) is even less successful in determining oppressive weather conditions. Although the oppressive AM of both HCA classifications (with 10 and 15 clusters) is associated with a mean relative excess mortality of 5.3%, it appears with a frequency of only 0.5% (less than 1 day per summer; Table 1b) which means that it covers a negligible fraction of days with pronounced excess mortality (Table 2b). A secondary oppressive AM occurs in the HCA with 15 clusters, but the percentage of days with mortality exceeding by at least 5%/10% the estimated baseline mortality is still very low when the main and secondary oppressive AMs are taken together (Table 2b).

In the NHCA classifications with 10 and 15 clusters, the oppressive AM is the warmest one among all AMs; it is associated with a mean wind speed close to zero, a below-average cloud amount (but not the smallest one among AMs) and a relatively large humidity (dew-point temperatures are much higher than in the oppressive AMs in the

Czech Republic, and dew-point deficit does not exceed 10ºC at 15 LT). The oppressive AM of the HCA classifications (which occurs with an extremely low frequency) is characterized by below-average temperatures and cloudy weather with strong northeasterly flow, so the mortality is not heat-related. In the HCA with 15 clusters, the hottest AM is in fact the secondary oppressive AM, associated with calm and humid weather with below-average cloud amount, and resembling in properties the oppressive AM of the NHCA classifications. We term this AM ‘moist tropical’ one.

Days with enhanced mortality are best covered by the oppressive AM from the NHCA with 10 clusters (Table 2b), which is an indicator of its superiority; however, a large fraction of days that are in fact associated to enhanced mortality is omitted.

If 1-day lag of mortality impacts after weather is considered, the mortality response in oppressive AMs is weaker (except for a rise in mean mortality in the oppressive AM of the HCA classifications, which occurs with an extremely low frequency). The association between oppressive AMs and mortality becomes very weak for 2-day lag of mortality after weather.

5. DISCUSSION AND CONCLUSIONS

The relatively large differences between properties of oppressive AMs in the Czech Republic and South Korea arise from climatological characteristics of the two areas, related among others to entirely different sea-land configuration. Over Europe, thermal properties of the moist tropical AM (originating over the subtropical North Atlantic) are moderated (when travelling northward and northeastward) by the cold southward Canary current, and the maritime AM is usually transformed to dry tropical AM before reaching central Europe. That is why ‘dry tropical AM’ is generally found to be the oppressive AM in all classifications. Over East Asia, the moist tropical AM (originating over the subtropical Northwestern Pacific) is modified (when travelling northward) by ocean surface due to the warm northward flowing Kuroshio current, and its properties (as to temperature and humidity) are quite distinct when reaching the Korean peninsula.

Based on the AMs-mortality links, two different classifications are selected for further examination of the mortality impacts within oppressive AMs in the population of the Czech Republic: NHCA with 6 to 10 AMs and HCA with 17 AMs. The main reason is that the oppressive AMs of these classifications cover a relatively large fraction of days with enhanced mortality, which is a useful feature as to possible predictive skills of health outcomes due to stressful weather. The number of clusters in the NHCA will be refined by means of a consequent set of experiments with the cluster analysis. In South Korea, the NHCA classification with 10 AMs is preferred for further examination of the mortality impacts within oppressive AMs; the HCA is not suitable for the identification of oppressive AMs in this region.

The finding that the link is strongest for unlagged data is in accord with Kyselý and Huth (2004) who found out strongest relationships between mortality and temperature when zero lag is considered.

Relationships between AMs and total mortality are notably weaker in the population of South Korea compared to the Czech Republic. This may be related to different demographic patterns (much larger percentage of population lives in large cities in South Korea than the Czech Republic) and associated confounding air-pollution effects; however, it is also possible that the variance of excess mortality within the oppressive AMs in the population of South Korea is largely due to meteorological and acclimation factors, including timing within a season (a larger mortality response is usually observed in early than late summer) and timing within a spell of hot/oppressive days (mortality effects tend to decrease with time during persistent heat waves).

In a following step of the research into weather-mortality links in both countries, regression analysis will be performed to find meteorological conditions responsible for mortality increases, since the variance in mortality within the oppressive air masses is large, and predictive skills of such models will be examined. These relations may be found useful also for the development of heat-watch-warning systems that are in operation at a growing number of places around the world (e.g. Ebi *et al.* 2004, Sheridan and Kalkstein 2004).

ACKNOWLEDGEMENTS: The study is supported by the Czech Science Foundation and Korea Research Foundation under projects 205/07/J044 and KRF-2006-C00005. Thanks are due to the staff of the National Institute of Public Health, Prague, the Institute of Health Information and Statistics, Prague, and the Korea National Statistical Office, Seoul, for providing daily mortality datasets; and L.Gaál and J.Hošek for assistance in drawing figures.

REFERENCES

Bower, D., G.R. McGregor, D.M. Hannah and S.C. Sheridan (2007), Development of a spatial synoptic classification scheme for western Europe, *International Journal of Climatology*, doi 10.1002/joc.1501 (in press).

Cegnar, T. and L.S. Kalkstein (2000), Heat watch/warning system in Rome – results and implementation. In: *Proc. 3rd International Conference on Applied Climatology* [CD-ROM], Pisa, Italy.

Ebi, K., T.J. Teisberg, L.S. Kalkstein, L. Robinson and R.F. Weiher (2004), Heat watch/warning systems save lives. *Bulletin of American Meteorological Society*, 85, 1067–1073.

Guest, C.S., K. Wilson, A. Woodward, K. Hennessy, L.S. Kalkstein, C. Skinner C and A.J. McMichael (1999), Climate and mortality in Australia: retrospective study, 1979-1990 and predicted impacts in five major cities in 2030, *Climate Research*, 13, 1–15.

IPCC (2007), Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the IPCC Fourth Assessment Report, Cambridge University Press, Cambridge.

Huth, R., I. Nemešová and N. Klimperová (1993), Weather categorization based on the average linkage clustering technique: An application to European mid-latitudes, *International Journal of Climatology*, 13, 817–835.

Kalkstein, L.S., P.F. Jamason, J.S. Greene, J. Libby and L. Robinson (1996a), The Philadelphia hot weather-health watch/warning system: development and application, summer 1995, *Bulletin of American Meteorological Society*, 77, 1519–1528.

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

Kyselý, J. (2002), Temporal fluctuations in heat waves at Prague-Klementinum, the Czech Republic, from 1901-1997, and their relationships to atmospheric circulation. *International Journal of Climatology* 22, 33–50.

Kyselý, J. (2004), Mortality and displaced mortality during heat waves in the Czech Republic, *International Journal of Biometeorology*, 49, 91–97.

Kyselý, J. and R. Huth (2004), Heat-related mortality in the Czech Republic examined through synoptic and 'traditional' approaches, *Climate Research*, 25, 265–274.

Moberg, A., P.D. Jones, D. Lister (2006), Indices for daily temperature and precipitation extremes in Europe analyzed for the period 1901-2000, *Journal of Geophysical Research*, 111, D22106

Nicholls, N. and L. Alexander (2007), Has the climate become more variable or extreme? Progress 1992-2006, *Progress in Physical Geography*, 31, 77–87.

Richman, M.B. (1986), Rotation of principal components, *Journal of Climatology*, 6, 293–335.

Sheridan, S.C. and L.S. Kalkstein (2004), Progress in heat watch-warning system technology, *Bulletin of American Meteorological Society*, 85, 1931–1941.

Smoyer, K.E., L.S. Kalkstein, J.S. Greene and H. Ye (2000), The impacts of weather and pollution on human mortality in Birmingham, Alabama and Philadelphia, Pennsylvania, *International Journal of Climatology*, 20, 881–897.

Tan, J., L.S. Kalkstein, J. Huang, S. Lin, H. Yin and D. Shao (2004), An operational heat/health warning system in Shanghai, *International Journal of Biometeorology*, 48, 157–162.