18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

Application of objective classifications of 'air masses' in modelling heat-related mortality in Korea

Kysely, J. [1], **R. Huth**[1], **L. Gaal**[1] **and J. Kim**[2]

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

Abstract: Objective classifications of weather types ('air masses', AMs) are frequently used for evaluating and predicting increased mortality due to heat stress in summer. The air-mass-based approach takes into account the entire weather situation rather than single elements; it identifies 'oppressive' AMs associated with elevated mortality in a given location/area, and applies regression models within the oppressive AMs in order to account for excess mortality. Principal component analysis and cluster analysis are used to define the AMs from basic input meteorological data.

The present study examines the applicability of the methodology for the area and population of South Korea. After presenting basic characteristics of the oppressive AMs of two selected classifications, recognized as superior in terms of the mean relative mortality increase in the oppressive AM and the coverage of days with large excess mortality by the oppressive AM, we focus on the development of regression models that account for variations in excess mortality within the oppressive AMs. Both meteorological and non-meteorological parameters are found to be important predictors in regression models for excess mortality within the oppressive AMs of the two classifications.

The results suggest that the coverage of days with large excess mortality by the oppressive AM is the most important criterion for the application into predicting elevated mortality risks. The classification with a relatively small number of AMs (6) shows better predictive skills in spite of smaller mean excess mortality on days classified with the oppressive AM. The relationships of excess mortality within the oppressive AM to meteorological and non-meteorological factors may be found useful also for the development of a heat-watch-warning system that is currently tested for Seoul and other large Korean cities.

The study is supported by the Korea Research Foundation and the Czech Science Foundation under joint project KRF-2006-C00005 / 205/07/J044.

Keywords: *air mass classifications, human mortality, biometeorology, heat stress, Korea*

1. INTRODUCTION

The most direct effects of weather on human health in mid-latitudes are observed during and after summer heat waves that lead to significant increases and intraseasonal shifts in mortality and morbidity (e.g. Koppe et al., 2004). Objective classifications of weather types, usually referred to as 'air masses' (AMs), have become widely-used in biometeorological studies that aim at identifying links between weather conditions and human health/mortality. The AM-based approach takes into account the entire weather situation rather than single elements; it identifies 'oppressive' AMs associated with elevated mortality in a given location/area, and applies regression models within the oppressive AMs in order to account for (and predict) excess mortality. (The definition of the oppressive AMs is not unique in literature; herein, the oppressive AMs are defined as those associated with mean excess mortality significantly different from zero, with the mean increase at least 3% relative to the baseline mortality.)

The idea behind the AM-based approach is that human physiology responds to the whole 'umbrella of air' and not single weather elements. The applications of AM classifications include analyses of historical datasets as well as the development of heat-watch-warning systems (HWWS; e.g. Kalkstein et al., 1996, Sheridan and Kalkstein, 2004) which determine whether a day, according to weather forecast, is likely to be associated with elevated mortality risks due to heat, and take action when an oppressive day is predicted.

The present study deals with mortality in the population of South Korea and examines applicability of two classifications leading to daily catalogues of AM types known as 'Temporal Synoptic Index' (Kalkstein, 1991); they are described in Section 2.2.

2. DATA AND METHODS

2.1 Mortality data

Daily data on total (all-cause) mortality in South Korea (population of 47.3 million in 2005) are examined over 15-yr period 1991-2005. The daily death counts are standardized to account for long-term trends in mortality (related mainly to demographic and health care changes), the seasonal and the weekly cycles; the procedure leads to daily series of 'excess mortality', i.e. deviations of recorded and expected (baseline) numbers of deaths for each day. The study is based on revised mortality datasets; see Kysely and Kim (2009) for details on the standardization procedure. Days with large accidental/disaster-related death toll were removed from the analysis. Excess mortality is examined in the whole population (all ages) and separately also in the elderly (persons aged 70+ years) which form the most vulnerable group according to many studies on weather-related mortality.

2.2 Air mass classifications

We make use of two AM classifications that were recognized as superior in a large set of classifications (differing in the input meteorological variables, the number of principal components [PCs] retained, and the number of clusters formed). The examined classifications are based on two variables, air temperature and dew-point deficit, pooled input data at 10 stations covering South Korea, and 2 retained PCs; non-hierarchical k-means cluster analysis was used to form the AMs since it was superior to hierarchical average-linkage clustering for the purpose of the study. The two examined classifications differ in the number of AMs (6 and 15); they are performed on data from mid-May to mid-September (4-month extended summer seasons) of 1991-2005.

Properties of the oppressive AMs of these two classifications, and their applicability into modelling and predicting elevated mortality risks in the population of South Korea are compared.

2.3 Regression models for excess mortality within the oppressive AM

To evaluate the impact of within-AM variations in meteorological elements, a stepwise multiple regression analysis was performed on all days classified with the oppressive AM of the selected classifications, with relative daily excess mortality in the whole population and the elderly (70+ years) as the dependent variables. The independent variables (Table 1) included weather elements as well as non-meteorological factors (day in sequence, time of season; Kalkstein, 1991; Sheridan and Kalkstein, 2004). Year was considered as another non-meteorological variable, to account for possible long-term changes in vulnerability to heat stress (possibly related to positive socio-economic development and better living conditions of most parts of the population; cf. Davis et al., 2003). Finally, we include also the numbers of days with the oppressive AM since the beginning of summer and in previous summer as possible predictors that may account for shorter-term and longer-term acclimation to oppressive weather conditions. Meteorological variables lagged by 1 and 2 days (t-1, t-2) and changes over 24h periods (d/dt) were also involved as possible predictors.

Table **1**. Independent variables in step-wise regression models for relative daily excess mortality within the oppressive AM.

Abbreviation	Variable
T3, T9, T15, T21	air temperature at 3, 9, 15 and 21 LT
Td3, Td9, Td15, Td21	dew-point temperature at 3, 9, 15 and 21 LT
HI3, HI9, HI15, HI21	heat index at 3, 9, 15 and 21 LT
T_AVG	average daily temperature
Td_AVG	average daily dew-point temperature
HI_AVG	average daily heat index
TCA_AVG	average daily total cloud amount
WS_AVG	average daily wind speed
X_1	variable X lagged by 1 day
X_2	variable X lagged by 2 days
dX	change in X over 24 hours, i.e. X-(X-1)
TOS	time of season (counter variable starting on June 1 = 1 in each season)
DIS	day in sequence in a spell of the oppressive AM (counter variable starting on the first day of a spell)
YEAR	time in years
OAM	number of days with the oppressive AM in previous summer
OAMy	number of days with the oppressive AM since the beginning of summer

3. RESULTS

3.1 Excess mortality and oppressive AMs

Box-plots of relative excess mortality in individual AMs of the two classifications are plotted in Figure 1. The classification that recognizes 15 AMs (C15 hereafter) yields the oppressive AM with very large mean relative excess mortality (6.7% in the whole population, 8.9% in the 70+ years group). On the other hand, the oppressive AM of the classification with the same input variables and settings except for 6 AMs (C6) has the largest coverage of days with mortality exceeding by at least 10% the baseline, i.e. the classification appears to be particularly useful for application into models that aim at predicting excess mortality (within HWWS; Sheridan and Kalkstein, 2004). The skill of the classifications in identifying weather conditions associated with mortality impacts is demonstrated by the fact that mean excess mortality is negative or close to zero in the other non-oppressive AMs.

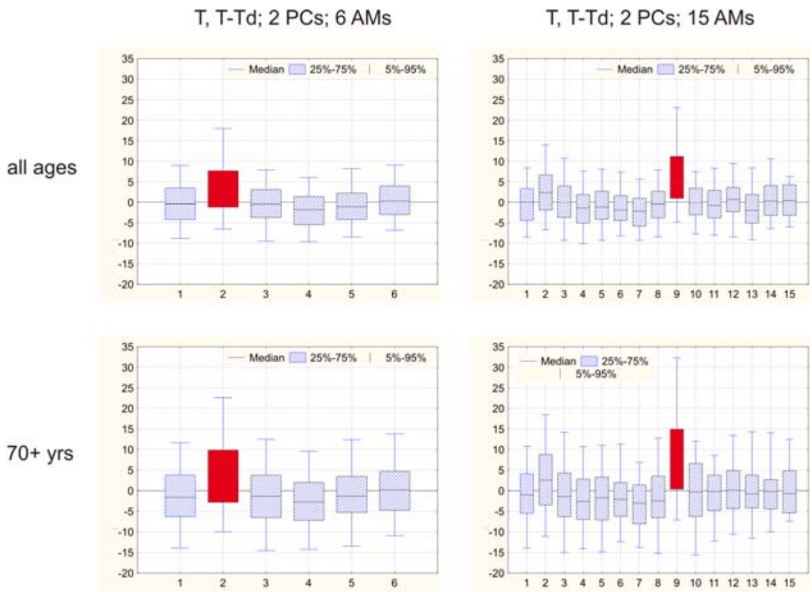

Figure 1. Box-plots of relative excess total mortality in individual AMs of the C6 (left) and C15 (right) classifications. Individual AMs are ranked according to their frequency of occurrence in each classification; the oppressive AMs are marked in red. Top: whole population, bottom: elderly population (70+ years).

Examples of the occurrence of the oppressive AMs of the C6 and C15 classifications are shown in Figure 2 for record-breaking hot summer of 1994 and another summer with less severe heat waves in 2004. The oppressive AM covers most days with pronounced excess mortality (note that daily excess mortality exceeds 200 deaths in the peak of the 1994 heat waves) in both seasons and both classifications; however, some days with relatively large excess mortality in 1994, after the peak of the heat wave, are not classified with the oppressive AM in C15. This is a consequence of the trade-off between the coverage of days with pronounced excess mortality (better in C6) and the separation from the rest of the sample/mean mortality increase on days classified with the AM (better/larger in C15).

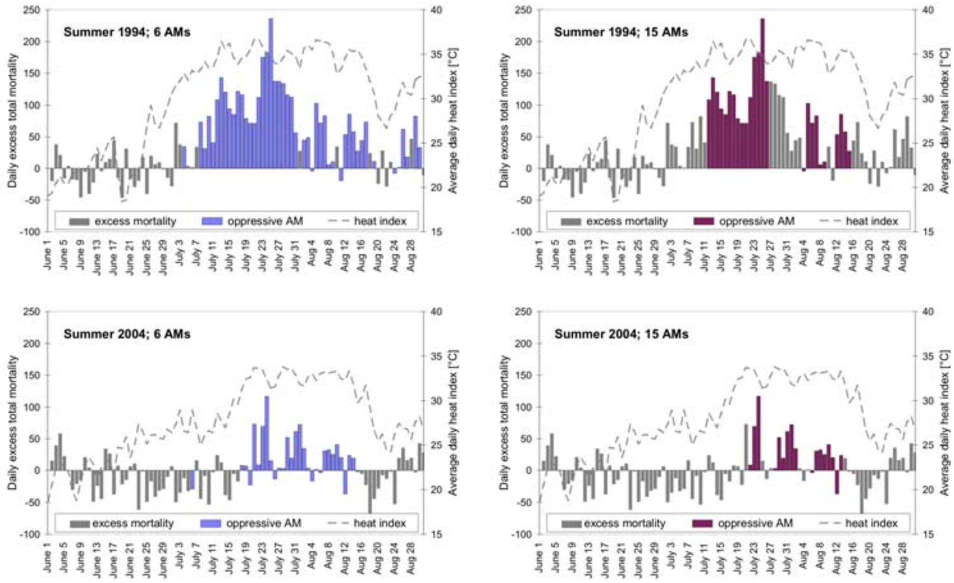

Figure 2. Occurrence of the oppressive AM (colour bars) during summers of 1994 and 2004 in South Korea, according to the C6 (left) and C15 (right) classifications. The total number of excess deaths due to heat was around 3350 (350) during the 1994 (2004) heat waves.

Scatter-plots of excess mortality against average daily heat index (Figure 3) show that the oppressive AM covers a large portion of days with pronounced heat-related mortality in C15, and nearly all in C6. The variance in mortality within the oppressive AMs is large and not all days classified with them are associated with excess mortality (Figure 3). This illustrates that the links are complex and mortality is affected not only by meteorological elements but also other factors like timing within a season, timing within a spell of oppressive days, longer-term changes in the public perception of heat, etc. A further analysis is needed to identify the relationships and find the conditions responsible for increased mortality risks.

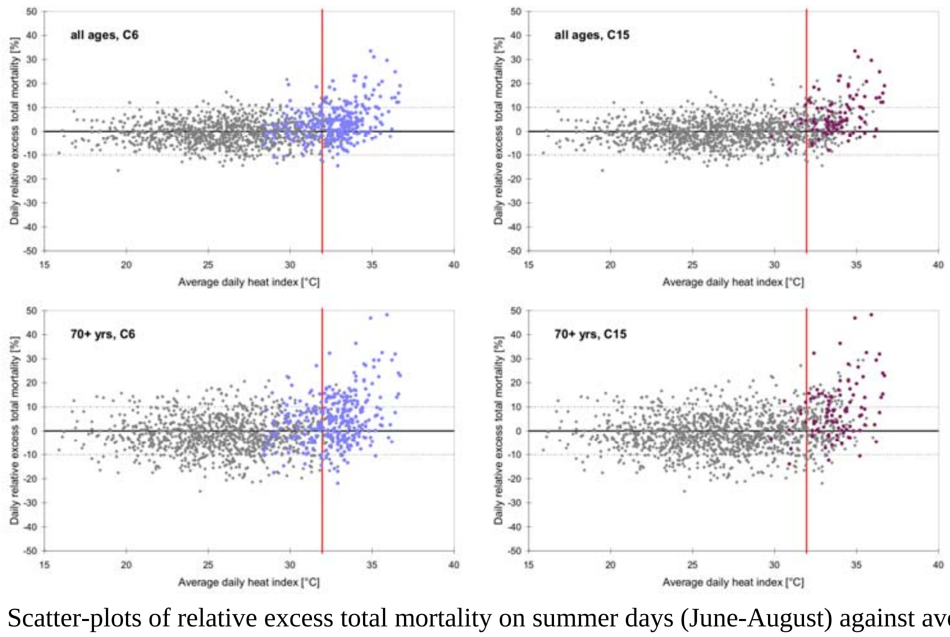

Figure 3. Scatter-plots of relative excess total mortality on summer days (June-August) against average daily heat index. The oppressive AMs of the C6 (left) and C15 (right) classifications are marked in colour. The dashed horizontal lines show limits of -10% and +10% relative excess mortality; the vertical line denotes the 80% quantile of average daily heat index in summer in the mean series for South Korea.

3.2 Regression models for excess mortality within the oppressive AMs

In the first step, linear regression models for relative excess daily mortality (in the whole population and the elderly) within the oppressive AMs of the C6 and C15 classifications were developed using the step-wise screening and the whole available period of data 1991-2005 (see Table 1 for the list of independent variables).

In C6, excess mortality is positively associated (Table 2) with day-time temperature and day-to-day change in night-time dew-point temperature, but non-meteorological factors are also important: mortality impacts decrease with the number of days with the oppressive AM both in previous summer and since the beginning of summer in a given year. For the whole population, mortality effects are found to decrease over time, too.

In C15, the regression models are more complex, with different predictors selected for the whole population and the elderly. Two non-meteorological factors are important: excess mortality in the oppressive AM increases with the day in sequence, and decreases with the time of season. A larger percentage of explained variance in C15 than C6 (Table 2) is related to a much smaller sample size (98 compared to 343). The C15 regression models appear to show some signs of overfitting, i.e. they may be too complex for given amount of data.

Table **2**. Regression models for relative excess mortality in the whole population (MOR_TOT) and the elderly (MOR_70+) within the oppressive AMs of the two selected classifications. Variables are explained in Table 1.

Classification	Dependent variable	Regression model	Percentage of explained variance [%]
C6	MOR_TOT	370.9 + 2.308 T15 + 0.995 dTd3 – 0.126 OAM – 0.090 OAMy – 0.218 YEAR	33
	MOR_70+	–89.3 + 3.177 T15 + 2.020 dTd3 – 0.153 OAM – 0.122 OAMy	27
C15	MOR_TOT	–24.1 + 2.930 dT9 + 2.142 Td15_1 – 1.636 TCAAVG_1 + 1.226 DIS – 0.243 TOS	60
	MOR_70+	–56.5 + 6.396 dT9 – 3.448 dTd15 – 2.853 TCAAVG_1 + 3.604 T3_1 + 1.107 DIS – 0.331 TOS	55

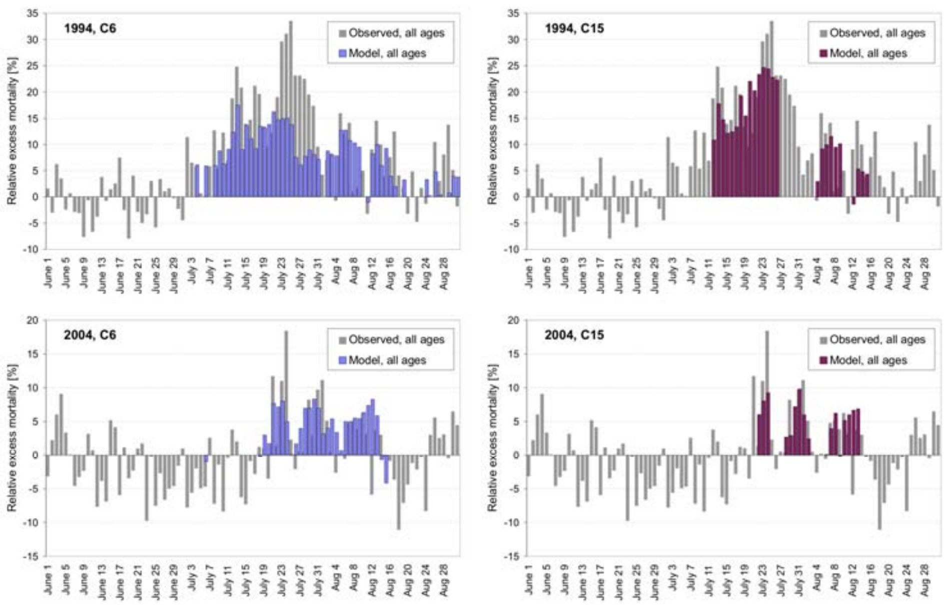

Figure 4. Observed excess mortality during summers of 1994 and 2004 in South Korea and modelled excess mortality on days with the oppressive AM of the C6 (left) and C15 (right) classifications.

Performance of the two models is compared in Figure 4 for the summers of 1994 and 2004. As expected, the model for the oppressive AM of the C15 classification produces a better fit but many days with excess mortality are not classified with the oppressive AM. The model for the C6 classification performs reasonably well except for the 8-day mortality peak in 1994, magnitude of which is largely underestimated.

If the set of 20 (50) days in June-August or July-August 1991-2005 with the largest recorded excess mortality is considered, the oppressive AM of the C6 classification captures most of them, and they are mostly associated with positive modelled excess mortality. Out of the 20 (50) days with the largest excess mortality in June-August, 19

(39) are associated with the oppressive AM of the C6 classification, and modelled relative excess mortality exceeds 5% on 16 (31) of them. We note that large excess mortality may also be related to other factors than oppressive weather conditions, so a 'perfect fit' is not expected.

With the view of their applications in HWWS, the models may be evaluated in terms of a skill score based on the number of 'successes' (excess mortality observed and modelled, Y/Y), 'false alerts' (excess mortality modelled but not observed, N/Y) and 'missing hits' (excess mortality observed but not modelled, Y/N). We use the threat score (TS; Wilks, 1995), computed as the number of correct forecasts of large excess mortality divided by the total number of cases when large excess mortality is observed and/or modelled.

The models' performance over July-August 1991-2005 is evaluated in Table 3 using different thresholds of excess mortality at which alerts would be issued, 3%, 5% and 10% relative to the expected number of deaths (the latter corresponds to more than 60 excess deaths in the Korean population). Since the models underestimate the number of days with large excess mortality, the performance is better when the bias is partly compensated for (the last two rows in Table 3). In C6, the days with modelled excess mortality exceeding by at least 5% the expected number of deaths cover 55% of days with observed excess mortality more than 10% above the baseline; however, TS is not large (0.30), particularly as the number of 'false alerts' exceeds that of 'successes'. An important finding is that the skill of the model for C15 is smaller than for C6 (Table 3); 'false alerts' are reduced at the expense of enhanced number of 'missing hits'.

Table **3**. Agreements (Y/Y, N/N) and disagreements (Y/N, N/Y) between observed/modelled days with large relative excess mortality (>=X% above the baseline) in July-August 1991-2005, according to the two selected classifications. TS stands for the threat score (Wilks, 1995).

a. Classification C6

OBS/MOD	Y/Y [%]	Y/N [%]	N/Y [%]	N/N [%]	TS	Bias
>=3%	10.3	20.5	6.2	62.9	0.28	0.54
>=5%	6.1	14.0	3.8	76.1	0.26	0.49
>=10%	1.6	5.7	0.2	92.5	0.21	0.25
>=5/3%	8.2	11.9	8.4	71.5	0.29	0.82
>=10/5%	4.0	3.3	5.9	86.8	0.30	1.35

b. Classification C15

OBS/MOD	Y/Y [%]	Y/N [%]	N/Y [%]	N/N [%]	TS	Bias
>=3%	5.2	25.7	1.8	67.3	0.16	0.23
>=5%	3.5	16.6	1.8	78.1	0.16	0.27
>=10%	1.9	5.4	0.4	92.3	0.25	0.32
>=5/3%	4.3	15.8	2.7	77.2	0.19	0.35
>=10/5%	2.7	4.6	2.7	90.0	0.27	0.74

3.3 Regression models for excess mortality – performance on an independent sample

In the second step, the models were developed using the same procedure but only 12 years of data (1991-2002), and their performance was tested on an independent 3-year sample (2003-2005).

There were altogether 30 days in July-August 2003-2005 with relative excess mortality exceeding by at least 5% the expected number of deaths; the oppressive AM of the C6 (C15) classification covers 17 (10) of them, and on all of them in both classifications, positive excess mortality was predicted. In C6, 12 out of 15 days with the largest excess mortality are classified with the oppressive AM, which suggests that the classification is a useful tool for finding stressful weather conditions.

The reproduction of day-to-day variations in mortality (when the oppressive AM is present) is not very successful in either classification (Figure 5), but C6 outperforms C15 again as the oppressive AM of the C6 classification covers more days with large excess mortality. Even in terms of the explained variance of the day-to-day changes of excess mortality within the oppressive AM, the C6 classification is superior.

We note that for the application in a HWWS, the prediction of *when* excess mortality may be expected is more important than the prediction of the magnitude of the excess mortality itself. The threat score of the prediction of days with large excess mortality is also larger in C6 than C15 (for any threshold of what is considered 'large' excess mortality), and in contrast to results for the dependent samples discussed above, it is higher for mortality in the elderly than the whole population. For the threshold of relative excess mortality at least 3% above the baseline, the predicted days cover almost half of the observed days with large excess mortality, and the rate of 'successes' against 'false alerts' is around 3, leading to a relatively high TS=0.40.

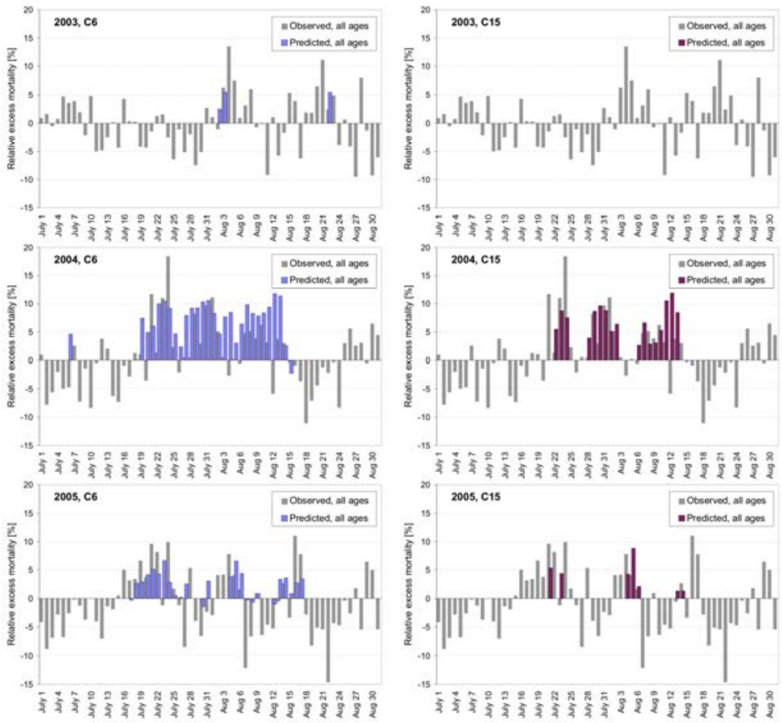

Figure 5. Observed excess mortality during July-August in 2003, 2004 and 2005 in South Korea and predicted excess mortality on days with the oppressive AM of the C6 (left) and C15 (right) classifications.

4. CONCLUSIONS

The study shows that results of the AM classifications, mainly with respect to the characteristics of the oppressive AM and its links to excess mortality, depend on specific settings of the analysis to a larger degree than previous studies suggest. This implies that more attention should be directed towards justifying and reasoning of the settings of the statistical procedures in further applications. The relationships of excess mortality within the oppressive AM to meteorological and non-meteorological factors, presented for the two selected classifications, may be found useful also for the development of the heat-watch-warning system that is currently tested for Seoul and other large Korean cities (Kalkstein et al., 2008). The classification with a relatively small number of AMs (6) shows much better predictive skills in spite of the smaller mean excess mortality in the oppressive AM.

ACKNOWLEDGEMENTS

The study was supported by the Korea Research Foundation and the Czech Science Foundation under projects KRF-2006-C00005 and 205/07/J044. Thanks are due to the staff of the Korea National Statistical Office, Seoul, for providing daily mortality datasets.

REFERENCES

Davis, R.E., P.C. Knappenberger, W.M. Novicoff, and P.J. Michaels, (2003), Decadal changes in summer mortality in U.S. cities. *International Journal of Biometeorology*, 47, 166-175.

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

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

Kalkstein, L.S., S.C. Sheridan, and Y.C. Au, (2008), A new generation of heat health warning systems for Seoul and other major Korean cities. *Meteorological Technology & Policy*, 1, 86-92.

Koppe, C., S. Kovats, G. Jendritzky, and B. Menne, (2004), Heat waves – risks and responses. World Health Organization, Regional Office for Europe, Copenhagen.

Kysely, J., and J. Kim, (2009), Mortality during heat waves in South Korea, 1991 to 2005: How exceptional was the 1994 heat wave? *Climate Research*, 38, 105-116.

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

Wilks, D.S., (1995), Statistical Methods in the Atmospheric Sciences: An Introduction. Academic Press, 467 pp.