(Statistical methods for connecting)
Health, heat stress, and vulnerability in urban populations

Stephan R. Sain
Geophysical Statistics Project
Institute for Mathematics Applied to Geosciences
National Center for Atmospheric Research
Boulder, CO USA

Matt Heaton, Tammy Greasby, Olga Wilhelmi, Mary Hayden, Andy Monaghan, Keith Oleson, Michael Barlage (NCAR); Johannes Feddema (KU); Chris Uejio (FSU); Deborah Bannerjee (Houston Department of Health and Human Services); more...


Supported by NSF and NASA.
• Disclaimer: This talk is about statistical methods and analysis.

• Motivation: Vulnerability — exposure, sensitivity, and adaptive capacity.

• Challenge: Health impacts are distributed unevenly.
   ★ Differentiated vulnerability.

• Goal: Incorporate high-resolution numerical model output, satellite data, parcel data, census data, survey data into a statistical model for public-health endpoints (heat-related mortality, hospitalizations, 911 calls, and heat-related attitudes).

• Postscript: Dealing with uncertainty...
The foundation of the statistical approach comes from two ideas:

Each block group has its own heat-related risk curve (differentiated vulnerability).

The spatial variation in these curves can be explained via covariates (e.g., census, parcel, etc.) and spatial random effects.
A hierarchical model
for connecting mortality to heat/demographics/etc.

Stage 1:  \( Y_{tb} \sim \mathcal{P}(E_b \exp\{\mu_{tb}\}) \)

\textit{Links mortality to risk}

Stage 2:  \( \mu_{tb} = \beta_0 + (\text{demo}_b)\beta + (\text{heat}_{tb})\eta_b + \phi_b + \delta_t \)

\textit{Links risk to demographics & heat stress}

Stage 3:  \( \eta_b = \gamma_0 + (\text{demo}_b^*)\gamma + \nu_b + \zeta_b \)

\textit{Links heat response to demographics}

- Breaks a complicated problem into smaller, more manageable pieces
  - Buy a house with a paperclip? (oneredpaperclip.blogspot.com)
A hierarchical model

After some careful consideration (and some math), the hierarchical model boils down to a Bayesian Poisson GLM.

- Covariates enter the model in a way to minimize collinearity.
- Spatial random effects enter through a collection of basis functions that depend on the spatial structure but minimize collinearity with covariates.
- Model includes main effects (demographics, heat, space) and interactions.

With this in mind, we are moving forward to generalize this statistical approach for data exploration, different spatial lattices, and additional locations.
A hierarchical model

Stage 1: \( Y_{tb} \sim \mathcal{P} \left( E_b \exp \{ \mu_{tb} \} \right) \)

- Mortality data from the Texas Department of State Health Services and listing ICD9 codes for cardiovascular, genitourinary, respiratory, nervous system, or hyperthermia as a contributing cause of death.
- Aggregate counts \( \{Y_{tb}\} \) to census block groups; focus only on (M)JJA(S).
- Expected counts \( \{E_b\} \) computed assuming constant mortality rate.
- \( \{\exp \{ \mu_{tb} \}\} \) are relative risk of heat-related mortality over space and time.
A hierarchical model

Stage 2: \[ \mu_{tb} = \beta_0 + (\text{demo}_b)'\beta + (\text{heat}_{tb})\eta_b + \phi_b + \delta_t \]

Demographic variables:

- Census: race, age, living alone, public transportation, poverty, education, etc.
- Parcel: air conditioning, age of residential buildings, pools, building quality, etc.

% in Poverty, 2000 Census.
A hierarchical model

Stage 2: \( \mu_{tb} = \beta_0 + (\text{demo}_b)' \beta + (\text{heat}_{tb}) \eta_b + \phi_b + \delta_t \)

- High Resolution Land Data Assimilation System (HRLDAS) coupled with an Urban Canopy Model (UCM) to better represent the physical processes involved in the exchange of heat, momentum, and water vapor in the urban environment.

- Heat stress measured through various heat indices: discomfort index, NWS heat index, humidex, apparent temperature, wet bulb globe temperature, daily max/min temp.
Variable selection for heat stress:

<table>
<thead>
<tr>
<th>Rank</th>
<th>Likelihood</th>
<th># Positive CIs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TMP:MIN</td>
<td>TMP:MIN</td>
<td>TMP:MIN</td>
</tr>
<tr>
<td>2</td>
<td>HI:MIN</td>
<td>HU:MAX</td>
<td>DI:MIN</td>
</tr>
<tr>
<td>3</td>
<td>DI:MIN</td>
<td>DI:MIN</td>
<td>WBGT:MIN</td>
</tr>
<tr>
<td>4</td>
<td>WBGT:MIN</td>
<td>AT:MAX</td>
<td>DI:MAX</td>
</tr>
</tbody>
</table>

Variable selection for Stage 2: $\mu_{tb} = \beta_0 + (demo_b)'\beta + (heat_{tb})\eta_b + \phi_b + \delta_t$

<table>
<thead>
<tr>
<th>Variable</th>
<th>Post. Mean</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>% older than 65 years</td>
<td>5.66</td>
<td>(5.49, 5.80)</td>
</tr>
<tr>
<td>% with no air conditioning</td>
<td>0.48</td>
<td>(0.37, 0.61)</td>
</tr>
<tr>
<td>% living alone</td>
<td>0.95</td>
<td>(0.80, 1.21)</td>
</tr>
<tr>
<td>% African American</td>
<td>0.55</td>
<td>(0.35, 0.81)</td>
</tr>
<tr>
<td>% in poverty</td>
<td>0.49</td>
<td>(0.36, 0.65)</td>
</tr>
<tr>
<td>% under 5 years</td>
<td>-1.93</td>
<td>(-2.06, -1.76)</td>
</tr>
</tbody>
</table>

Variable selection for Stage 3: $\eta_b = \gamma_0 + (demo^*_b)'\gamma + \nu_b$

<table>
<thead>
<tr>
<th>Variable</th>
<th>Post. Mean</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>% older than 65 years</td>
<td>1.80</td>
<td>(1.39, 2.22)</td>
</tr>
<tr>
<td>% Caucasian, older than 65, Living Alone</td>
<td>0.60</td>
<td>(0.38, 0.83)</td>
</tr>
</tbody>
</table>
Figure 1: Posterior mean of cumulative relative risks averaged over time \((8 \times 92)^{-1} \sum_{y,t} \exp\{\mu_{ytb}\}\).

Block groups in white are those where a 95% credible interval contained zero (i.e., no significant increase or decrease in risk from \(E_{by}\)). Block groups with relative risk greater than zero are at an elevated risk of non-accidental mortality.

5.2. Heat Exposure Risk Factor Results

Next, consider the portion of the cumulative relative risk that is due to heat exposure; specifically, consider inference for the term \((\text{heat}) \cdot \mu_{ytb}\) in Equation (2). Table 3 displays the heat exposure variables from Table 1 ranked according to deviance information criterion (Spiegelhalter et al., 2002), number of 95% credible intervals strictly greater than zero, and total rank.

Consider, first, ranking the variables based solely on the deviance information criterion (DIC). The DIC is a measure of model goodness of fit. Models with low DIC are preferred to models with high DIC. According to Table 3, the heat variable that gives the best model fit is daily minimum temperature (TMP:MIN). However, the two next best models used minimum heat index (HI:MIN) and discomfort index (DI:MIN) and had DIC values only 2.74 and 4.23 greater than the model using TMP:MIN suggesting nearly equal fits for these three variables. Apart from daily minimum temperature, these other heat variables are composite temperature/humidity indices suggesting that non-accidental mortality may be explained by a combination of variables related to high heat.

Identifying a heat variable which is positively associated with non-accidental mortality risk can be beneficial for heat hazard preparedness and response purposes. For example, the National Weather Service (NWS) issues heat warnings if the heat index exceeds certain thresholds. This method for issuing heat warnings suggests that by identifying a heat exposure variable that is positively associated with relative risk, public health officials will be able to take appropriate actions should the heat variable reach dangerous levels. Considering this aspect, the second column of Table 3 ranks the heat exposure variables based on the...
Figure 2: Map of posterior mean of time-averaged heat risk \((8 \times 92)^{-1} \sum_{y,t} (heat_{ytb}) \eta_b\) using daily minimum temperature (TMP:MIN). Block groups shaded white are those where a 95% credible interval contained zero (i.e. no significant increase or decrease in risk due to heat exposure). While Figure 1 is useful in identifying areas with high overall relative risk, Figure 2 is perhaps more useful in assessing public heat vulnerability because it portrays those block groups where relative risk is inflated due specifically to heat exposure. Figure 2 highlights several inner-city and suburban block groups with an elevated risk due to heat exposure. For example, the group of census blocks to the north-east of the city center have elevated relative risk due to heat exposure yet the overall relative-risk of these same block groups in Figure 1 is less than zero.

5.3. Demographic Risk Factor Results

Using TMP:MIN as the heat exposure variable, Table 4 displays the selected demographic variables from the forward selection algorithm described in Section 4. The chosen model had a BIC value of -2219421 compared to an null model BIC value of -2211342 (an difference of -8079). The next smallest BIC value was -2219446. This small difference in BIC suggests that several variables may equally explain trends in relative risk but we choose to focus on the model that achieved the best fit.

Table 4 also displays the posterior mean of the percent change in relative non-accidental mortality risk given a 1% increase in the demographic variable holding all else constant (the coefficients in Equation (2)). For example, according to Table 4, if the percent of the population within a block group older than age 65 increased by 1%, the relative risk of non-accidental mortality for that block group would increase by 5.66% (95% credible interval of 5.49% to 5.80%).

The variables identified in Table 4 are enlightening and highlight important potential vulnerability factors. The percent of the population 65 years and older and percent of...
Posterior mean of $\eta_b$ (sensitivity), given $\eta_b$ is credibly greater than zero.
Heat maps (exposure).

avg min temp = 22

avg min temp = 23.3

avg min temp = 23.8

avg min temp = 24.2

avg min temp = 24.4

avg min temp = 24.7

avg min temp = 24.9

avg min temp = 25.2

avg min temp = 25.6

avg min temp = 26.2
Total mortality w/ prediction intervals.
Total mortality and block groups with high (>2.5 deaths/100K people) average mortality rates (exposure + sensitivity).
Probability of at least one 911 call from an analysis of $n = 2054$ heat-related 911 calls from 2006-2010 via a marked point pattern.
Questions?

ssain@ucar.edu
http://www.image.ucar.edu/~ssain

Thank You!
