A Microsimulation Model of Population Heat Exposure

Samarth Swarup\textsuperscript{1}, Julia M. Gohlke\textsuperscript{2}, and James R. Bohland\textsuperscript{3}

\textsuperscript{1}Network Dynamics and Simulation Science Lab, Biocomplexity Institute, \\
\textsuperscript{2}Department of Population Health Sciences, Virginia-Maryland College of Veterinary Medicine, \\
\textsuperscript{3}Global Forum on Urban and Regional Resilience, Virginia Tech, Blacksburg, VA, USA \\
\{swarup,jgohlke,jayjon\}@vt.edu

Abstract. Exposure to extreme heat is an important cause of morbidity and mortality throughout the world. The elevation of temperatures and increases in extended periods of high temperatures due to climate change means that heat exposure as a health concern will increase. Thus, methodologies that researchers and practitioners can use to assess heat exposure among various population groups will become increasingly important. Human exposure to heat can be measured by wearable sensors and the maintenance of activity logs of a sample of individuals over the course of their daily activities. We introduce an alternative methodology that we believe offers great promise for evaluating heat exposure of the entire population of a city, region or state. Microsimulation modeling of the daily activity patterns of a synthetic population for the state of Alabama demonstrates the efficacy of this approach in measuring “potential” heat exposure, and in evaluating the significance of demographic attributes, in this case household income, that correlate with different exposure levels.

Keywords: synthetic population, heat exposure

1 Introduction

Climate change is increasing the number of days of extreme heat. Extreme heat events have become more frequent and intense and heat waves everywhere in the U.S. are projected to become more intense [20, 17]. Exposure to extreme heat is associated with earlier death [2], increased hospitalizations [13], and adverse birth outcomes including preterm birth [12].

However, it is difficult to obtain estimates of population-wide exposures to extreme heat because temperatures can vary widely spatio-temporally. This means that approaches that rely on physically instrumenting subjects to record their exposure as they move about during the day have trouble scaling up to the entire population. On the other hand, approaches that use census data in combination with large-scale temperature data are only estimating the exposure at
people’s home locations, not taking into account daily mobility patterns, nor other exogenous factors such work site conditions.

In this paper we explore the use of a synthetic population in an agent-based model to assess the heat exposure of different segments of the population in the State of Alabama, USA. Alabama is a large state with a range of topography, urban/rural spaces, coastal/inland differences, and high summer temperatures. A synthetic population of Alabama is a representation of the demographics, typical daily activities, and activity locations of the entire population of the region. It is generated by combining multiple sources of data in a series of steps, as described later in the paper. The approach represents a means of measuring what we have termed “potential extreme heat exposure” that has some important advantages over current methodologies because it enables researchers to estimate exposures for an entire population of a region stratified by socio-economic and other attributes. Potential extreme heat exposure occurs in our model when a person is located at a time and position during their daily activities where the temperature is 90°F or greater, though this threshold can easily be changed to represent different levels of extremity. The exposure is potential because we currently do not assess environmental mitigating factors, e.g., presence of air conditioning, whether the individual is within a structure or outside, in direct sun versus shade, or whether one is within a vehicle.

Assessment of actual rather than potential exposure is required to assess the effects of high temperatures on specific morbidity or mortality impacts. However, our purpose in this paper is to demonstrate the potential agent-based modeling of a synthetic population provides for assessing exposure. The present work is the essential building block of a longer research program using synthetic populations to assess hazards exposure.

In this article we provide a context for understanding the importance of assessing heat exposure as a health risk, discuss the relative advantages gained through the use of our model from both research and policy-planning perspectives, describe in detail the structure of the model, present empirical evidence of its applicability for measuring potential heat exposure for the state of Alabama, and end with a discussion of how the approach can be enhanced in the future to provide even greater utility.

2 The health/extreme heat relationship

Individuals have multiple encounters with many different health hazards in the course of their life span. Clearly not every encounter leads to a change in morbidity because that relationship is influenced in part by the intensity of the exposure, or dosage, of the hazardous condition. Our focus is on a particular hazard, high temperatures, which are foundational to the occurrence of extreme heat events (EHEs), which result in more deaths annually in the United States, 7,800 from 1999-2009, than other extreme weather events [18]. Furthermore, Keim forecasts that of the six disasters most likely to be influenced by climate change, extreme heat is the one with the greatest probability of increasing mortality, severe injuries, and widespread chronic illnesses [11]. It has both direct and secondary pathways for impacting health.
No single criterion exists for defining an extreme heat event. EHEs typically are measured by a combination of high temperature and/or humidity and the duration of those temperatures for an extended period of days [2].

Exposure to high temperatures is a function of both environmental and socio-demographic contextual factors. Environmental context becomes important because micro-thermal regimes vary spatially, particularly in urban areas where heat islands arise because of differentials in surface materials. Thus, every environment has a particular geographic distribution of high temperature sites. Whether individuals encounter those sites is a function of how they traverse the geography of their place during their daily activities.

For example, in Figure 1a, an individual begins the day at point A for approximately 7 hours at which point (B) s/he moves to C for another activity (such as work), and remains there for much of the day before moving on to point E at the end of the day. By tracking an individual’s activity across a region of varying temperature levels, we can estimate the potential exposure of an individual based on thermal conditions at specific locations A, C, and E, for a defined time of day. Using a synthetic population and agent-based modeling the amount of time an individual is at a place or during their movement where temperatures exceed 90°F can be determined for any individual. The total number of hours a day an individual could be exposed to temperatures exceeding 90°F gives a daily accounting of potential exposure levels for every individual in the population. Daily profiles of thermal exposure have been measured in small subsamples of populations through the use of thermal sensors attached to individuals [4, 14]. Karner et al. used a simulation approach in assessing heat exposure for a specific population, cyclists using specific routes [10]. With a synthetic population and agent based modeling, exposure potentials can be obtained for all individuals in the population across all locations in a place. The current model does not account for time spent indoors vs. outdoors, though this can be incorporated if the appropriate data become available, which would allow for studying occupational and other activity-related effects.
3 Methodology

Our approach is to use a detailed, individual-level agent-based model known as a synthetic population, which was previously generated [1] as follows (also see Figure 1b):

- **Generating agents with demographics**: We begin by using data from the American Community Survey (ACS), which provides demographic distributions at the US Block Group level and a 5% sample of records at the Public Use Micro Area (PUMA) level. These are combined using the iterative proportional fitting technique to create a disaggregated population for the region [3]. Since the ACS provides household-level data, the generated agent population is also grouped into households, in addition to having characteristics such as age, gender, household income, and more.

- **Assigning activities**: In the next step, we assign typical daily activities to each agent by integrating data from the National Household Travel Survey (NHTS). Activity schedules are assigned by learning a Classification and Regression Tree (CART) model [5] from the NHTS data, which uses demographic variables to split the NHTS records into clusters. The dependent variable chosen is the total time spent outside the home. This results in a typical daily schedule for each agent, including start times and durations for each activity.

- **Assigning activity locations**: Each activity for each agent is assigned an appropriate location. Home locations are constructed using data on residence types in the area from the ACS and road network data from HERE (formerly NAVTEQ). School locations are obtained from the National Center for Education Statistics (NCES). Business locations are obtained from Dun & Bradstreet. Other points of interest, such as parks, tourist attractions, etc., are obtained from HERE. Activity locations are assigned to activities using a gravity model [7].

The resulting model is a dynamic representation of human mobility and interaction over the course of a normative day. The synthetic population is specific to a geographic region because of its dependence on “contingent realities” for the area—the demographics of the people who live there and the distribution of actual activity locations. It provides a plausible, bottom-up mechanism for generating large-scale structured representations of populations without making assumptions as are common in more stylized models. The resulting data set is very rich and can be applied to many problem domains [16, 15].

For this study, we use a previously generated synthetic population for Alabama, with 4.37 million residents [1], based on data from the year 2009. Methods to calculate any metric of heat exposure can be employed, such as the total duration for which the agent is exposed to temperatures above a certain threshold, or the average temperature over the day to which the agent is exposed either at a specific site or during travel to that site. For this study we determined the number of hours per day an individual was exposed to temperatures in excess of 90°F during the summer of 2009 and then calculated the average number of
hours per day over 90°F each individual could have experienced an extreme heat exposure. We did the same calculations for a threshold of 95°F also, but we only report on the analysis for 90°F as the results for 95°F demonstrate very similar distributional patterns across geographic areas and income groups.

3.1 Temperature calculations

For temperature data, we use the North American Land Data Assimilation System (NLDAS) data set [21]. NLDAS provides an estimate of the air temperature at 2 meters above the surface at a spatial resolution of 0.125° and an hourly temporal resolution for the continental USA. While the spatial resolution is not very high, the hourly temporal resolution is needed for computing heat exposure. Other data sources, such as MODIS [9], provide high spatial resolution, but very low temporal resolution. One possibility for extending the present work would be to combine the two data sources to obtain high spatial and temporal resolution. The NLDAS data for 2009 (June 1 – Sep 30) were downloaded from NASA’s GES DISC system\(^1\). The calculation of average hours/day above 90°F for each individual in the synthetic population is done as follows.

The daily activity schedule for each individual is a list of (activity type, start time, duration, latitude, longitude) tuples. For example, if a person is at home at the start of the day and leaves for work at 8 a.m., their first activity would be written as (HOME, 00:00:00, 08:00:00, <lat, lon of their home location>). If they arrive at work at 9 am and stay there till 5 p.m., their second activity would be written as (WORK, 09:00:00, 08:00:00, <lat, lon of their work location>). Note that the third item is 08:00:00 to indicate that the duration of the work activity is 8 hours. Also note that travel activities are not explicitly noted in the activity schedule, but we account for potential exposure during travel as will be explained further below. The travel mode information in NHTS could be used to refine this step in the future.

To calculate the time above 90°F for any activity, we do interpolation as follows. Let \(t_{\text{start}}\) be the start time of a given activity and \(t_{\text{end}}\) its end time \((t_{\text{end}} = t_{\text{start}} + \text{duration})\). The NLDAS data give us the temperature at the hour marks before and after \(t_{\text{start}}\) and \(t_{\text{end}}\). Let \(t_{\text{start}}^+\) denote the hour mark after \(t_{\text{start}}\) and \(t_{\text{start}}^-\) denote the hour mark before \(t_{\text{start}}\). For example, if \(t_{\text{start}} = 9:15\) a.m., \(t_{\text{start}}^- = 9:00\) a.m. and \(t_{\text{start}}^+ = 10:00\) a.m. From the NLDAS data, we can look up the temperature \(T_{\text{start}}^-\) corresponding to \(t_{\text{start}}^-\) and \(T_{\text{start}}^+\) corresponding to \(t_{\text{start}}^+\). We then interpolate to find the temperature at \(t_{\text{start}}\), as

\[
T_{\text{start}} = T_{\text{start}}^- + \frac{(t_{\text{start}} - t_{\text{start}}^-)}{(t_{\text{start}}^+ - t_{\text{start}}^-)} \times (T_{\text{start}}^+ - T_{\text{start}}^-). \tag{1}
\]

Since \((t_{\text{start}}^+ - t_{\text{start}}^-)\) is always 1 hour, this simplifies to,

\[
T_{\text{start}} = T_{\text{start}}^- + (t_{\text{start}} - t_{\text{start}}^-) \times (T_{\text{start}}^+ - T_{\text{start}}^-). \tag{2}
\]

\(^{1}\) http://disc.sci.gsfc.nasa.gov/uui/datasets?keywords=NLDAS
where \((t_{\text{end}} - t_{\text{start}})\) is expressed as a fraction of an hour. We calculate the temperature at the end of the activity period, \(T_{\text{end}}\), in a similar manner. Now we have the temperature at \(t_{\text{start}}, t_{\text{end}}\), and every hour mark in-between (from the NLDAS data). If two consecutive temperature values in this list are above the threshold, we assume that the temperature was above the threshold for the entire duration. If only one out of two consecutive values is above the threshold, then we calculate the duration above the threshold by linear interpolation in a similar way to the temperature calculation above. Thus for every activity during the day, for every person, we can calculate the duration above the temperature threshold. We do not do spatial interpolation between the centroids of the NLDAS cells because the temperature values in the cells do not represent measurements at the center, but rather an average value for the cell.

For travel activities, we do exactly the same thing, with the only difference being that \(T_{\text{start}}\) is taken to be the temperature at the origin location and \(T_{\text{end}}\) the temperature at the destination location.

We repeat this calculation for every synthetic individual, for every day between June 1, 2009 and Sep 30, 2009 and calculate the average hours/day the ambient temperature is over 90°F.

![Graphs](image)

(a) Average hours/day with ambient temperature > 90°F. (b) Average hours/day of ambient temperature >90°F by household income.

**Fig. 2**

Our initial simulation for the entire synthetic population of Alabama provides a baseline for additional analyses. The distributional plot for the initial analysis (Figure 2a) illustrates several important outcomes concerning potential heat exposure in the state. First, the majority of residents experienced between three and six hours per day of extreme heat. While the overall number declined after six hours, a sizable number or residents were at locations where on average seven or more hours per day of high heat exposure could occur. Finally, the overall distribution was biased towards the higher hourly amounts. While over 10,000 individuals had no direct exposure to extreme heat, smaller numbers had one hour or less, particularly in comparison to those who averaged more than six hours.

Our model was run for the entire state with the results analyzed by the household income (Figure 2b), since income would impact access to mitigating factors that could reduce exposure (e.g. air conditioning) or reduce health effects of exposure (health care). The results indicate an income disparity between...
the higher and lower exposure levels in the state. Those experiencing the lower number of hours/day of extreme heat exposure (left of the first dashed line in Fig. 2b) had an average annual household income of $63,268 while those at the higher end of the exposure level (right of the second dashed line in Fig. 2b) had an average annual household income of $42,409; a differential in excess of $20,000. Some variations from that trend were evident (Figure 2b) but an income bias in the plot of household income to potential exposure clearly existed. A more sophisticated model could be used to identify the trend breaks precisely, but that would not change the basic conclusion about income disparities in heat exposure.

Geographic and climate differences might account for some of this disparity. In addition, the disparity can be due to multiple other reasons. It has been shown that tree cover tends to be higher in higher-income neighborhoods [8]. Increasing altitude is also generally associated with increasing socioeconomic status [19]. Both of these factors can contribute to lower temperatures. From another perspective, Deryugina and Hsiang have shown that daily average temperatures are negatively associated with economic performance. They estimate that “the productivity of an individual declines roughly 1.7% for each 1°C (1.8°F) increase in daily average temperature above 15°C (59°F)” [6].

4 Discussion

The goal of this research was to test a new methodology for assessing extreme heat exposure. The use of agent-based modeling of a synthetic population provides researchers and practitioners a new approach to analyzing what will increasingly be an important health threat with expected changes in thermal regimes due to climate change. The results presented here represent a strong proof of the concept of using the approach to assess potential extreme heat exposure. More importantly, enhancing the methodology would make a transition from potential to actual possible. We believe our approach offers five important advantages over existing methods.

First, since the entire population of a region is used, albeit a synthetic one, sampling error is not an issue. The limitations and errors in the estimates arise from the availability and quality of the data from which synthetic populations are constructed, which can be refined as higher quality data sources become available.

Second, since the synthetic population reflects the entire population and includes the composition of a region’s residents, exposures can be calculated for different strata in the population, making multiple cross-classifications feasible such that socio-economic (and other) comparisons in heat exposure are calculable. The ability to identify multivariate disparities in exposures makes heat exposure calculations more specific, which could ultimately make health risk calculations more specific. This methodology could offer public health officials an important tool for developing responses and mitigation practices that are better suited to specific groups.

Third, good estimates can be achieved for potential and actual heat exposure for groups that heretofore have been difficult to obtain. Many of these are the
most vulnerable to mortality and morbidity outcomes from extreme heat, such as low income groups, the elderly, or socially isolated individuals.

Fourth, exposure to extreme cold as well as extreme heat can be estimated readily. Because both extreme heat and cold create health risks, a platform that readily permits either offers major advantages to assessment of health risks.

Fifth, the use of an agent-based model using a synthetic population can be used as a planning tool to assess the consequence of different policy approaches to mitigating the health impacts of extreme heat exposures for different scales of places: neighborhood, city, region or state. For example, the current strategy for reducing exposure levels and mitigating health effects is to move higher risk groups to cooling shelters. Typically these are public facilities, such as libraries, administrative offices, or arenas. In some instances these facilities are not well sited for the more vulnerable populations. Agent-based models would enable policy makers to simulate exposure levels of different at risk groups under different weather conditions to measure where the potential impacts to groups are greater. Those “what if” scenarios can lead to more informed decisions on where best to locate cooling centers, how to deploy public transportation to maximize the use of these facilities, or whether to announce high risk days for different areas of a region or different segments of the population.

5 Conclusion

Advancing the utility of the methodology will require enhancing the platform to include additional social and health attributes of the synthetic population. Surveys of time spent in non-climate-controlled environments can be integrated to address potential vs. actual exposure. Existing health status can be assigned to individuals within the population to capture an important dimension of vulnerability. Status measures can, in turn, be linked to comorbidity probabilities enabling one to track mortalities or morbidities occurring as a consequence of different exposure conditions.

Additionally, regional environmental and infrastructure elements can be added to the platform. The travel infrastructure will enable investigators to incorporate exposures associated with specific infrastructures and routes that are used in travelling to work or recreational movement. The spatial resolution of the model could be increased by integrating high spatial resolution temperature data like MODIS [9] with the high temporal resolution temperature data (NLDAS) used here.

The model could produce multi-day exposure profiles from combinations of hazards including heat, pollution, chemicals or other environmental risk factors. Multiple risk conditions are possible now but only for static or fixed locations. The use of synthetic populations in modeling expands our ability to analyze long term (3-5 years) and short-term exposures under multi-hazard conditions. The ability to achieve that level of inquiry will provide public health and other officials a tool for coping with both episodic as well as incremental environmental threats. Such an approach will have great value as the variability in weather conditions increases because of climate change.
Acknowledgments.

S.S. was supported in part by DTRA CNIMS Contract HDTRA1-11-D-0016-0001 and NIH Grant 1R01GM109718. J.G. was supported in part by a grant from NIH/NIEHS - R01ES0230

References


