

LETTER • OPEN ACCESS

Online signals of heat vulnerability: revealing city-level response to extreme heat using Google Trends

To cite this article: Jangho Lee 2025 *Environ. Res. Lett.* **20** 124034

View the [article online](#) for updates and enhancements.

You may also like

- [Avoided population exposure to extreme heat under two scenarios of global carbon neutrality by 2050 and 2060](#)
Yadong Lei, Zhili Wang, Xiaoye Zhang et al.
- [Extreme heat preparedness and response implementation: a qualitative study of barriers, facilitators, and needs among local health jurisdictions in the United States](#)
Jessica C Kelley, Cat Hartwell, Evan C Mix et al.
- [Inclusion of child-relevant data in the development and validation of heat vulnerability indices: a commentary](#)
Kate R Weinberger, Blean Girma, Jane E Clougherty et al.



The Electrochemical Society
Advancing solid state & electrochemical science & technology



**249th
ECS Meeting**
May 24-28, 2026
Seattle, WA, US
*Washington State
Convention Center*

Spotlight Your Science

***Submission deadline:
December 5, 2025***

SUBMIT YOUR ABSTRACT

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

OPEN ACCESS

RECEIVED
18 September 2025REVISED
3 November 2025ACCEPTED FOR PUBLICATION
18 November 2025PUBLISHED
28 November 2025

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.

Online signals of heat vulnerability: revealing city-level response
to extreme heat using Google Trends

Jangho Lee

Department of Earth and Environmental Sciences, University of Illinois Chicago, Chicago, IL 60607, United States of America

E-mail: jholee@uic.edu**Keywords:** Google Trends, crowdsourced data, heat adaptation, urban heat, heat resilienceSupplementary material for this article is available [online](#)

Abstract

Extreme heat poses a growing public risk, yet understanding real-time, population-level human response remains a major challenge. This study leverages Google Trends search activity (2016–2024) across 30 major US metropolitan areas to quantify digital responses to heat. Using generalized additive models, two key response metrics are identified: a temperature ‘threshold’ that triggers public concern and a ‘slope difference’ that measures the intensity of the subsequent reaction. The analysis first establishes that a city’s median climate primarily determines its baseline threshold, while the variability of climate governs the slope difference, both reflecting long-term adaptation to local conditions. After accounting for these climatic effects, the study finds that socioeconomic factors explain a significant portion of the remaining variance in digital heat responsiveness. Specifically, the climate-adjusted threshold is higher in cities with greater social vulnerability, indicating a dangerous delay in awareness. This interplay creates five distinct urban typologies, revealing that temperate coastal cities exhibit a highly reactive pattern while arid inland cities show a response buffered by modern infrastructure. These findings uncover a critical disconnect wherein the populations most at risk are often the last to digitally engage with the threat. This digital proxy provides a new, near real-time method for monitoring public risk perception, offering a vital resource for developing more equitable public health interventions against the escalating threat of extreme heat.

1. Introduction

Extreme heat is among the most severe consequences of climate change and is reported as the leading cause of weather-related mortality (NWS 2024). Rising background temperatures due to climate change are further intensified in cities by the urban heat island effect, which elevates local temperatures above surrounding rural areas (Phelan *et al* 2015, Du *et al* 2021, Lee 2024). This amplification is particularly consequential given that more than 80% of the US population resides in urban areas, which occupy less than 5% of the nation’s land surface; globally, more than half of the population lives in cities that cover less than 3% of land area. The combination of climate-driven warming, urban amplification, and dense population exposure makes cities especially vulnerable to

extreme heat (Reid *et al* 2009, Leal Filho *et al* 2018, Tuholske *et al* 2021).

However, quantifying how urban populations are actually exposed to heat remains challenging. Two limitations are particularly important. The first is the lack of fine-scale temperature measurements within cities. Urban temperatures can vary substantially even across short distances, yet most available observations do not capture this heterogeneity (Mirzaei and Haghighat 2010, Venter *et al* 2021). Satellite-derived land surface temperature is often used in urban heat studies because of its relatively fine spatial resolution, sometimes on the order of tens of meters (Zhang *et al* 2010, Chang *et al* 2023, Lee *et al* 2024). However, satellite data typically trade off spatial and temporal resolution and measure land surface rather than air temperature, making them related

but not directly comparable (Hsu *et al* 2021, Venter *et al* 2021). Urban canopy models and reanalysis products provide alternative estimates, but these are model-based and subject to their own limitations. For instance, urban canopy models often struggle to capture the true heterogeneity of urban forms (Ching 2013, Demuzere *et al* 2017, Nazarian *et al* 2022), and reanalysis products typically lack the spatial resolution and urban-specific physics needed to accurately represent the microclimates where people actually live (Urban *et al* 2021, Lee and Dessler 2024).

The second limitation concerns the measurement of actual human exposure. Even under the same ambient temperature, exposure can differ dramatically across individuals and cities (Kuras *et al* 2017, Yang *et al* 2018). For example, an office worker in a building with air conditioning may experience far less heat stress than a construction worker laboring outdoors. Similarly, cities vary in the extent of infrastructure and resources available to mitigate heat exposure (Heaviside *et al* 2017). As a result, translating ambient conditions into realized human exposure remains inherently difficult. To overcome this, researchers have employed various methods, from using GPS and wearable sensors for direct personal monitoring (Pham *et al* 2020, Friedrich *et al* 2025) to developing agent-based models that simulate daily human mobility and activity (Sonnenschein *et al* 2022, Muslimin and Brasier 2024, Du *et al* 2025). Nevertheless, this challenge is particularly critical, as it represents the gap that the present study seeks to address.

Digital behavioral data such as internet search activity and social media content offer a valuable complementary perspective (Grasso *et al* 2017, Young *et al* 2021, Kim and Kim 2023, Zander *et al* 2023, Lyu *et al* 2024). Unlike traditional meteorological or health surveillance data, these sources directly reflect how populations perceive and respond to heat in real time. Search queries provide indicators of public concern and information-seeking behavior, while social media posts can capture subjective experiences, coping strategies, and localized reports of heat stress. Importantly, these signals are generated at high temporal frequency and often cover broad spatial extents, making them especially useful for characterizing population-level responses in urban areas where direct measurements of exposure are limited.

Previous studies have demonstrated the potential of digital behavioral data for assessing public responses to extreme heat. Analyses of social media platforms, particularly Twitter (now rebranded as X), have shown that heat-related posts can serve as indicators of both perceived risk and adverse health outcomes (Zander *et al* 2023, Elkefi and Tounsi 2024). Google Trends (GT) has likewise been associated with health impacts, such as increased emergency department visits during periods of extreme heat (Li *et al* 2016, Green *et al* 2018, Bogdanovich *et al* 2023, Kim

and Kim 2023). These findings suggest that digital signals do not only reflect abstract concern but can function as population-level proxies of exposure by capturing when and where people actively seek information or assistance in response to heat. This proxy function is especially valuable in urban contexts, where exposure varies across neighborhoods and individuals in ways that are difficult to measure directly. In doing so, digital behavioral data provide a complementary lens on human vulnerability and adaptation that is not easily captured by conventional meteorological or epidemiological observations.

Building on this foundation, the present study investigates how digital behavioral responses to heat vary across diverse urban contexts. The analysis focuses on three distinct domains of public attention—weather awareness, health concerns, and energy demand—captured through internet search activity. A central goal of this work is to consider the diversity of urban responses within a unified framework. Cities differ in their climatic conditions, socioeconomic environments, and adaptive capacities, and these differences shape how populations perceive and respond to heat. By comparing patterns across major metropolitan areas, the study highlights both commonalities and divergences in sensitivity, offering insights into the factors that may drive urban adaptation or vulnerability.

2. Data and methods

2.1. GT data

This study employs internet search activity from GT as a proxy for public attention to heat-related conditions. GT allows the retrieval of normalized indices of search activity for specified terms or topics within defined temporal windows and geographic regions. The platform distinguishes between search terms, which correspond to exact character strings entered by users (e.g. ‘heat wave’), and topics, which are broader categories defined by Google that aggregate semantically related queries across languages, spelling variations, and synonyms (e.g. the topic ‘Heat wave’ encompasses searches for ‘extreme heat wave’, ‘hot outside’, and ‘ola de calor’ in Spanish). Because topics provide a more comprehensive measure of collective search behavior compared to single search terms, this study primarily employed topic-based queries.

Daily-level GT data can only be retrieved for short time intervals of approximately 200 d. Crucially, each of these queries is independently normalized on a 0–100 scale, where 100 represents the peak search interest only within that specific time window. To create a continuous and consistently scaled daily time series for the full nine-year period (2016–2024), a ‘stitching’ method using overlapping queries was therefore employed. By downloading consecutive, overlapping segments, the shared period was used to rescale and merge each window relative

to its neighbors. This process produces a single long-term record that preserves the proportional search interest across the entire nine years. This complete, stitched time series for each search term was then re-normalized on a final 0–100 scale. This ensures that the final time series for each topic is internally consistent, making search interest values comparable across the full study period. This period was chosen to maximize data availability while extending to the most recent year of observation.

GT also enables city-level targeting, and this analysis focuses on thirty major metropolitan areas in the continental United States: New York City (NY), Los Angeles (CA), Chicago (IL), Houston (TX), Phoenix (AZ), Philadelphia (PA), San Diego (CA), Dallas–Fort Worth (TX), Seattle (WA), Denver (CO), Washington, D.C., Boston (MA), Jacksonville (FL), Columbus (OH), Charlotte (NC), San Jose (CA), Austin (TX), Indianapolis (IN), Nashville (TN), Portland (OR), Detroit (MI), Oklahoma City (OK), Atlanta (GA), Miami (FL), San Antonio (TX), El Paso (TX), Louisville (KY), Baltimore (MD), Milwaukee (WI), and Minneapolis (MN). These metropolitan areas were selected for their large populations and broad geographic distribution. Collectively, they account for more than 40% of the total US population and span the full range of US climate zones, making them a robust representation of urban exposure to extreme heat.

An initial set of fifty candidate keywords was compiled to capture various dimensions of heat stress and heat perception. Many of these, however, produced insufficient search activity for robust analysis across all cities. The final selection included nine representative topics: Air conditioning, Dehydration, Emergency department, HVAC system, Humidity, Power outage, Sunburn, Heat wave, and Heat index. Among these, Humidity, Heat wave, and Heat index reflect weather-related searches; Dehydration, Emergency department, and Sunburn capture health-related concerns; and Air conditioning, HVAC system, and Power outage correspond to energy-related behaviors. For each city, category-specific indices were created by averaging the normalized series within each classification, yielding three composite GT indicators—weather, health, and energy—over the nine-year study period.

2.2. Meteorological data

Meteorological conditions were characterized using the GridMet dataset (Abatzoglou 2013), which provides daily gridded surface meteorology over the conterminous United States at a spatial resolution of 4 km. GridMet is produced by statistically downscaling the North American Land Data Assimilation System (NLDAS) (Cosgrove *et al* 2003) to the high-resolution climatological surfaces provided by the PRISM (Parameter-elevation Regressions on Independent Slopes Model) dataset (Daly and Bryant

2013). NLDAS supplies the temporal variability from reanalysis and observationally constrained land surface models at relatively coarse resolution, while PRISM provides fine-scale spatial detail by accounting for topographic and geographic influences on long-term climate normals. By merging these two sources, GridMet generates daily gridded meteorological fields at 4 km resolution over the conterminous United States, offering both the spatial detail and temporal consistency needed for climate and impact studies.

From GridMet, three variables were obtained: daily maximum temperature (T_{\max}), daily minimum temperature (T_{\min}), and specific humidity (sph). Daily mean values were estimated as the average of the corresponding maximum and minimum, and specific humidity was assumed to remain constant throughout the day. These variables were used to derive dry-bulb temperature (DBT) and wet-bulb temperature (WBT), with the latter calculated using the empirical approximation (Stull 2011). In addition, four widely used composite indices of heat stress were computed. The simplified wet-bulb globe temperature (SWBGT) provides an approximation of combined temperature and humidity effects using a reduced formula (Liljegren *et al* 2008). Humidex (Rana *et al* 2013, Diaconescu *et al* 2023) reflects perceived temperature by incorporating both air temperature and vapor pressure. The Heat Index (HI), commonly used by NOAA and the US National Weather Service, estimates perceived heat stress under shaded, light-wind conditions (Hawkins *et al* 2017). The Discomfort Index (DI), originally developed in occupational health research, integrates temperature and humidity to represent thermal discomfort thresholds (Thom 1959, Epstein and Moran 2006). For each metric, daily mean and maximum values were calculated in 4 km resolution.

For each city, daily time series were then constructed by averaging all grid cells within a 50 km radius of the city center, providing representative local meteorological conditions. The final dataset therefore consists of twelve daily metrics for each city: mean and maximum values for DBT, WBT, SWBGT, Humidex, HI, and DI.

2.3. Socioeconomic data

Socioeconomic data were obtained from the American Community Survey (ACS) 5 year estimates for 2023 via the US Census Bureau API. The dataset includes comprehensive demographic, economic, housing, and social indicators for each of the 30 cities in the analysis. Demographic variables include total population, age distribution (percentage 65 years and over, percentage under 18 years), and sex ratio. Economic indicators comprise median household income, poverty rate, unemployment rate, and uninsured population percentage. Educational attainment is measured as the percentage of population 25 years

and older with less than high school education and the percentage with bachelor's degree or higher. Housing characteristics include tenure status (owner-occupied and renter-occupied percentages), year built (percentage built 2010 or later), multi-unit structures (percentage of 5+ units), crowded housing (percentage of occupied units with more than 1.0 persons per room), vacancy rate, and households with no vehicle. Racial and ethnic composition variables include the percentage of White, Black, and Asian populations, as well as Hispanic or Latino percentage. Urban characteristics are represented by population density (persons per square mile), calculated using land area data from the 2023 Census Gazetteer. All percentage variables were computed from raw ACS counts and represent proportions of relevant population denominators, with quality checks implemented to handle missing values and ensure data consistency across all geographic units.

2.4. Generalized additive model (GAM)

This study employs GAMs as the primary statistical framework to assess the nonlinear relationships between meteorological conditions and search activity. GAMs extend generalized linear models by replacing linear predictors with smooth functions, allowing flexible estimation of nonlinear effects without imposing strict parametric assumptions (Wood *et al* 2015, Hastie 2017).

For each city, three GT indices—weather, health, and energy—were treated as dependent variables, while each of 12 meteorological predictors were considered. Separate models were constructed for each outcome–predictor pair. The general form of the model is expressed as equation (1),

$$GT = s(\text{Heat}) + s(\text{Time}) + \varepsilon. \quad (1)$$

In equation (1), GT represents the target GT data (weather, health, or energy), and $s(\text{TemperaHeatture})$ is a spline function of the heat metric predictor. $s(\text{Time})$ is a spline function of time, and ε is the residual error. Spline function for temperature is specified with 8 knots, and the spline function for time is specified with 3 knots per year (total of 30 knots) to account for long-term trends and seasonality. Model estimation was performed using restricted maximum likelihood, which provides stable smoothing parameter selection (Wood 2004). To focus on the heat related outcomes, only the warm season (from March to October) data was utilized for the GAM analysis. Sensitivity tests with varying numbers of initial knots and degrees of freedom yielded no substantive differences in the results.

To identify the most relevant predictors for each domain of search activity, the 12-heat metrics were systematically evaluated. For each city and domain (weather, health, and energy), all candidate metrics were fit as predictors in separate GAMs, and

their explanatory performance was assessed. Model performance was quantified using a combination of standard criteria, including the Akaike information criterion (AIC), the coefficient of determination (R^2), and the area under the receiver operating characteristic curve (AUC). These metrics were then ranked within each city, and the top-ranked predictor was selected as the representative heat metric overall.

2.5. Threshold detection and piecewise linear fit

Threshold detection was performed using piecewise linear regression to identify optimal breakpoints in the temperature-GT response relationships. For each city and domain (weather, health, energy), the algorithm systematically tests breakpoint candidates across the temperature range (0 °C–30 °C) in 500 equally spaced intervals. At each candidate breakpoint, the data is split into two segments, and separate linear regressions are fitted to each segment. The optimal threshold is determined by selecting the breakpoint that minimizes the combined sum of squared errors from both regression lines. This approach identifies the temperature point where the relationship between temperature and GT search volume exhibits the most significant change in slope, effectively capturing the threshold at which heat-related search behavior intensifies. The resulting threshold represents the critical temperature point, while the slope difference (post-threshold slope minus pre-threshold slope) quantifies the magnitude of behavioral change beyond this threshold.

3. Results and discussion

3.1. Heat-related online responses across domains

This study first examined the relationship between daily mean WBT and online search activity across three domains of public attention: weather awareness, health concerns, and energy demand (figure 1). Daily mean WBT is selected as a metric of heat because it was identified as the most informative predictor among multiple candidate indices as it consistently yielded the best explanatory performance across AIC, R^2 and AUC rankings. A full comparison of all 12 evaluated heat metrics can be found in supplementary section S1. Although WBT was not the single top-performing metric in every city or domain, it achieved the strongest overall performance, including the highest mean R^2 , the lowest mean AIC, and the most frequent city-level wins across the three domains.

As seen in the figure 1, the GAMs were able to capture the exposure-response function of the GT search results. The exposure–response functions derived from the GAMs (figures 1(a), (f) and (k)) reveal two distinct features: thresholds and slope differences. Thresholds mark the point at which search activity begins to increase sharply in response to heat, while slope differences represent the amount of

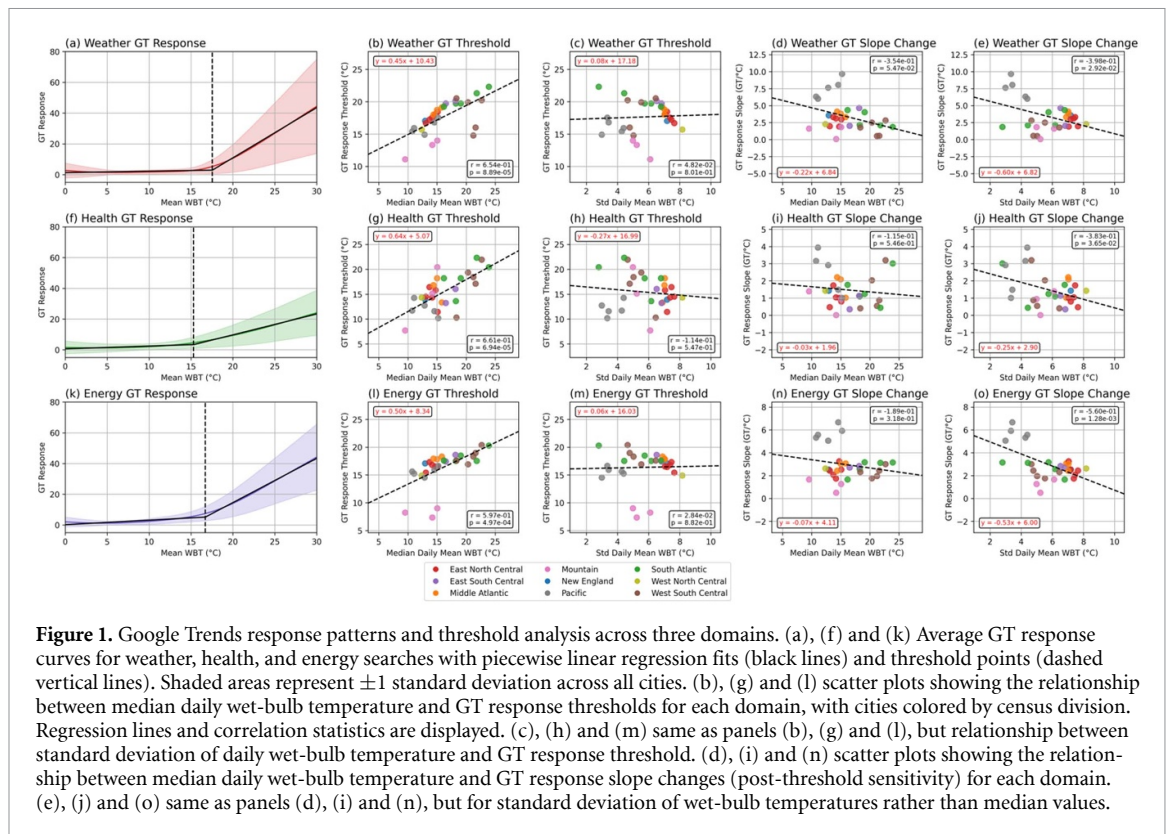


Figure 1. Google Trends response patterns and threshold analysis across three domains. (a), (f) and (k) Average GT response curves for weather, health, and energy searches with piecewise linear regression fits (black lines) and threshold points (dashed vertical lines). Shaded areas represent ± 1 standard deviation across all cities. (b), (g) and (l) scatter plots showing the relationship between median daily wet-bulb temperature and GT response thresholds for each domain, with cities colored by census division. Regression lines and correlation statistics are displayed. (c), (h) and (m) same as panels (b), (g) and (l), but relationship between standard deviation of daily wet-bulb temperature and GT response threshold. (d), (i) and (n) scatter plots showing the relationship between median daily wet-bulb temperature and GT response slope changes (post-threshold sensitivity) for each domain. (e), (j) and (o) same as panels (d), (i) and (n), but for standard deviation of wet-bulb temperatures rather than median values.

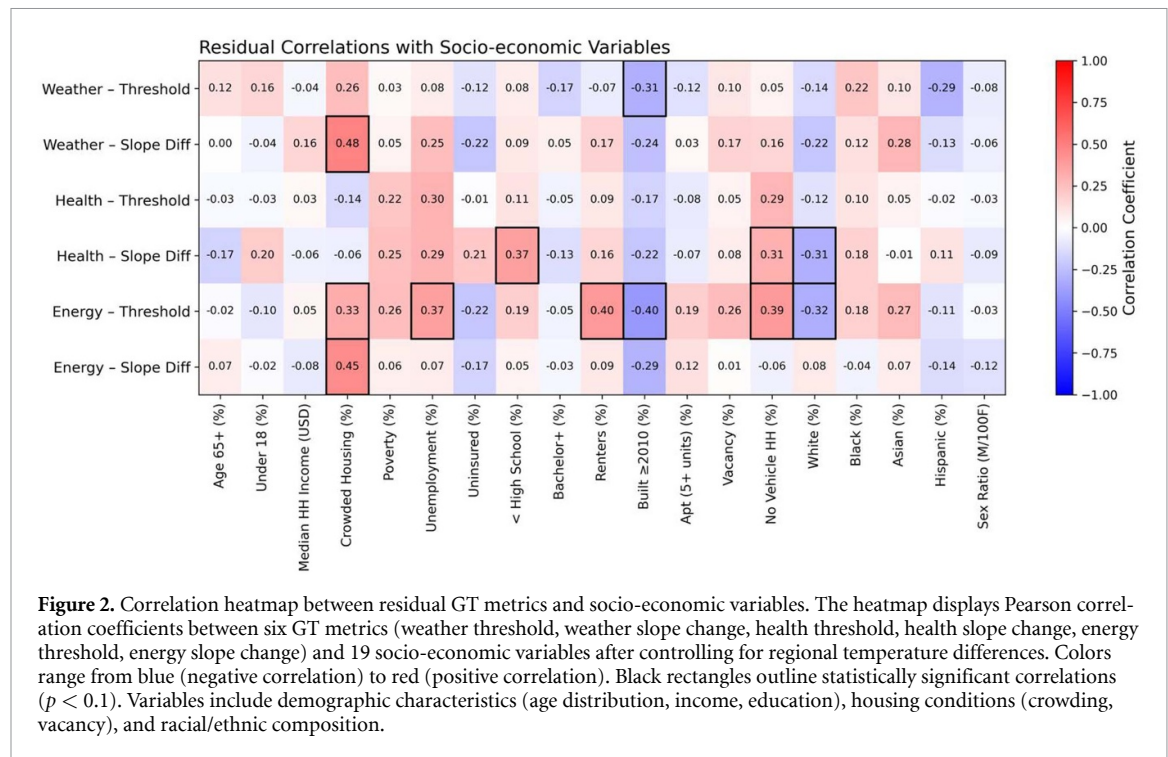
increase in search intensity once this threshold is surpassed. These two parameters together offer insight into how urban populations recognize and respond to escalating heat. Looking at the city-average curves (figures 1(a), (f) and (k)), the threshold for weather, health, and energy related searches were 17.5, 15.3, and 16.7 °C WBT, respectively.

Across 30 US metropolitan areas, thresholds exhibit a systematic pattern. Cities with higher warm-season (Mar to Oct) median WBT, such as Houston, Miami, and Phoenix, display substantially elevated thresholds compared to cooler climates like Boston, Seattle, or Minneapolis. This trend, where warmer cities with higher median WBT have higher thresholds, is consistent in all sectors (figures 1(b), (g) and (l)). This finding indicates that populations in hotter cities do not engage in search activity until substantially more severe conditions occur, a pattern that can be considered as a long-term adaptation to local climatic baselines with physiological acclimatization and social, infrastructural adaptation. From a resilience perspective, this suggests that digital signals carry embedded evidence of acclimatization: people in warmer environments require a higher environmental trigger before engaging in information-seeking behaviors. On the other hand, climate variability (figures 1(c), (h) and (m))—quantified as the standard deviation of daily mean WBT across the warm season—does not show significant relationship with the location of the thresholds.

Slope differences, show weaker associations with median climate (figures 1(d), (i) and (n)) but stronger links to climatic variability (figures 1(e), (j) and (o)). Cities characterized by greater WBT variability exhibit smaller differences in pre- and post-threshold slopes, indicating that higher day-to-day fluctuations of climate dampen the incremental increase in search activity once heat awareness is triggered. This finding suggests that climatic variability, rather than average conditions, plays a larger role in modulating the strength of behavioral responsiveness, possibly because irregular exposure blurs distinctions between ‘normal’ and ‘extreme’ heat events.

Regional anomalies further underscore this point. Cities in the Mountain West, including Denver, El Paso, and Phoenix, display response curves that deviate from the broader national trend. The dry climate in these regions lowers WBT relative to DBT, sometimes producing curves that understate the discomfort experienced by local populations. As a result, thresholds and slopes estimated from WBT alone may not fully capture behavioral responses in arid climates where solar radiation, wind exposure, and cultural practices play a larger role in shaping heat stress.

Together, these results suggest two key insights. First, thresholds scale in relation to background median climate, reflecting a behavioral normalization to recurring heat exposure. Second, slopes are more influenced by climatic variability, suggesting that shorter term fluctuations shape how sharply people react once heat levels surpass awareness thresholds.



3.2. Socioeconomic factors of digital sensitivity

While the local median and variability of climate primarily determine a city's threshold and slope characteristics in its digital heat response, a substantial residual component remains unexplained. This residual underscores the critical role of non-climatic factors in shaping a population's reactive capacity. To isolate these influences, we detrended the six digital sensitivity metrics by removing the effects of both the median and variability (standard deviation) of wet-bulb and DBTs. This extended climate adjustment accounts for differences in baseline thermal conditions as well as day-to-day fluctuations. The resulting climate-adjusted residuals were then correlated with a suite of city-level socioeconomic indicators, and the statistically significant relationships ($p < 0.1$) are summarized in figure 2 and discussed below.

Thresholds—representing the point at which search activity begins to rise sharply with heat—show clear associations with indicators of social and infrastructural vulnerability. In the energy domain, higher thresholds are found in cities with larger shares of crowded housing ($r = 0.33$, $p = 0.08$), unemployment ($r = 0.37$, $p = 0.05$), renters' ratio ($r = 0.40$, $p = 0.03$), and households lacking vehicle access ($r = 0.39$, $p = 0.03$). These relationships indicate that populations facing housing pressure, economic precarity, or limited mobility tend to engage only under more extreme heat conditions. Such delayed awareness may reflect social normalization of discomfort in chronically stressed environments.

Conversely, cities with a larger share of newer housing stock (built after 2010) exhibit lower thresholds for both energy ($r = -0.40$, $p = 0.03$)

and weather searches ($r = -0.31$, $p = 0.09$). Modern infrastructure may facilitate earlier recognition of thermal stress through better indoor comfort expectations, greater access to digital information, and improved energy communication systems. Similarly, a higher proportion of white residents is weakly associated with lower thresholds in the energy domain ($r = -0.32$, $p = 0.09$), consistent with broader disparities in adaptive capacity and access to risk information. Taken together, these findings suggest that early digital engagement is more common in cities with better infrastructure, higher socioeconomic resources, and lower structural vulnerability.

The slope difference of the response, defined as the rate of increase in search activity once the threshold is surpassed, reflects how strongly populations react once aware of danger. Across the domains, this slope difference displays distinct yet systematic links to socioeconomic structure. In the weather and energy domains, the steepest slope differences are observed in cities with higher housing density and crowded conditions ($r = 0.48$ and 0.45 , respectively). These findings indicate that dense living environments amplify collective attention and accelerate digital responses once heat becomes salient. Such amplification likely arises from shared exposure, overlapping communication networks, and stronger social contagion of awareness.

In contrast, the health domain shows a different pattern. Slope differences increase with the share of residents with less than high-school education ($r = 0.37$, $p = 0.04$) and with households lacking vehicle access ($r = 0.31$, $p = 0.09$), while decreasing with the white population share ($r = -0.31$,

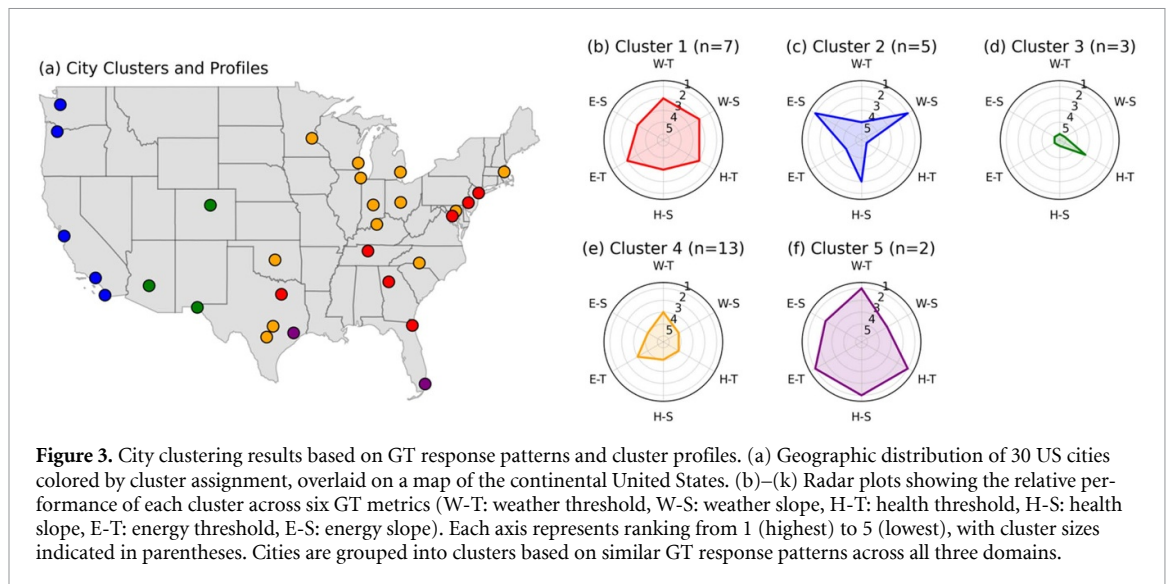


Figure 3. City clustering results based on GT response patterns and cluster profiles. (a) Geographic distribution of 30 US cities colored by cluster assignment, overlaid on a map of the continental United States. (b)–(k) Radar plots showing the relative performance of each cluster across six GT metrics (W-T: weather threshold, W-S: weather slope, H-T: health threshold, H-S: health slope, E-T: energy threshold, E-S: energy slope). Each axis represents ranking from 1 (highest) to 5 (lowest), with cluster sizes indicated in parentheses. Cities are grouped into clusters based on similar GT response patterns across all three domains.

$p = 0.10$). These correlations suggest that, for health-related searches, the most reactive digital responses occur in populations facing informational or mobility constraints—cities that may recognize the risk later but respond intensely once the threat becomes unavoidable.

Together, the results outline a dual social structure in digital sensitivity to heat. Cities with stronger infrastructure, economic affluence, and higher education tend to display lower thresholds but smaller slope differences, reflecting moderated or buffered responses after awareness is triggered. In contrast, socially or infrastructurally vulnerable cities exhibit higher thresholds (delayed awareness) and steeper slope differences, denoting rapid, reactive surges in attention once conditions become critical.

3.3. Typologies of digital response to heat

To identify common patterns in how urban populations respond to extreme heat, this study applied k-means clustering to the six digital sensitivity metrics. The optimal number of clusters was determined to be five ($k = 5$). The six metrics used for clustering are combinations of three search domains (W for Weather, H for Health, E for Energy) and two response characteristics (T for Threshold, S for Slope difference). The clustering result and their characteristics are depicted in figure 3.

Cluster 1, primarily composed of cities in the Eastern and Southern US, demonstrates a moderately adapted response to heat. Their search thresholds are consistently slightly above the 30-city average, particularly for health-related topics (H-T Avg: 18.46 vs Overall Avg: 15.41), indicating a degree of acclimatization shaped by their humid subtropical and continental climates. Once this higher threshold for concern is met, their response slopes are average, suggesting a standard, non-extreme reaction. As the

socioeconomic analysis did not identify any statistically significant distinguishing variables for this group, their behavioral pattern appears to be primarily shaped by long-term climatic adaptation, establishing them as a key baseline for comparison.

Cluster 2 represents the five West Coast cities. They exhibit a pattern of high alert (low threshold and high slope difference), likely due to their generally temperate climates where extreme heat is a less frequent event. Their populations begin searching for information at relatively cool temperatures, reflected in search thresholds that are consistently lower than average. Once this low trigger point is crossed, their digital response is extremely strong, with the highest response slopes of any cluster for both weather (W-S Avg: 7.87 vs Overall Avg: 3.34) and energy. This intense, information-seeking reaction is likely enabled by their distinct socioeconomic profile, which is characterized by significantly lower rates of poverty, uninsured individuals, and housing vacancy.

Cluster 3 are Denver, El Paso, and Phoenix. This group of three arid-climate cities presents a digital response that is fundamentally different from the other typologies. Their engagement with heat shows a nearly flat, almost linear pattern, lacking the distinct non-linear activation point seen in other cities. Because of this, the very concept of a search ‘threshold’—a clear point where public concern suddenly activates—is less meaningful for this cluster. Consequently, this study records the lowest possible response slopes of any group, signifying a near-total absence of an escalating digital reaction as heat intensifies. This profoundly muted response is likely a result of effective infrastructural buffering. The socioeconomic data shows these cities have a significantly higher percentage of modern housing (built since 2000). This suggests that residents are so well-insulated from the heat’s impact by high-quality infrastructure, particularly effective air conditioning,

that the urgency to seek further information online never materializes, even as objective heat conditions worsen.

As the largest and most geographically diverse cluster, Cluster 4 is defined by a consistently silent digital response. Its thirteen cities exhibit average to low search thresholds, indicating a normal point of initial awareness of heat. However, their response slopes are consistently low across all domains. This muted reaction is not driven by any specific climate type, nor were any significant socioeconomic variables found to distinguish this cluster. Their defining trait is a subdued behavioral response whose drivers are not captured by the variables in this study, pointing to the influence of other local factors, such as the nature of media coverage or public messaging, in shaping a city's collective digital reaction.

Cluster 5 consists of Houston and Miami. This pair of hot, humid subtropical cities presents a unique, highly adapted, yet acutely health-aware pattern. They exhibit the highest search thresholds of any cluster across all domains, a clear signal of extreme acclimatization to their chronically hot climate. Despite this high tolerance, their response slope for health topics is the most intense of any group (H-S Avg: 3.64 vs. Overall Avg: 1.35). This suggests that while residents are accustomed to general heat, their lived experience has conditioned them to be highly vigilant about specific health risks when conditions become extreme. This health-vigilant behavior is layered on top of a socioeconomic profile of significant vulnerability, including substantially lower median household incomes, creating a unique digital signature defined by the intersection of climatic adaptation and acute health risk awareness.

4. Summary and conclusion

As climate change amplifies the frequency and severity of extreme heat events, understanding the real-time human response has become a critical priority. Traditional approaches often overlook the dynamic interplay between physical exposure and public perception. This study demonstrates that digital search behavior can serve as a powerful, near-real-time proxy for collective heat awareness. By modeling search patterns across 30 US cities, it reveals that the pathway from heat exposure to public action is shaped by each city's distinctive climatic, infrastructural, and socioeconomic landscape. A central contribution of this work is the disentanglement of two interacting components of urban heat response: climate-driven adaptation and socioeconomically driven vulnerability. The analysis shows that a city's median climate largely determines its baseline search threshold, while climate variability governs the response slope.

Upon this climate-defined baseline, analysis with climate-adjusted residuals expose the hidden layer of socioeconomic influence: cities with older housing, limited mobility, and higher unemployment tend to be the last to engage digitally with rising heat risk, whereas well-resourced cities respond earlier and more actively. The results also highlight how education and housing density strengthen the responsiveness slope, suggesting that knowledge resources and spatial connectivity enable faster collective awareness once heat becomes salient. Finally, clustering of the climate-adjusted residuals identifies five urban typologies that synthesize these climatic and social patterns. These typologies represent gradients rather than fixed categories, illustrating that digital heat resilience varies widely across contexts. Cities with chronically hot climates and high thresholds may require targeted messaging to counteract complacency, whereas cooler regions with low thresholds may need calibrated communication to prevent undue alarm during rare heat events.

This research refines our understanding of how climate shapes urban heat awareness. While a city's baseline sensitivity largely reflects its local climate, the key contribution lies in analyzing climate-adjusted residuals, which reveal how socioeconomic and infrastructural factors drive differences in digital responsiveness among climatically similar cities. By clustering these residuals, the study identifies five distinct urban typologies that synthesize gradients of adaptation and vulnerability rather than fixed categories. Such diversity highlights that no single model of public heat response applies universally—each city's digital awareness reflects a unique intersection of climatic exposure, infrastructure, and social capacity. Rather than prescribing uniform solutions, these findings offer a diagnostic foundation for designing region-specific adaptation and communication strategies. The results underscore equity dimensions: delayed awareness among vulnerable populations calls for targeted communication systems that move beyond general alerts to reach digitally underserved neighborhoods. Structuring results across Weather, Health, and Energy domains enables a multi-sectoral lens for near-real-time monitoring of public concern, advancing beyond static vulnerability mapping toward dynamic, data-driven preparedness.

Finally, the results indicate that adaptation may lag the accelerating pace of climate change. As people normalize heat relative to familiar conditions, awareness thresholds risk adjusting too slowly to increasing extremes. Sustaining resilience thus requires viewing adaptation as a moving target—where public awareness, infrastructure, and institutions evolve faster than the changing climate itself. In this sense, the digital sphere does not merely mirror physical exposure; it forms an active feedback

channel through which societal concern, communication, and policy response can co-evolve. Recognizing and operationalizing this digital-to-physical feedback loop represents the next frontier in aligning collective perception with the realities of a warming world.

This study is not without limitations. GT data are aggregated and may not represent all demographic groups. Although prior work links search data to real-world outcomes like ER visits or power outages, this evidence remains limited; future research should validate these digital signals against localized health and infrastructure data. The nine-year record also restricts detection of long-term changes in public awareness, highlighting the need for longer time series. Socioeconomic indicators used here, while broad, may not fully capture adaptive capacity—metrics such as AC penetration could add valuable insight. Finally, differences in formal heat-action plans across cities, such as in San Diego, Phoenix, or New York, suggest that institutional preparedness may shape digital awareness; integrating such planning data offers an important avenue for future study.

In conclusion, by decoding the digital echoes of our collective response to heat, this study provides a new lens through which to view and measure urban heat sensitivity. It demonstrates that the digital landscape is not a passive mirror of the physical world but an active arena where vulnerability and adaptation are expressed. By learning to listen to these digital signals, we can better anticipate and address the tangible harms of a warming planet, ultimately building more just and effective strategies for the climate challenges ahead.


Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://trends.google.com/trends/>; <https://data.census.gov/>; www.climatologylab.org/gridmet.html.

Acknowledgment

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research's Urban Integrated Field Laboratories CROCUS project research activity, under Award Number DE-SC0023226.

Author contribution

Jangho Lee  0000-0002-8942-1092
Conceptualization (lead), Data curation (lead),
Formal analysis (lead), Investigation (lead),
Methodology (lead), Resources (equal),
Software (equal), Supervision (equal),
Validation (equal), Visualization (equal), Writing –

original draft (equal), Writing – review & editing (equal)

References

- Abatzoglou J T 2013 Development of gridded surface meteorological data for ecological applications and modelling *Int. J. Climatol.* **33** 121–31
- Bogdanovich E, Guenther L, Reichstein M, Frank D, Ruhrmann G, Brenning A, Denissen J M C and Orth R 2023 Societal attention to heat waves can indicate public health impacts *Weather Clim. Soc.* **15** 557–69
- Chang Y, Xiao J, Li X and Weng Q 2023 Monitoring diurnal dynamics of surface urban heat island for urban agglomerations using ECOSTRESS land surface temperature observations *Sustain. Cities Soc.* **98** 104833
- Ching J K 2013 A perspective on urban canopy layer modeling for weather, climate and air quality applications *Urban Clim.* **3** 13–39
- Cosgrove B A *et al* 2003 Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project *J. Geophys. Res.* **108**
- Daly C and Bryant K 2013 *The PRISM Climate and Weather System—an Introduction* (PRISM Climate Group) p 2
- Demuzere M *et al* 2017 Impact of urban canopy models and external parameters on the modelled urban energy balance in a tropical city *Q. J. R. Meteorol. Soc.* **143** 1581–96
- Diaconescu E, Sankare H, Chow K, Murdock T Q and Cannon A J 2023 A short note on the use of daily climate data to calculate Humidex heat-stress indices *Int. J. Climatol.* **43** 837–49
- Du H *et al* 2021 Simultaneous investigation of surface and canopy urban heat islands over global cities *ISPRS J. Photogramm. Remote Sens.* **181** 67–83
- Du Y, Zhang A, Zhen Q, Wei S, Jing H, Zheng Y and Zhang Q 2025 Remapping spatio-temporal cumulative heat stress of centralized rural community: an agent-based modeling approach *Build. Environ.* **284** 113429
- Elkefi S and Tounsi A 2024 Examining public perceptions and concerns about the impact of heatwaves on health outcomes using Twitter data *J. Clim. Change Health* **17** 100320
- Epstein Y and Moran D S 2006 Thermal comfort and the heat stress indices *Ind. Health* **44** 388–98
- Friedrich J, Schick T S, Mess F and Blaschke S 2025 Wearable device-based interventions in heat-exposed outdoor workers—a scoping review and an explanatory intervention model *BMC Public Health* **25** 2893
- Grasso V, Crisci A, Morabito M, Nesi P and Pantaleo G 2017 Public crowdsensing of heat waves by social media data *Adv. Sci. Res.* **14** 217–26
- Green H K, Edeghere O, Elliot A J, Cox I J, Morbey R, Pebody R, Bone A, McKendry R A and Smith G E 2018 Google search patterns monitoring the daily health impact of heatwaves in England: how do the findings compare to established syndromic surveillance systems from 2013 to 2017? *Environ. Res.* **166** 707–12
- Hastie T J 2017 Generalized additive models *Statistical Models in S* pp 249–307 (available at: <https://www.taylorfrancis.com/chapters/edit/10.1201/9780203738535-7/generalized-additive-models-trevor-hastie>)
- Hawkins M D, Brown V and Ferrell J 2017 Assessment of NOAA National Weather Service methods to warn for extreme heat events *Weather Clim. Soc.* **9** 5–13
- Heaviside C, Macintyre H and Vardoulakis S 2017 The urban heat island: implications for health in a changing environment *Curr. Environ. Health Rep.* **4** 296–305
- Hsu A, Sheriff G, Chakraborty T and Many D 2021 Disproportionate exposure to urban heat island intensity across major US cities *Nat. Commun.* **12** 2721
- Kim Y and Kim Y 2023 Global regionalization of heat environment quality perception based on K-means

- clustering and Google trends data *Sustain. Cities Soc.* **96** 104710
- Kuras E R *et al* 2017 Opportunities and challenges for personal heat exposure research *Environ. Health Perspect.* **125** 085001
- Leal Filho W, Icaza L E, Neht A, Klavins M and Morgan E A 2018 Coping with the impacts of urban heat islands. A literature based study on understanding urban heat vulnerability and the need for resilience in cities in a global climate change context *J. Clean. Prod.* **171** 1140–9
- Lee J 2024 Assessment of US urban surface temperature using GOES-16 and GOES-17 data: urban heat island and temperature inequality *Weather Clim. Soc.* **16** 315–29
- Lee J, Berkelhammer M, Wilson M D, Love N and Cintron R 2024 Urban land surface temperature downscaling in Chicago: addressing ethnic inequality and gentrification *Remote Sens.* **16** 1639
- Lee J and Dessler A E 2024 Improved surface urban heat impact assessment using GOES satellite data: a comparative study with ERA-5 *Geophys. Res. Lett.* **51** e2023GL107364
- Li T, Ding F, Sun Q, Zhang Y and Kinney P L 2016 Heat stroke internet searches can be a new heatwave health warning surveillance indicator *Sci. Rep.* **6** 37294
- Liljegren J C, Carhart R A, Lawday P, Tschopp S and Sharp R 2008 Modeling the wet bulb globe temperature using standard meteorological measurements *J. Occup. Environ. Hyg.* **5** 645–55
- Lyu F, Zhou L, Park J, Baig F and Wang S 2024 Mapping dynamic human sentiments of heat exposure with location-based social media data *Int. J. Geogr. Inf. Sci.* **38** 1291–314
- Mirzaei P A and Haghighat F 2010 Approaches to study urban heat island—abilities and limitations *Build. Environ.* **45** 2192–201
- Muslimin R and Brasier N 2024 An analysis of outdoor thermal comfort impact on pedestrian walking patterns at sub-hour timescales with agent-based modelling *Smart Sustain. Built Environ.* (<https://doi.org/10.1108/SASBE-06-2023-0158>)
- Nazarian N *et al* 2022 Integrated assessment of urban overheating impacts on human life *Earth's Future* **10** e2022EF002682
- NWS 2024 Weather related fatality and injury statistics (available at: www.weather.gov/hazstat)
- Pham S, Yeap D, Escalera G, Basu R, Wu X, Kenyon N J, Hertz-Picciotto I, Ko M J and Davis C E 2020 Wearable sensor system to monitor physical activity and the physiological effects of heat exposure *Sensors* **20** 855
- Phelan P E, Kaloush K, Miner M, Golden J, Phelan B, Silva III H and Taylor R A 2015 Urban heat island: mechanisms, implications, and possible remedies *Annu. Rev. Environ. Resour.* **40** 285–307
- Rana R, Kusy B, Jurdak R, Wall J and Hu W 2013 Feasibility analysis of using humidex as an indoor thermal comfort predictor *Energy Build.* **64** 17–25
- Reid C E, O'Neill M S, Gronlund C J, Brines S J, Brown D G, Diez-Roux A V and Schwartz J 2009 Mapping community determinants of heat vulnerability *Environ. Health Perspect.* **117** 1730–6
- Sonnenschein T, Scheider S, de Wit G A, Tonne C C and Vermeulen R 2022 Agent-based modeling of urban exposome interventions: prospects, model architectures, and methodological challenges *Exposome* **2** osac009
- Stull R 2011 Wet-bulb temperature from relative humidity and air temperature *J. Appl. Meteorol. Climatol.* **50** 2267–9
- Thom E C 1959 The discomfort index *Weatherwise* **12** 57–61
- Tuholske C, Caylor K, Funk C, Verdin A, Sweeney S, Grace K, Peterson P and Evans T 2021 Global urban population exposure to extreme heat *Proc. Natl Acad. Sci.* **118** e2024792118
- Urban A *et al* 2021 Evaluation of the ERA5 reanalysis-based Universal Thermal Climate Index on mortality data in Europe *Environ. Res.* **198** 111227
- Venter Z S, Chakraborty T and Lee X 2021 Crowdsourced air temperatures contrast satellite measures of the urban heat island and its mechanisms *Sci. Adv.* **7** eabb9569
- Wood S N 2004 Stable and efficient multiple smoothing parameter estimation for generalized additive models *J. Am. Stat. Assoc.* **99** 673–86
- Wood S N, Goude Y and Shaw S 2015 Generalized additive models for large data sets *J. R. Stat. Soc. C* **64** 139–55
- Yang L E, Hoffmann P, Scheffran J, Rühle S, Fischereit J and Gasser I 2018 An agent-based modeling framework for simulating human exposure to environmental stresses in urban areas *Urban Sci.* **2** 36
- Young J C, Arthur R, Spruce M and Williams H T P 2021 Social sensing of heatwaves *Sensors* **21** 3717
- Zander K K, Rieskamp J, Mirbabaie M, Alazab M and Nguyen D 2023 Responses to heat waves: what can Twitter data tell us? *Nat. Hazards* **116** 3547–64
- Zhang P, Imhoff M L, Wolfe R E and Bounoua L 2010 Characterizing urban heat islands of global settlements using MODIS and nighttime lights products *Can. J. Remote Sens.* **36** 185–96