5-2022

Sentiment analysis of weather-related tweets from cities within hot climates

Yuliya DZYUBAN
Singapore Management University, ydzyuban@smu.edu.sg

Graces N. Y. CHING
Singapore Management University, gracesching@smu.edu.sg

Sin Kang YIK
Singapore Management University, skyik@smu.edu.sg

Adrian J. TAN
Singapore Management University

Peter J. CRANK
Singapore Management University, petercrank@smu.edu.sg

See next page for additional authors

Citation

DZYUBAN, Yuliya; CHING, Graces N. Y.; YIK, Sin Kang; TAN, Adrian J.; CRANK, Peter J.; BANERJEE, Shreya; PEK, Rachel Xin Yi; and Winston T. L. CHOW. Sentiment analysis of weather-related tweets from cities within hot climates. (2022). *Weather, Climate, and Society*. 14, (4), 1133-1145.
Available at: https://ink.library.smu.edu.sg/cis_research/37

Sentiment Analysis of Weather-Related Tweets from Cities within Hot Climates

YULIYA DZYUBAN,[a] GRACES N. Y. CHING,[a] SIN KANG YIK,[a] ADRIAN J. TAN,[a] PETER J. CRANK,[a]
SHREYA BANERJEE,[a] RACHEL XIN YI PEK,[a] AND WINSTON T. L. CHOW[a,b]

[a] *College of Integrative Studies, Singapore Management University, Singapore*
[b] *School of Social Sciences, Singapore Management University, Singapore*

(Manuscript received 26 October 2021, in final form 19 May 2022)

ABSTRACT: Evidence exists that exposure to weather hazards, particularly in cities subject to heat island and climate change impacts, strongly affects individuals' physical and mental health. Personal exposure to and sentiments about warm conditions can currently be expressed on social media, and recent research noted that the geotagged, time-stamped, and accessible social media databases can potentially be indicative of the public mood and health for a region. This study attempts to understand the relationships between weather and social media sentiments via Twitter and weather data from 2012 to 2019 for two cities in hot climates: Singapore and Phoenix, Arizona. We first detected weather-related tweets, and subsequently extracted keywords describing weather sensations. Furthermore, we analyzed frequencies of most used words describing weather sensations and created graphs of commonly occurring bigrams to understand connections between them. We further explored the annual trends between keywords describing heat and heat-related thermal discomfort and temperature profiles for two cities. Results showed significant relationships between frequency of heat-related tweets and temperature. For Twitter users exposed to no strong temperature seasonality, we noticed an overall negative cluster around hot sensations. Seasonal variability was more apparent in Phoenix, with more positive weather-related sentiments during the cooler months. This demonstrates the viability of Twitter data as a rapid indicator for periods of higher heat experienced by public and greater negative sentiment toward the weather, and its potential for effective tracking of real-time urban heat stress.

SIGNIFICANCE STATEMENT: Social media such as Twitter allow individuals to broadcast their opinions in real time, including perceptions and sensations related to weather events. Evidence from two cities exposed to hot weather—one equatorial and one desert subtropical—indicates that tweets were sensitive to seasonal temperature differences even within a small range. For Twitter users exposed to no strong temperature seasonality, generally negative sentiments to hot weather were seen year-round. In Phoenix with more pronounced seasonality, tweets were more positive in sentiment during the cooler months. This result shows promise for the medium as a rapid real-time indicator—or a snapshot—for societal sentiment to weather events.

KEYWORDS: Social Science; Annual variations; Climate variability; Tropical variability; Seasonal effects; Societal impacts

1. Introduction

Heat exposure is a known health hazard globally and is the top weather-related cause of mortality in the United States each year (Vaidyanathan et al. 2020; EPA 2021). Increasing global temperatures from climate change are very likely to result in greater human exposure to hot weather and extreme heat in the future (Ebi et al. 2021), especially in cities and settlements where the combination of local urban heat islands and regional warming exacerbates risks of urban "overheating" (Santamouris and Kolokotsa 2015; Wouters et al. 2017). Furthermore, heat stress and other related illnesses are a significant threat to human well-being (McGregor and Vanos 2018). Yet, in addition to the physical illness threat from heat there is a psychological health concern component. Research has shown heat in combination with urbanization is connected to various psychological and mental health outcomes such as irritability and aggression (Kuo and Sullivan 2001), while March et al. (2008) have shown the effect of urbanization leading to various forms of psychosis. Urban–rural differences in schizophrenia occurrence were confirmed as being statistically higher by Vassos et al. (2016). Furthermore, Wang et al. (2014) found that there are more manic mental health disorders (such as schizophrenia, manic depressive, and bipolar disorders) hospital admissions during heatwaves in Toronto, Canada, than during other times of the year. Other research in cities within Vietnam, Australia, Taiwan, Israel, Brazil, and South Korea show similar results of heat being connected to negative mental health outcomes (Aviv et al. 2011; Hansen et al. 2008; Kim et al. 2014; Sung et al. 2011; Sung et al. 2013; Trang et al. 2016; Volpe and Del Porto 2006). Thus, connections between extreme heat and increased hospitalizations for mental health conditions have been made before, though specific causal pathways have yet to be determined.

Social media can be useful resources to gauge the public's emotions, opinions, and mood on various topics. These include their expression toward heat and weather events. Twitter is a relatively popular social media platform that is increasingly being used in academic research into sentiment analysis on public opinion, with major reasons being data accessibility and clear

Denotes content that is immediately available upon publication as open access.

Corresponding author: Yuliya Dzyuban, ydzyuban@smu.edu.sg

DOI: 10.1175/WCAS-D-21-0159.1

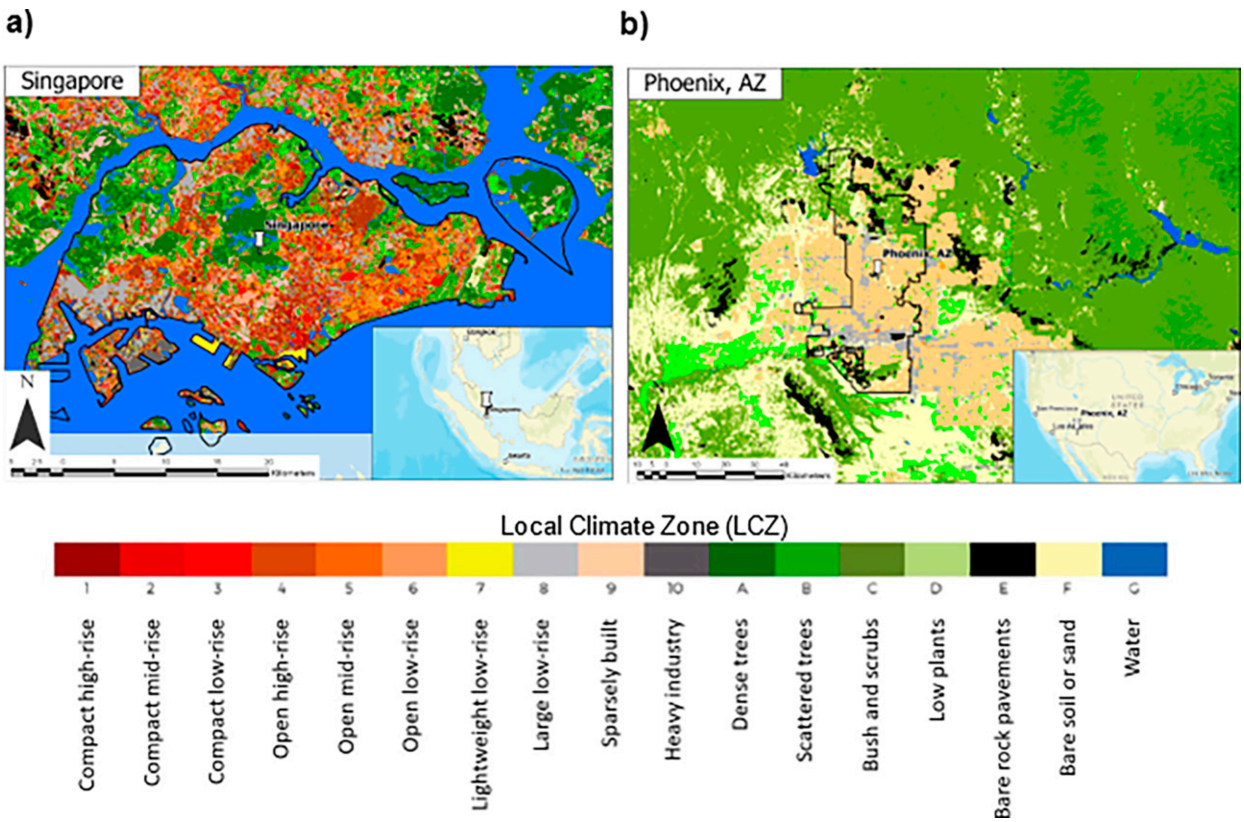

FIG. 1. Geographic location and land-use zone classifications of (a) Singapore [adopted from Mughal et al. (2019)] and (b) Phoenix [adopted from Demuzere et al. (2020)].

geotagging of tweets by region and time (Dahal et al. 2019). Several studies examined how Twitter can indicate people's moods in relation to the weather (Bujisic et al. 2019; Giuffrida et al. 2020), as well as to understand the public's opinion on climate change (Veltri and Atanasova 2015; Fownes et al. 2018; Cody et al. 2015). Roberts (2017) used Twitter to gather the public's opinion on the use of green spaces in urban areas, while Murakami et al. (2016) found strong relationships between heat-related tweets and temperature, and temperature differences, in the Tokyo metropolitan area, and showed that social media data can be a useful resource for estimating intraurban temperature variations.

Understanding public perception, opinions, and mood associated with temperature can complement existing urban climate observational and modeling platforms in collecting relevant data and information toward developing heat mitigation policies and strategies in the warming world. This is especially acute for settlements in hot climates facing increasing climate risks from intensifying heat. For this study, we selected two warm-climate cities: Singapore (a city-state in Southeast Asia) and Phoenix, Arizona. Both of them are considered to be hot but differ in their climate classification, cultural, and political contexts. By examining tweets originating within these two cities, the paper focuses on the following research questions:

1) What is the relationship between weather and weather-related tweets in cities with hot climates?
2) How and what are the most common sentiments associated with weather in cities with hot climates?
3) How do contextual differences between the cities influence weather-related Twitter usage?

2. Method

a. Study areas

This paper focuses on two cities: Singapore (Köppen–Geiger climate: Af) and Phoenix (Köppen–Geiger climate: BWh) (Kottek et al. 2006); both cities with hot climates but differing in atmospheric moisture levels. Singapore is an equatorial humid tropical city (1.35°N, 103.82°E), whereas Phoenix is a desert city (33.45°N, 112.07°W) situated in the hot, arid midlatitudes (Fig. 1). Singapore exhibits consistently high and uniform mean monthly temperatures (long-term annual mean ~27.5°C) and high total annual precipitation (long-term mean total annual rainfall ~2165.9 mm) (Roth and Chow 2012; Meteorological Service Singapore 2021). The climate is characterized by the northeast (NE) and southwest (SW) monsoons from December to early March and from June to September, respectively, separated by intermonsoonal periods from late March to May and from October to November. Phoenix, on the other hand, experiences hot summers and mild winters. Annual average maximum temperatures typically exceed 30.8°C, with summer temperatures exceeding 43.8°C in June. The average annual precipitation is 210.88 mm (Connors et al. 2013; U.S. Climate Data 2022).

Singapore is comparatively denser than Phoenix in terms of population and mean building heights. Both cities have urban heat islands of substantial size and intensity (Roth and Chow 2012; Brazel et al. 2007). With an expected growth in urbanization, Chow et al. (2012) point out that this would worsen the thermal comfort of the inhabitants and increase their vulnerability to heat.

b. Data collection

Twitter is a popular social media and microblogging platform that enables users to write, post, and share messages (also known as "tweets"). It was launched in 2006, and it has reaches 397 million active users globally (Statista 2021a), with the United States having the largest Twitter user base of 73 million (Statista 2021b). In Asia, Twitter is projected to grow to 40 million users by 2025, and in Singapore, it is the sixth most popular social media application with over 1.89 million (Statista 2021c). With the widespread use of Twitter in both regions, the information generated by users offers an opportunity to investigate individuals' sentiments and perceptions on various topics.

We collected two data types in this study. First, 83 085 tweets (34 899 for Singapore, and 48 186 for Phoenix) including "weather" were queried from 2012 to 2019 using Twitter's application program interface (https://developer.twitter.com/en/docs/twitter-api). Second, meteorological data containing hourly weather information from the same time period were extracted from government websites such as the Meteorological Services Singapore (http://www.weather.gov.sg/home/) and National Weather Service (https://www.weather.gov/) for Singapore and Phoenix, respectively.

The Twitter datasets including only tweets made in English were scraped using the R AcademicTwitteR package (https://cran.r-project.org/web/packages/academictwitteR/index.html). Information requested from Twitter includes the following:

- author "ID" (a unique identifier given to each user),
- tweet time (date and time at which the tweet was created),
- Twitter text [up to 140 (before November 2017) or 280 characters of the status update], and
- text source (whether it was shared through the Twitter mobile platform or other social media platforms).

c. Data analysis

A random sample of 400 tweets selected from the raw dataset for each country indicated a relevancy of 42.5% and 78% for Phoenix and Singapore, respectively (refer to appendix B for relevance criteria). The Twitter dataset underwent a filtering process to clean up the data collected. A set of inclusion and exclusion criteria rules were created and applied to improve the relevancy of the dataset. First, any duplicates, official weather announcements (e.g., tweets from municipal weather agencies) and bot advertisements were removed as our research focuses on tweets made by individuals. The tweets were removed by manually identifying sources and users that gave repetitive

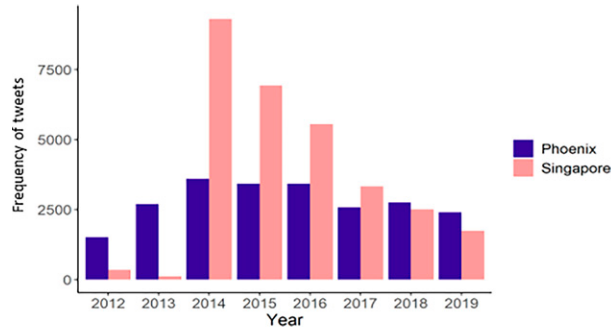

FIG. 2. Annual frequency of weather-related tweets for Singapore (coral fill) and Phoenix (indigo fill) between 2012 and 2019.

and frequent weather forecasts tweets. The next layer of filtering involved assessing and consolidating rules for tweets that were irrelevant based on the random sampling of the raw dataset. A list of irrelevant words, hashtags, and phrases that showed frequent occurrences were identified. In these cases, there were alternative uses of the word "weather" such as "under the weather," "weather the storm," and "fair weather fans" (e.g., sports-related tweets in Arizona). Other common topics that include asking "how's the weather" and tweets referring to locations outside of Singapore or Phoenix were omitted to capture people's perception of their current weather. Retweets of official weather advisories by individuals, alternative spellings to common words (e.g., hot vs hawt), and acronyms (e.g., gtg, omg) were included as long as they indicated a clear opinion toward the weather. In Singapore, colloquial slangs and phrases are commonly used in tweets; hence they were deemed relevant. After applying the inclusion and exclusion criteria described (refer to appendix B), a total of 52 240 tweets—29 832 from Singapore and 22 408 from Phoenix—remained in the dataset. A relevance check on 400 random tweets for each city showed 95% relevance for Singapore and 87% for Phoenix, which we deemed acceptable for further analysis.

Text string data were tokenized and cleaned, with one and two tokens per row and stop words removed with the R-tidytext "stop_words" function. The two-tokens-per-row dataset was used to generate bigrams for two-word connections with occurrences above 30 times using R-widydr.

We screened the one-token-per-row dataset for 100 weather descriptors typically used to describe weather sensations. Seventy-six climatic adjectives adopted from Liu et al. (2020) provide microclimatic descriptors about thermal intensity, thermal sensation, solar radiation, wind, humidity, and thermal pleasure, with an additional 24 words we deemed as important weather descriptors (The list of these descriptors is located in appendix A). Of 20 most often used weather-related descriptors for both cities, we selected words describing "hot" and "cold" discomfort. Heat-related sentiments included "hot," "heat," "sweat," "sweating," "scorching," "burning," "melting"; and cold-related sentiments were "cold," "freezing," "chill," "polar," and "snowy." We screened bigrams with occurrences above 20 times for non-thermal-comfort-related meanings of the weather-related descriptors that we used to identify hot- and cold-related sensations. Since bigram "hot chocolate"

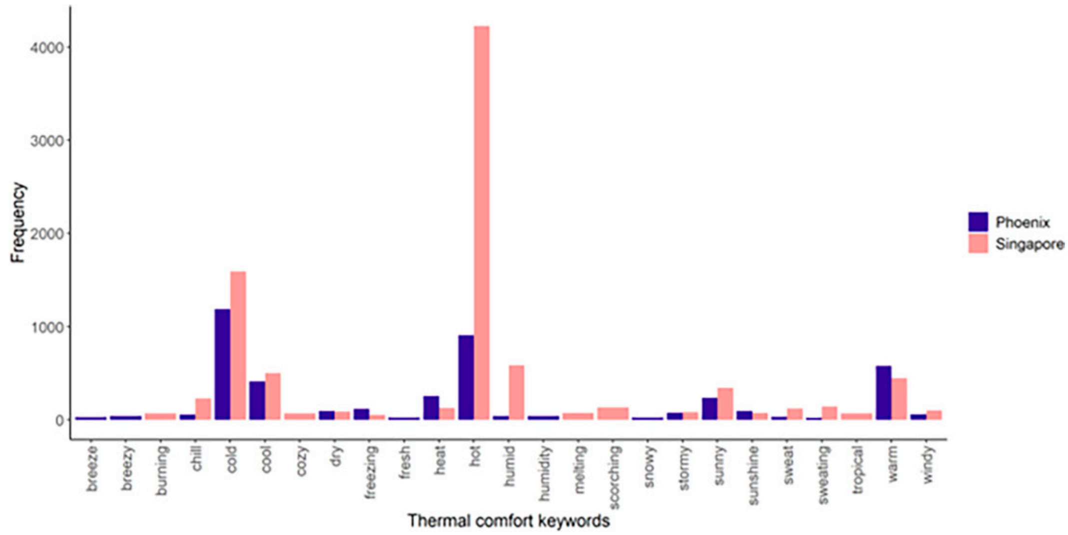

FIG. 3. Frequency of the top 20 most often used thermal-comfort-related descriptors for Singapore (coral fill) and Phoenix (indigo fill) between 2012 and 2019.

occurred above 20 times in our dataset, we excluded tweets containing this word combination from the dataset classifying hot and cold sensations. Remaining keywords classified into hot and cold sensations were used to generate monthly and hourly time series of hot and cold sensations and air temperature (Ta) and wet bulb temperature (Tw) means and to conduct Spearman's rho correlation analysis. Ta is a commonly used weather variable in similar studies (Young et al. 2021; Murakami et al. 2016). Tw is useful for understanding the evaporative cooling potential of the body, which is crucial in hot environments and can be restricted in climates with high atmospheric moisture content (Davis et al. 2016).

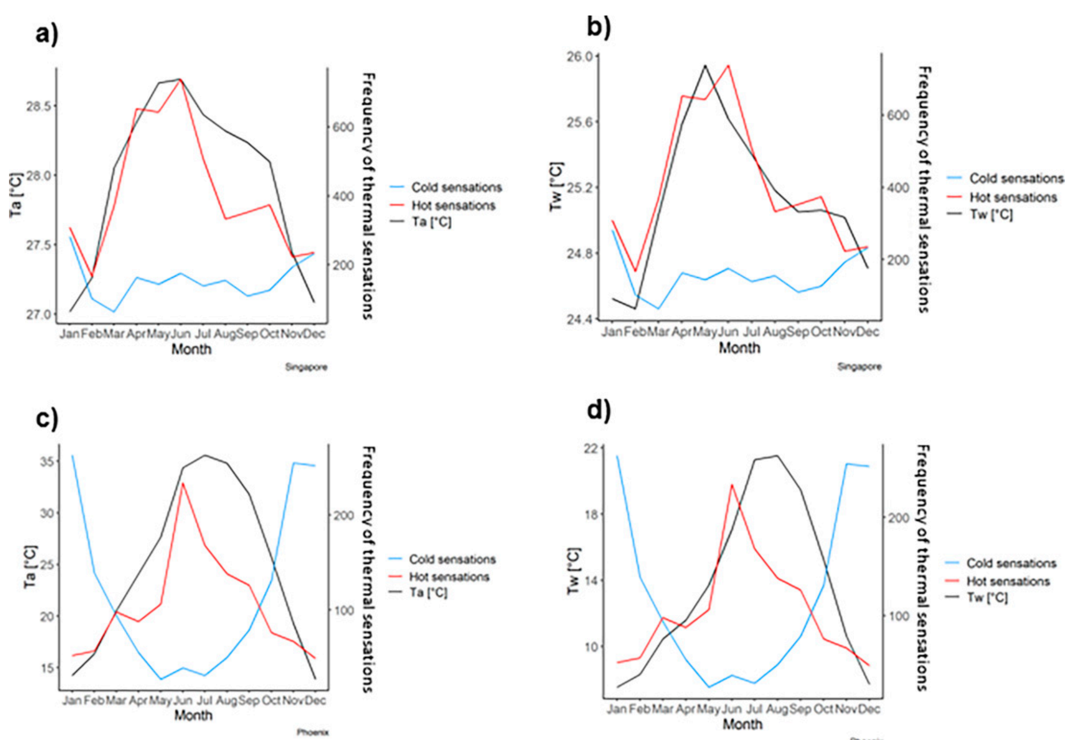

FIG. 4. Monthly time series for (left) Ta and (right) Tw and frequency of hot- and cold-expressing thermal sensations for (a),(b) Singapore and (c),(d) Phoenix between 2012 and 2019.

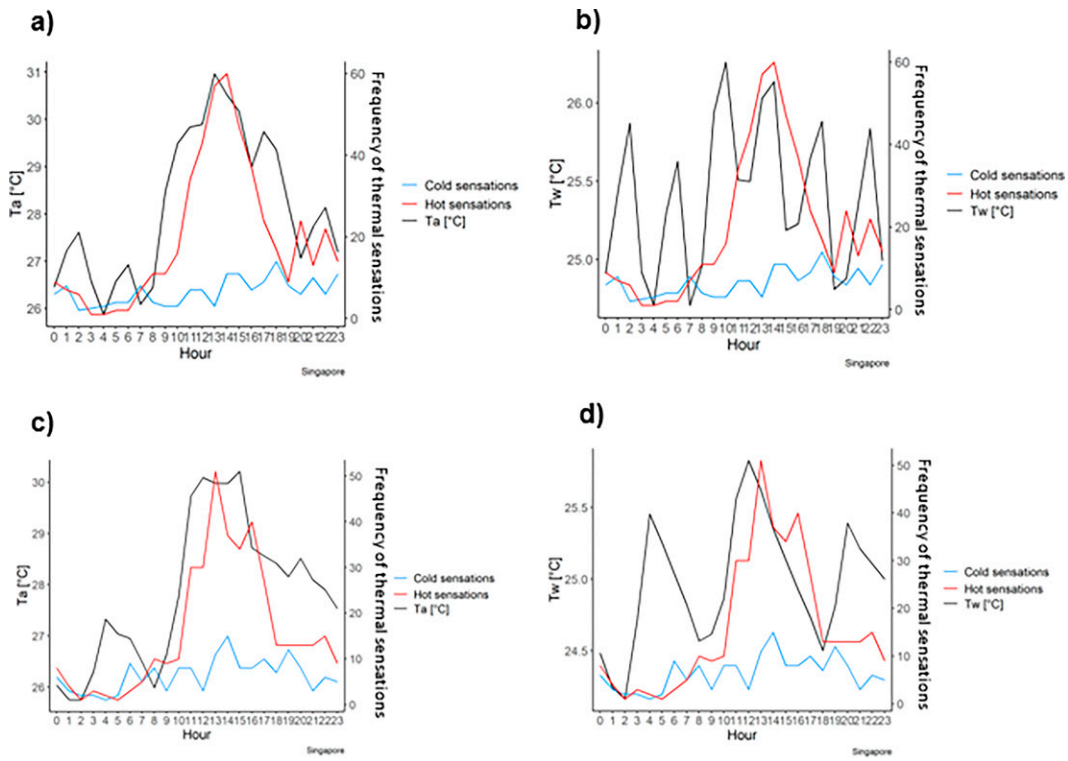

FIG. 5. Hourly time series for (left) Ta and (right) Tw and frequency of hot- and cold-expressing thermal sensations for Singapore (a),(b) intermonsoon (April–May and October–November) and (c),(d) monsoon season (December–March and June–September) between 2012 and 2019. Data are in local time (UTC + 8 h).

To ascertain lexical patterns of weather-related tweets across countries and seasons, string data were used to generate word clouds with occurrences above 100 times. Text stemming was also employed to identify and consolidate words with the same base word. For instance, "sweating" and "sweaty" would be grouped under the same stem with the meaning "sweat." Then, the text string data were organized according to seasonal time periods for each city. For Singapore, the periods were monsoon (December–March and June–September) and intermonsoon periods (April–May and October–November); for Phoenix, the periods were spring/fall season (March, April, October, and November), summer (May–September), and winter (December–February). Furthermore, word clouds were produced to understand the commonly used language during those seasonally important time periods. Analysis in this study was performed using RStudio, version 1.4.1106.

3. Results

Tweets including the word "weather" and location set to the respective city amounted to 48 186 tweets for Phoenix and 34 899 for Singapore. After removing duplicates and tweets from official weather channels, the remaining 22 408 tweets for Phoenix and 29 832 for Singapore were used for analysis. Singapore had fewer tweets between 2012 and 2013 than did Phoenix (Fig. 2), with a sharp increase in 2014, coinciding with an extreme weather (drought) event that year (McBride et al. 2015), and a downward trend since then. Phoenix had a relatively equal distribution of weather-related tweets throughout the period of interest with a slight increase between 2014 and 2016.

Among 20 most frequently used words for Phoenix (Fig. 3), "cold" and "hot" were used most often: 1181 and 883 times, respectively. For Singapore, "hot" was used 4179 times, followed by "cold" (n = 1579). Other words used above 300 times were "warm," "cool," "heat," and "sunny" for Phoenix and "humid," "cool," "warm," and "sunny" for Singapore.

a. *Relationship between weather and thermal sensations*

Aggregated monthly weather time series for both cities showed relationships with thermal sensations. The rise in hot sensations coincided with the increase in Ta and decreased in the following months even though the temperature remained high. Hot sensations for both cities peaked in June and were the lowest in winter months.

In Singapore (Figs. 4a,b), the slope of rising hot sensations closely followed the slopes for Ta and Tw. We observed two peaks in hot sensations, the first one occurring in April, at the onset of the rise in Ta and Tw, the second, higher peak, in hot sensations coincided with the peak in Ta while the peak in Tw occurred one month earlier, in May. Overall, Tw had a slightly stronger relationship with hot sensations than did Ta (Tw: $R^2 = 0.90$, $p < 0.01$; Ta: $R^2 = 0.85$, $p < 0.01$). Cold sensations, however, were not significantly related to Ta and Tw in Singapore.

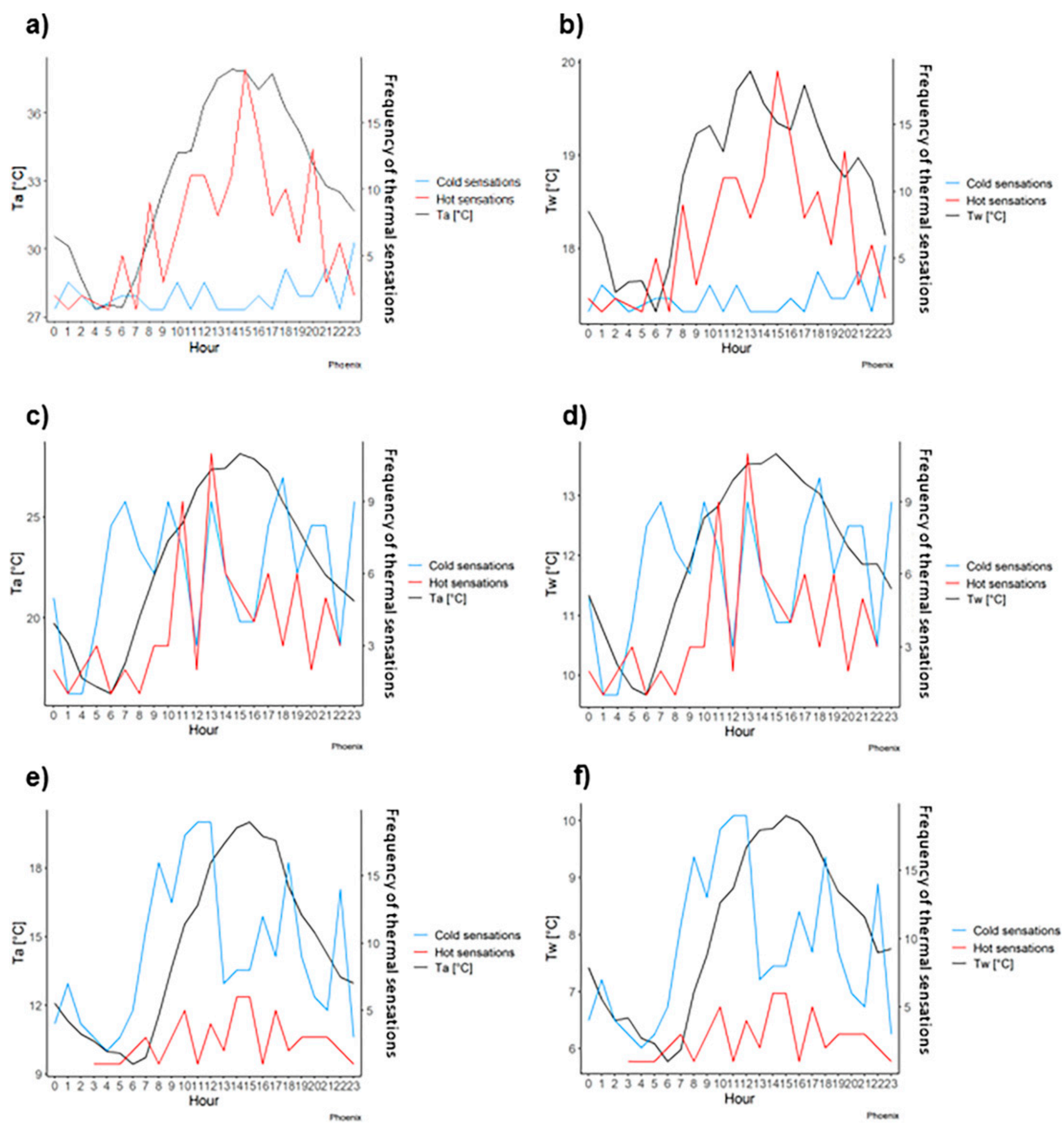

FIG. 6. Hourly time series for (left) Ta and (right) Tw and frequency of hot- and cold-expressing thermal sensations for Phoenix (a),(b) summer season (May–September); (c),(d) spring/fall season (March, April, October, and November); and (e),(f) winter season (December–February) between 2012 and 2019. Data are in local time (UTC − 7 h).

In Phoenix (Figs. 4c,d), both hot and cold sensations tweets were significantly correlated with Ta and Tw. Unlike Singapore, there was one pronounced peak in hot sensations in June and Ta had a slightly stronger relationship with both hot (Ta: $R^2 = 0.95, p < 0.01$; Tw: $R^2 = 0.87, p < 0.001$) and cold sensations (Ta: $R^2 = -0.85, p < 0.01$; Tw: $R^2 = -0.75, p < 0.01$). Peak in tweets referring to hot sensations preceded the peak in Ta and Tw by one and two months, respectively. Peaks in tweets referring to cold sensations coincided with lowest Ta and Tw.

Seasonal hourly time series showed that the peak in hot sensations occurred between 1400 and 1500 local time (LT) in Singapore, and generally followed a similar pattern to Ta for both intermonsoon ($R^2 = 0.86, p < 0.01$) and monsoon ($R^2 = 0.89, p < 0.01$) seasons, respectively (Fig. 5). Tw was not significantly correlated with hot sensations during intermonsoon but was weakly correlated with hot sensations during the monsoon season ($R^2 = 0.49, p = 0.018$). Cold sensations were not associated with hourly weather trends in Singapore.

In Phoenix, aggregated hourly series showed that hot sensations had a strong relationship with Ta ($R^2 = 0.79, p < 0.01$) and Tw ($R^2 = 0.70, p < 0.01$) in summer with peak occurrences of hot sensations at 1500 LT (Fig. 6a). Hourly Ta had a slightly stronger relationship with hot sensations during the spring/fall season than Tw (Ta: $R^2 = 0.69, p < 0.01$; Tw: $R^2 = 0.67, p < 0.01$), but the overall trend had a moderate strength, and we did not observe a distinctive peak for hot sensations (Figs. 6c,d). Both hourly hot and cold sensations were moderately positively correlated with Ta and Tw in winter (hot: Ta: $R^2 = 0.54,$

$p = 0.02$; Tw: $R^2 = 0.48$, $p < 0.046$; cold: Ta: $R^2 = 0.45$, $p = 0.03$; Tw: $R^2 = 0.42$, $p = 0.044$) (Figs. 6e,f).

b. *Lexical patterns by country and season*

Bigrams (Fig. 7) showed two-word sequences of words related to "weather" such as "hot weather" or "weather sucks." For bigrams with occurrences above 30 times, there was a dominant cluster connected to weather around "hot" for Singapore with dominant negative sensations. Interestingly, bigram "hot chocolate" was popular in Singapore, mainly indicating positive attitude and preference toward cooler weather, reflected in sentiments like, "what a perfect weather for a perfect hot chocolate while sitting outdoor"; "this weather calls for cuddles and hot chocolate." However, no such cluster emerged for Phoenix, instead there was a cluster related to temperature degrees, climate change, and Phoenix Sky Harbor airport.

To further understand participants' sentiments to weather, we generated word clouds (Fig. 8) for combined NE and SW monsoon and intermonsoon periods in Singapore (Figs. 8a,b), as well as summer, winter, and spring/fall periods in Phoenix (Figs. 8c–e). Word clouds for the two seasons in Singapore are similar, where "hot" and "sleep" are the most popular words in both. However, "rain," "nice," and "cold" are slightly more prominent during the cooler monsoon season. Word clouds for Phoenix have distinct lexical patterns, with "degree" and "hot" being the most frequent in summer, "beautiful" and "perfect" in spring/fall, and "cold," "perfect," and "beautiful" in winter.

4. Discussion

Previous studies have shown that social media data can elucidate relationships between weather and expressed sentiments (Baylis et al. 2018; Young et al. 2021). In this study, we found relationships between Ta and Tw and thermal comfort expressing sensations.

There were clear seasonal differences in frequency of thermal comfort expressing sensations and Ta for both cities. Even though mean seasonal changes in Ta were relatively small in Singapore (mean annual Ta ranges between 27° and 29°C), our analysis showed a strong relationship between Ta and prevalence of hot-related sensations, demonstrating that Twitter users in Singapore were potentially sensitive to small variations in weather. Seasonal trends were even more pronounced in Phoenix, where frequency of both hot and cold-related sensations was significantly correlated with mean Ta and Tw. The peak in hot sensations for Phoenix and the first peak for Singapore occurred at the onset of the Ta and Tw rise, signaling a higher sensitivity to heat at the start of the season that can be explained by seasonal alliesthesia, a feeling of (dis)-pleasure resulting from the change in thermal conditions (Cabanac 1971). Similarly, previous studies showed that the frequency of heat-related tweets was sensitive to not only an increase in temperatures but also to the difference between the temperatures (Murakami et al. 2016).

Hourly analysis showed that hot-related sensations closely followed Ta during monsoon and intermonsoon seasons in Singapore, and hot summer Phoenix months. These results could potentially indicate that Twitter users were sensitive to diurnal temperature differences and quick in expressing heat-related discomfort almost immediately. While we observed moderate relationships between Ta and hot sensations for spring/fall, and positive relationships for both hot and cold sensations for winter months in Phoenix, we think it might be generally related to the time of the day when Twitter users are more active coinciding with diurnal temperature changes. A small number of weather-related tweets in winter might have affected the robustness of the analysis. Relationships between hourly sentiments and Tw were not as pronounced for Singapore, possibly due to a high variability in hourly Tw, especially during the intermonsoon seasons characterized by afternoon thunderstorms from strong daytime heating and convection of moist surface air. Tw for Phoenix was much lower with the shape of the curve closely resembling Ta, indicating low atmospheric moisture content and high evaporative capacity.

Furthermore, linguistic analysis of most frequently used words in tweets presented in word clouds also showed little seasonal variation during monsoon and intermonsoon in Singapore. On the other hand, word clouds for Phoenix show more seasonal variation, indicating an overall positive sentiment toward cooler weather during spring/fall and winter months. This is also reflected in literature, for instance, Morrissey et al. (1996) study in northern Australia reveals that humidity and temperature are important factors influencing mood and behavior, and a study in the United States (Yang et al. 2015) demonstrates that climate and seasonality can impact mood and depression rates.

Although this study has shed light on the viability of Twitter as an indicator of public sentiment toward the weather, it is important to note that the results are specific to the two cities studied. Different countries and cultures have many variations in the usage and restrictions of social media, which will lead to varying results with the same method applied. Moreover, each city comes with its own unique weather and climate patterns, as well as populations with their own unique threshold of thermal (dis)comfort (Alhadad et al. 2019). Current study exhibits clear differences in weather-related Twitter usage between geographically, culturally, and politically distinct locations. First of all, while the number of yearly weather-related tweets in Phoenix was relatively consistent, we observed a spike in Singapore in 2014 followed by a downward trend since (Fig. 2). This spike in the data is not related to overall Twitter users in Singapore, which has a slight rising trend (Fig. 9a). However, January–March 2014 was hit by a record long dry spell (McBride et al. 2015), which is also evidenced by the lowest annual Tw (Fig. 9b). Thus, we believe this factor could have contributed to the rise in weather-related tweets, indicating sensitivity to regional climatic events.

Furthermore, bigram map showed one dominant cluster connected to weather around "hot" for Singapore with dominant negative sensations and use of profanities to express weather sentiment. However, no such cluster emerged for Phoenix. Instead, clusters associated with temperature degrees and climate change mentions emerged. This fact indicates potential differences in the use of Twitter between the two cities. In the United States, Twitter is widely used by weather channels and public

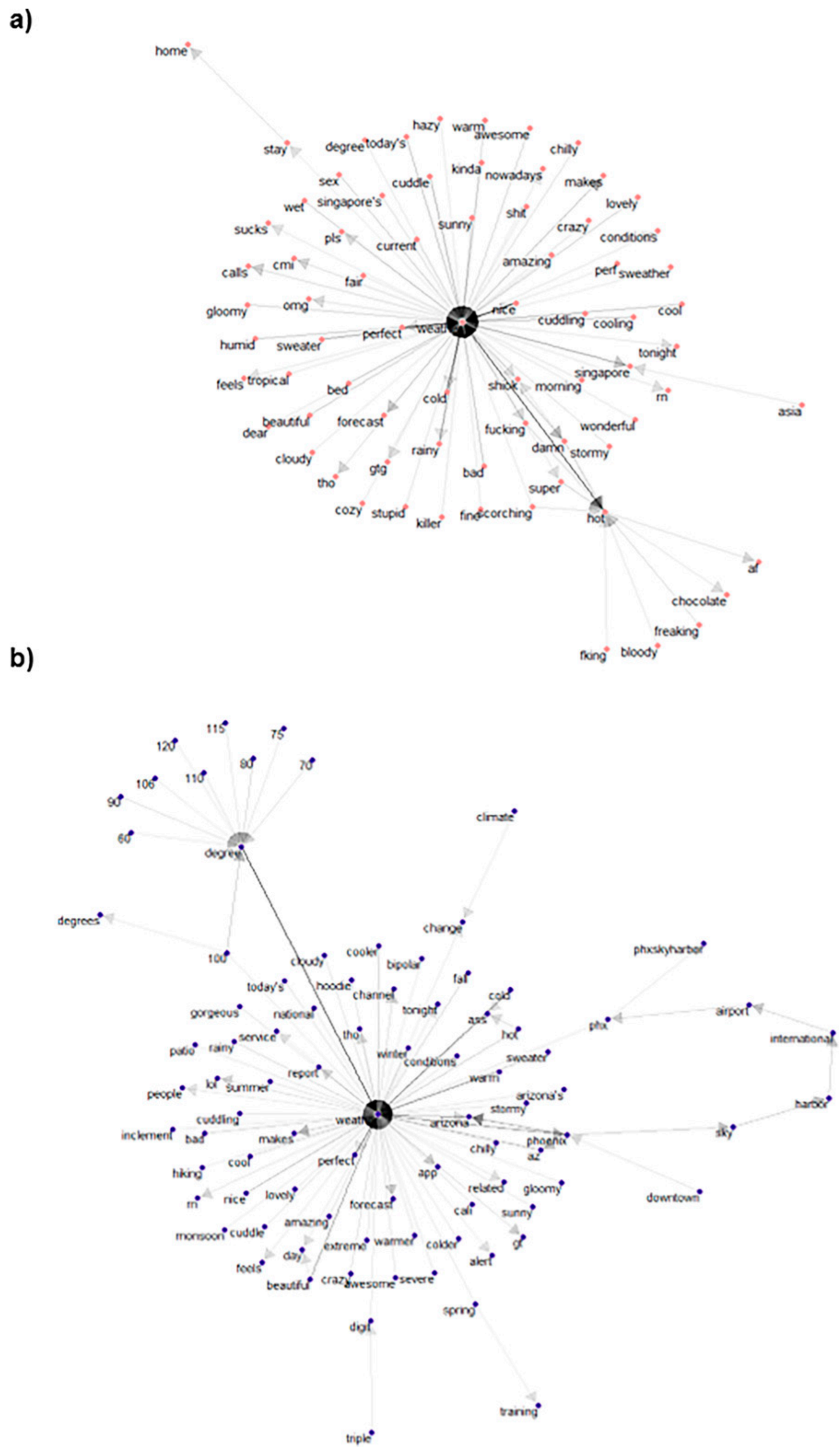

FIG. 7. Bigrams of weather-related tweets ($N > 30$) for (a) Singapore and (b) Phoenix between 2012 and 2019.

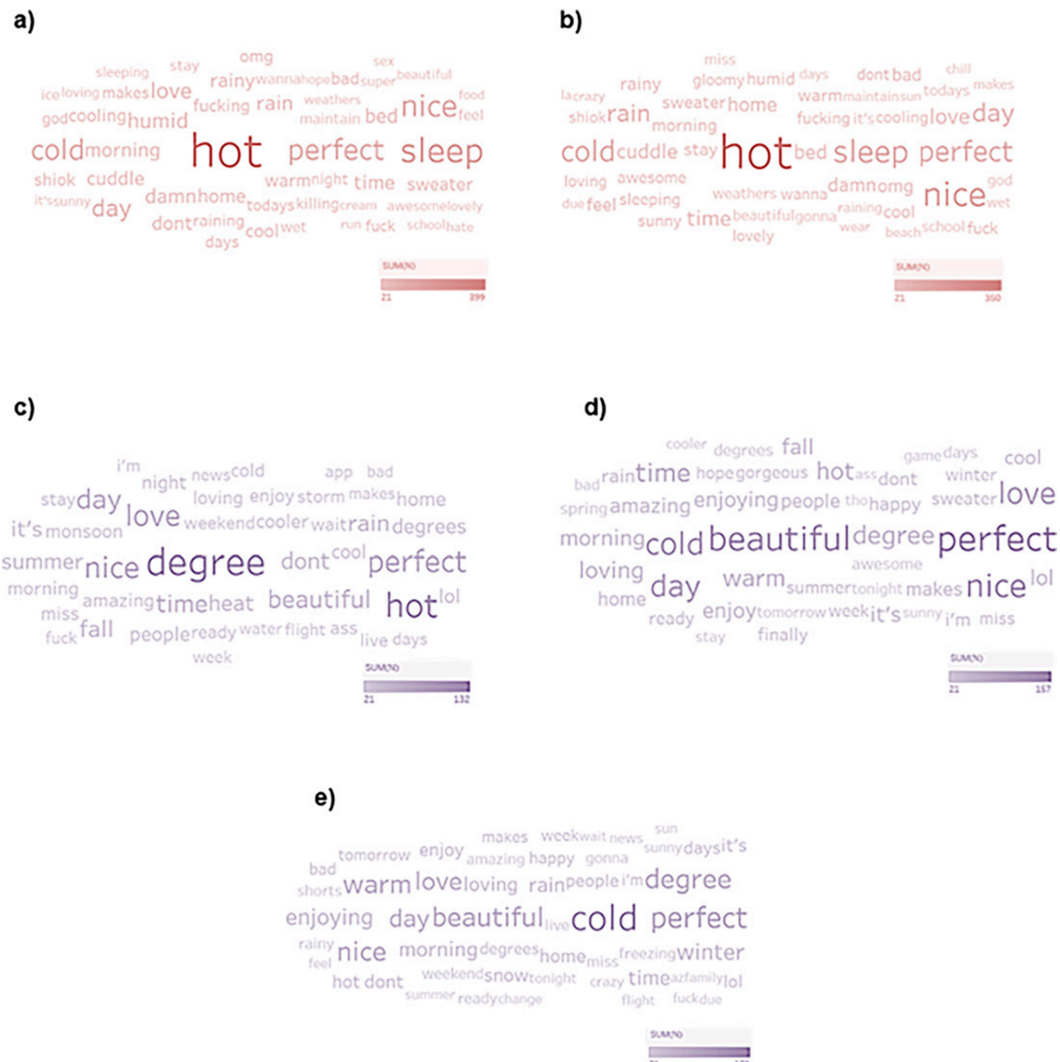

FIG. 8. Word clouds with $N < 20$ for (a) intermonsoon (April–May and October–November) and (b) monsoon (December–March and June–September) seasons in Singapore and (c) summer (May–September), (d) spring/fall (March, April, October, and November), and (e) winter (December–February) seasons in Phoenix between 2012 and 2019.

offices (Gaztelumendi et al. 2018) to send weather warning messages resulting in a higher awareness and mentions of specific temperatures referring to weather advisories among the public. However, the use of Twitter for advisories in Singapore is limited.

5. Limitations

Although Twitter data provides a large source of information to determine the perceptions of individuals, the representativeness of Twitter samples remains as a limitation given the difference in user demographics between cities. For instance, Mislove et al. (2021) state that the general Twitter user in the United States is likely to be Caucasian, male, and located in major cities. Differences in the use of social media platforms should be accounted for cross-city comparisons between different countries and cultural contexts. Future studies need to take into account the study area's unique climate, demographics, language, thermal thresholds, and preferred social media platform.

Furthermore, countries with multiple commonly spoken languages present challenges for text analysis and comparison. In our study, both Phoenix and Singapore have more than one commonly used language. Singapore has four official languages—English, Malay, Tamil, and Mandarin. English is the most widely spoken vernacular across all ethnic groups (Singstat 2020). Singlish, a colloquial language of English mixed with Chinese dialects and Malay, is also used in Singapore. However, there are limited resources to analyze this language variant, and English is still critical to understand the sentiment alongside other languages (Lo et al. 2017). In Phoenix, approximately 70% of the

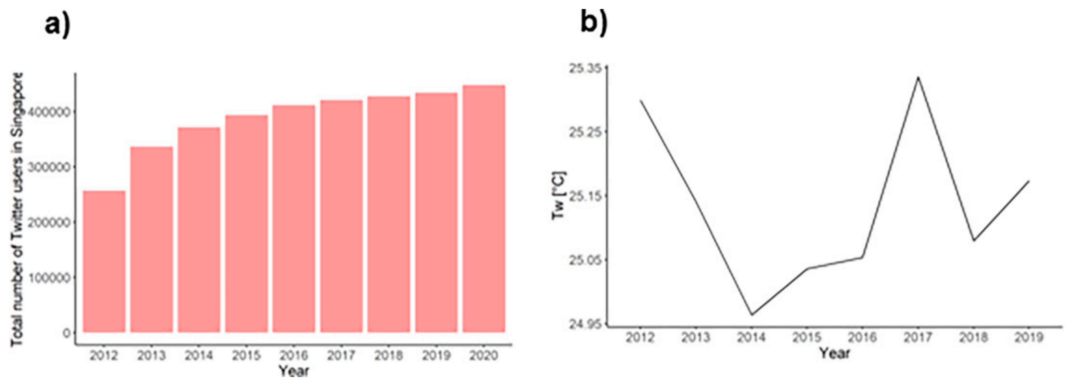

FIG. 9. (a) Frequency of Twitter users and (b) mean annual Tw for Singapore between 2012 and 2019.

population speak mainly English in their household, followed by Spanish (U.S. Census Bureau 2019). Thus, for comparability between both cities, English has been selected for this analysis.

6. Conclusions

We demonstrated strong relationships between both diurnal and seasonal changes in temperature and number of thermal comfort-related sensations. In Singapore, hot sensations were influenced by Ta and Tw, even though monthly temperature changes were small. In Phoenix, both hot and cold sensations were sensitive to weather changes. Singapore heat expressing sensations were also strongly influenced by diurnal changes in Ta throughout the year, while for Phoenix, there was a strong relationship during summer. We also showed the indication of seasonal alliesthesia in both cities since hot-related tweets peaked at the beginning of the seasonal rise in Ta and Tw and went down in the subsequent months.

The most used weather descriptors for Singapore and Phoenix were "hot" and "cold." Discomfort related to high atmospheric moisture in Singapore was evident through such descriptors as "humid," "sweating," and "melting." Word clouds showed no clear lexical seasonal pattern in Singapore. In Phoenix, on the other hand, we observed clear seasonal trends with more mentions of "degrees" and "hot" in summer versus more positive sentiments during spring/fall and winter, with such words as "perfect," "beautiful," and "cold."

Bigrams showed a negative cluster related to hot weather in Singapore. In Phoenix, we observed clusters involving the mentions of temperature degrees, which are likely associated to extreme weather advisories, as well as mention of climate change. This brings us to the point that, while in both cities tweets have strong correlations with weather, it has to be interpreted according to regional contexts. For instance, in this case, Twitter demonstrates the public's dissatisfaction with weather in Singapore, while at the same time it also reflects how local government weather advisories influence public weather perception in Phoenix. Furthermore, we hypothesized that weather-related tweets in Singapore were influenced by

TABLE A1. List of 100 weather descriptors used for initial screening of Twitter data.

1	Airy	21	Cold	41	Glowing	61	Scorching	81	Sunny
2	Alpine	22	Cool	42	Gusty	62	Searing	82	Sunshine
3	Arctic	23	Cosy	43	Heat	63	Shivering	83	Sweat
4	Arid	24	Cozy	44	Heated	64	Siberian	84	Sweating
5	Balmy	25	Crisp	45	Hot	65	Sizzling	85	Sweaty
6	Barren	26	Dank	46	Humid	66	Smelting	86	Swelter
7	Billowing	27	Dehydrated	47	Humidity	67	Snowy	87	Sweltering
8	Blaze	28	Desiccated	48	Icy	68	Snug	88	Temperate
9	Blazing	29	Draft	49	Lukewarm	69	Soaking	89	Tepid
10	Blistering	30	Drafty	50	Melting	70	Soupy	90	Toast
11	Blustery	31	Dry	51	Mild	71	Steam	91	Toasty
12	Boiling	32	Dusty	52	Mistral	72	Steaming	92	Torrid
13	Breeze	33	Feverish	53	Moist	73	Steamy	93	Tropical
14	Breezy	34	Fiery	54	Muggy	74	Sticky	94	Turbulent
15	Brisk	35	Freezing	55	Parched	75	Stormy	95	Unbearable
16	Burn	36	Fresh	56	Parching	76	Sultry	96	Unbearably
17	Burning	37	Frigid	57	Polar	77	Summery	97	Warm
18	Chill	38	Frosty	58	Radiant	78	Sunburn	98	Windswept
19	Chilled	39	Frozen	59	Roasting	79	Sunlight	99	Windy
20	Clammy	40	Glacial	60	Scalding	80	Sunlit	100	Wintry

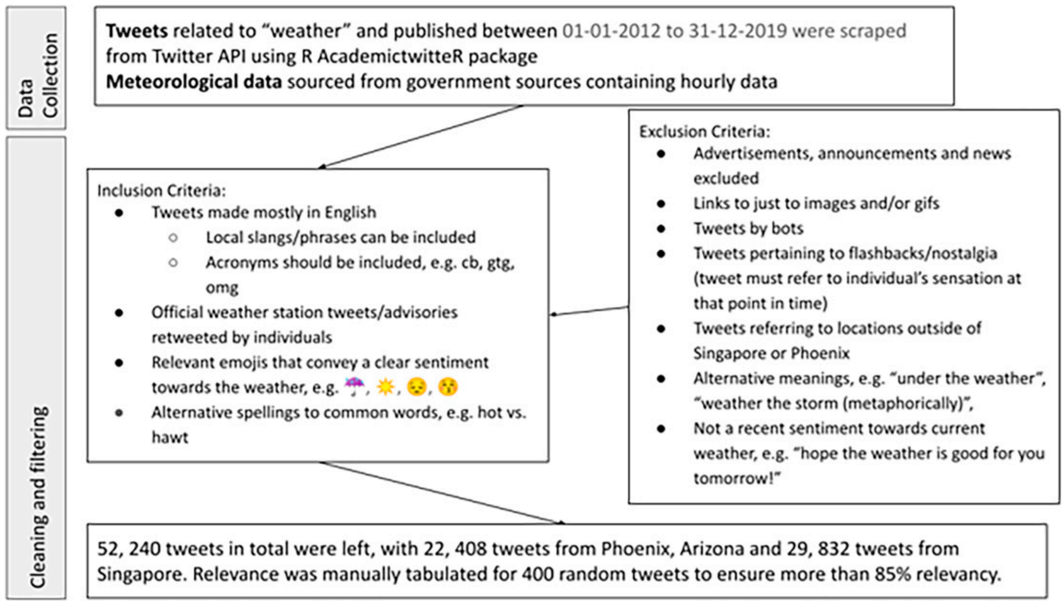

FIG. B1. Description of data-cleaning and relevance-assessment process.

the 2014 drought, which is reflected in the spike of tweets, indicating sensitivity to regional climate events. Also, lower overall mean sentiment score in Singapore is possibly due to the higher annual mean Ta and little seasonal variability.

Acknowledgments. The research was conducted under the "Cooling Singapore 2.0" project, led by the Singapore–ETH Centre (SEC) with the Singapore Management University (SMU), the Singapore–MIT Alliance for Research and Technology (SMART), TUMCREATE (established by the Technical University of Munich), the National University of Singapore (NUS), and the Cambridge Centre for Advanced Research and Education in Singapore (CARES). This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its "Cooling Singapore 2.0 Funding Initiative."

Data availability statement. Because of privacy and ethical concerns, supporting data cannot be made available.

APPENDIX A

Weather Descriptors

This appendix presents Table A1, which contains the list of weather descriptors used for the initial screening of Twitter data. It contains 76 climatic adjectives that were adopted from Liu et al. (2020), along with an additional 24 weather descriptors added by the authors.

TABLE B1. Relevance results.

	Raw data	Cleaned data
Singapore	78.5%	95%
Phoenix, Arizona	42.45%	86.75%

APPENDIX B

Information about Relevance

This appendix provides an overview of the process of collecting, cleaning, and filtering data (Fig. B1) to assure their relevance toward the subsequent analysis. It also presents the relevance results for Singapore and Phoenix for raw and cleaned data (Table B1).

REFERENCES

Alhadad, S. B., P. M. S. Tan, and J. K. W. Lee, 2019: Efficacy of heat mitigation strategies on core temperature and endurance exercise: A meta-analysis. *Front. Physiol.*, **10**, 71, https://doi.org/10.3389/fphys.2019.00071.

Aviv, A., G. Bromberg, Y. Baruch, Y. Shapira, and D. M. Blass, 2011: The role of environmental influences on schizophrenia admissions in Israel. *Int. J. Soc. Psychiatry*, **57**, 57–68, https://doi.org/10.1177/0020764009348444.

Baylis, P., N. Obradovich, Y. Kryvasheyeu, H. Chen, L. Coviello, E. Moro, M. Cebrian, and J. H. Fowler, 2018: Weather impacts expressed sentiment. *PLOS ONE*, **13**, e0195750, https://doi.org/10.1371/journal.pone.0195750.

Brazel, A., P. Gober, S. J. Lee, S. Grossman-Clarke, J. Zehnder, B. Hedquist, and E. Comparri, 2007: Determinants of changes in the regional urban heat island in metropolitan Phoenix (Arizona, USA) between 1990 and 2004. *Climate Res.*, **33**, 171–182, https://doi.org/10.3354/cr033171.

Bujisic, M., V. Bogicevic, H. G. Parsa, V. Jovanovic, and A. Sukhu, 2019: It's raining complaints! How weather factors drive consumer comments and word-of-mouth. *J. Hosp. Tour. Res.*, **43**, 656–681, https://doi.org/10.1177/1096348019835600.

Cabanac, M., 1971: Physiological role of pleasure. *Science*, **173**, 1103–1107, https://doi.org/10.1126/science.173.4002.1103.

Chow, W. T. L., D. Brennan, and A. J. Brazel, 2012: Urban heat island research in Phoenix, Arizona: Theoretical contributions and policy applications. *Bull. Amer. Meteor. Soc.*, **93**, 517–530, https://doi.org/10.1175/BAMS-D-11-00011.1.

Cody, E. M., A. J. Reagan, L. Mitchell, P. S. Dodds, and C. M. Danforth, 2015: Climate change sentiment on Twitter: An unsolicited public opinion poll. *PLOS ONE*, **10**, e0136092, https://doi.org/10.1371/journal.pone.0136092.

Connors, J. P., C. S. Galletti, and W. T. L. Chow, 2013: Landscape configuration and urban heat island effects: Assessing the relationship between landscape characteristics and land surface temperature in Phoenix, Arizona. *Landscape Ecol.*, **28**, 271–283, https://doi.org/10.1007/s10980-012-9833-1.

Dahal, B., S. A. P. Kumar, and Z. Li, 2019: Topic modeling and sentiment analysis of global climate change tweets. *Soc. Network Anal. Min.*, **9**, 24, https://doi.org/10.1007/s13278-019-0568-8.

Davis, R. E., G. R. McGregor, and K. B. Enfield, 2016: Humidity: A review and primer on atmospheric moisture and human health. *Environ. Res.*, **144**, 106–116, https://doi.org/10.1016/j.envres.2015.10.014.

Demuzere, M., S. Hankey, G. Mills, W. Zhang, T. Lu, and B. Bechtel, 2020: Combining expert and crowd-sourced training data to map urban form and functions for the continental US. *Sci. Data*, **7**, 264, https://doi.org/10.1038/s41597-020-00605-z.

Ebi, L. K., and Coauthors, 2021: Hot weather and heat extremes: Health risks. *Lancet*, **398**, 698–708, https://doi.org/10.1016/S0140-6736(21)01208-3.

EPA, 2021: Climate change indicators: Heat-related deaths. U.S. Environmental Protection Agency, accessed 15 October 2021, https://www.epa.gov/climate-indicators/climate-change-indicators-heat-related-deaths.

Fownes, J. R., C. Yu, and D. B. Margolin, 2018: Twitter and climate change. *Sociol. Compass*, **12**, e12587, https://doi.org/10.1111/soc4.12587.

Gaztelumendi, S., M. Martija, and O. Principe, 2018: Twitter and weather services. *Adv. Sci. Res.*, **15**, 239–243, https://doi.org/10.5194/asr-15-239-2018.

Giuffrida, L., H. Lokys, and O. Klemm, 2020: Assessing the effect of weather on human outdoor perception using Twitter. *Int. J. Biometeor.*, **64**, 205–216, https://doi.org/10.1007/s00484-018-1574-7.

Hansen, A., P. Bi, M. Nitschke, P. Ryan, D. Pisaniello, and G. Tucker, 2008: The effect of heat waves on mental health in a temperate Australian city. *Environ. Health Perspect.*, **116**, 1369–1375, https://doi.org/10.1289/ehp.11339.

Kim, S.-H., S.-N. Jo, H.-N. Myung, and J.-Y. Jang, 2014: The effect of pre-existing medical conditions on heat stroke during hot weather in South Korea. *Environ. Res.*, **133**, 246–252, https://doi.org/10.1016/j.envres.2014.06.003.

Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel, 2006: World map of the Köppen-Geiger climate classification updated. *Meteor. Z.*, **15**, 259–263, https://doi.org/10.1127/0941-2948/2006/0130.

Kuo, F. E., and W. C. Sullivan, 2001: Aggression and violence in the inner city: Effects of environment via mental fatigue. *Environ. Behav.*, **33**, 543–571, https://doi.org/10.1177/00139160121973124.

Liu, S., N. Nazarian, J. Niu, M. A. Hart, and R. de Dear, 2020: From thermal sensation to thermal affect: A multi-dimensional semantic space to assess outdoor thermal comfort. *Build. Environ.*, **182**, 107112, https://doi.org/10.1016/j.buildenv.2020.107112.

Lo, S. L., E. Cambria, R. Chiong, and D. Cornforth, 2017: Multilingual sentiment analysis: From formal to informal and scarce resource languages. *Artif. Intell. Rev.*, **48**, 499–527, https://doi.org/10.1007/s10462-016-9508-4.

March, D., S. L. Hatch, C. Morgan, J. B. Kirkbride, M. Bresnahan, P. Fearon, and E. Susser, 2008: Psychosis and place. *Epidemiol. Rev.*, **30**, 84–100, https://doi.org/10.1093/epirev/mxn006.

Mcbride, J. L., S. Sahany, M. E. E. Hassim, C. M. Nguyen, S.-Y. Lim, R. Rahmat, and W.-K. Cheong 2015: The 2014 record dry spell at Singapore: An intertropical convergence zone (ITCZ) drought. *Bull. Amer. Meteor. Soc.*, **96** (12), S126–S130, https://doi.org/10.1175/BAMS-D-15-00117.1.

McGregor, G. R., and J. K. Vanos, 2018: Heat: A primer for public health researchers. *Public Health*, **161**, 138–146, https://doi.org/10.1016/j.puhe.2017.11.005.

Meteorological Service Singapore, 2021: Climate of Singapore. National Environmental Agency, accessed 6 October 2021, http://www.weather.gov.sg/climate-climate-of-singapore/.

Mislove, A., S. Lehmann, Y. Y. Ahn, J. P. Onnela, and J. Rosenquist, 2021: Understanding the demographics of Twitter users. *Fifth Int. AAAI Conf. on Weblogs and Social Media*, Barcelona, Spain, AAAI, 554–557, https://ojs.aaai.org/index.php/ICWSM/article/view/14168.

Morrissey, S. A., P. T. F. Raggatt, B. James, and J. Rogers, 1996: Seasonal affective disorder: Some epidemiological findings from a tropical climate. *Aust. N. Z. J. Psychiatry*, **30**, 579–586, https://doi.org/10.3109/00048679609062653.

Mughal, M. O., X.-X. Li, T. Yin, A. Martilli, O. Brousse, M. A. Dissenga, and L. K. Norford, 2019: High-resolution, multi-layer modelling of Singapore's urban climate incorporating local climate zones. *J. Geophys. Res. Atmos.*, **124**, 7764–7785, https://doi.org/10.1029/2018JD029796.

Murakami, D., G. W. Peters, Y. Yamagata, and T. Matsui, 2016: Participatory sensing data tweets for micro-urban real-time resiliency monitoring and risk management. *IEEE Access*, **4**, 347–372, https://doi.org/10.1109/ACCESS.2016.2516918.

Roberts, H. V., 2017: Using Twitter data in urban green space research: A case study and critical evaluation. *Appl. Geogr.*, **81**, 13–20, https://doi.org/10.1016/j.apgeog.2017.02.008.

Roth, M., and W. T. L. Chow, 2012: A historical review and assessment of urban heat island research in Singapore. *Singapore J. Trop. Geogr.*, **33**, 381–397, https://doi.org/10.1111/sjtg.12003.

Santamouris, M., and D. Kolokotsa, 2015: On the impact of urban overheating and extreme climatic conditions on housing, energy, comfort and environmental quality of vulnerable population in Europe. *Energy Build.*, **98**, 125–133, https://doi.org/10.1016/j.enbuild.2014.08.050.

Singstat, 2020: Census of Population 2020—Statistical Release 1: Demographic characteristics, education, language and religion. Department of Statistics Singapore Doc., 247 pp., https://www.singstat.gov.sg/-/media/files/publications/cop2020/sr1/cop2020sr1.pdf.

Statista, 2021a: Most popular social networks worldwide as of July 2021, ranked by number of active users (in millions). We Are Social, Hootsuite, and DataReportal, accessed 18 October 2021, https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users.

——, 2021b: Twitter. Accessed 18 October 2021, https://www.statista.com/study/9920/twitter-statista-dossier.

——, 2021c: Social media in Asia Pacific. Accessed 18 October 2021, https://www.statista.com/study/75172/social-media-in-asia-pacific.

Sung, T.-I., M.-J. Chen, C.-Y. Lin, S., C. Lung, and H., J. Su, 2011: Relationship between mean daily ambient temperature range and hospital admissions for schizophrenia: Results from a national cohort of psychiatric inpatients. *Sci. Total Environ.*, **410–411**, 41–46, https://doi.org/10.1016/j.scitotenv.2011.09.028.

——, ——, and H.-J. Su, 2013: A positive relationship between ambient temperature and bipolar disorder identified using a national cohort of psychiatric inpatients. *Soc. Psychiatry Psychiatr. Epidemiol.*, **48**, 295–302, https://doi.org/10.1007/s00127-012-0542-5.

Trang, P. M., J. Rocklöv, K. B. Giang, G. Kullgren, and M. Nilsson, 2016: Heatwaves and hospital admissions for mental disorders in northern Vietnam. *PLOS ONE*, **11**, e0155609, https://doi.org/10.1371/journal.pone.0155609.

U.S. Census Bureau, 2019: Language spoken at home. American Community Survey, accessed 15 October 2021, https://data.census.gov/cedsci/table?q=arizona language&tid=ACSST1Y2019.S1601.

U.S. Climate Data, 2022: Climate Phoenix—Arizona. U.S. Climate Data, accessed 15 October 2021, https://www.usclimatedata.com/climate/phoenix/arizona/united-states/usaz0166.

Vaidyanathan, A., J. Malilay, P. Schramm, and S. Saha, 2020: Heat-related deaths—United States, 2004–2018. *Morb. Mortal. Wkly. Rep.*, **69**, 729–734, https://doi.org/10.15585/mmwr.mm6924a1.

Vassos, E., E. Agerbo, O. Mors, and C. B. Pedersen, 2016: Urban-rural differences in incidence rates of psychiatric disorders in Denmark. *Br. J. Psychiatry*, **208**, 435–440, https://doi.org/10.1192/bjp.bp.114.161091.

Veltri, G. A., and D. Atanasova, 2015: Climate change on Twitter: Content, media ecology and information sharing behaviour. *Public Understanding Sci.*, **26**, 721–737, https://doi.org/10.1177/0963662515613702.

Volpe, F. M., and J. A. Del Porto, 2006: Seasonality of admissions for mania in a psychiatric hospital of Belo Horizonte, Brazil. *J. Affective Disord.*, **94**, 243–248, https://doi.org/10.1016/j.jad.2006.03.025.

Wang, X., E. Lavigne, H. Ouellette-Kuntz, and B. E. Chen, 2014: Acute impacts of extreme temperature exposure on emergency room admissions related to mental and behavior disorders in Toronto, Canada. *J. Affective Disord.*, **155**, 154–161, https://doi.org/10.1016/j.jad.2013.10.042.

Wouters, H., and Coauthors, 2017: Heat stress increase under climate change twice as large in cities as in rural areas: A study for a densely populated midlatitude maritime region. *Geophys. Res. Lett.*, **44**, 8997–9007, https://doi.org/10.1002/2017GL074889.

Yang, W., L. Mu, and Y. Shen, 2015: Effect of climate and seasonality on depressed mood among Twitter users. *Appl. Geogr.*, **63**, 184–191, https://doi.org/10.1016/j.apgeog.2015.06.017.

Young, J. C., R. Arthur, M. Spruce, and H. T. P. Williams, 2021: Social sensing of heatwaves. *Sensors*, **21**, 3717, https://doi.org/10.3390/s21113717.