ARTICLE

Spatiotemporal patterns of US drought awareness

Sungyoon Kim[1], Wanyun Shao [2,3] & Jonghun Kam [1,3]*

ABSTRACT   Drought is a creeping climatological phenomenon with persistent precipitation deficits. Unlike rapid onset natural hazards such as floods and wildfires, the intangible and gradual characteristics of drought cause a lack of social response during the onset. The level of awareness of a local drought increases rapidly through mass media reports and online information searching activities when the drought reaches its peak severity. This high level of local drought awareness drives concerns for water shortage and support for water policy. However, spatiotemporal patterns of national-scale drought awareness have never been studied due to constraints imposed by time-consuming and costly survey data collection and surveys' limited sample sizes. Here, we present the national-scale study to reveal the spatiotemporal patterns of drought awareness over the contiguous United States (CONUS) using Google Trends data and Principal Component Analysis (PCA). Results show that the first two PC modes can explain 48% (38% for PC1 and 10% for PC2) of the total variance of state-level drought awareness. We find that the PC1 mode relates to a national pattern of drought awareness across the CONUS. The spatiotemporal patterns further imply that residents in the Northeastern US region are the most aware of the emergence of drought, regardless of the geographic location of the occurrence. The results illustrate how big data, such as search query and social media data, can help develop an effective and efficient plan for drought mitigation in the future.

[1] Department of Civil, Construction, and Environmental EngineeringThe University of Alabama, Tuscaloosa, AL 35487, USA. [2] Department of Geography, The University of Alabama, Tuscaloosa, AL 35487, USA. [3] Center for Complex Hydrosystems Research, Tuscaloosa, AL 35487, USA. *email: jkam@eng.ua.edu

Introduction

Drought is the least understood natural hazard due to the complexity of the generating mechanisms. Previous studies have been heavily focused on identifying potential sources of predictability (Schubert et al., 2008, Hoerling et al., 2013, Seager et al., 2015) and evaluating the current prediction skill of existing climate models (Seager et al., 2009, Yuan et al., 2011). In addition, detection and attribution of US regional droughts have been assessed (Seager et al., 2014, Cook et al., 2009, 2015). Despite the advanced understanding of the generating mechanisms and predictability, recent US droughts caused multi-billion-dollar economic losses (Smith and Matthews, 2015). This raises a concern about a lack of the effectiveness and efficiency of the current US drought monitoring and forecasting systems in generating timely social responses (Hayes et al., 2004). A better understanding of the key drivers and mechanisms of social response during the occurrence of drought crisis is warranted.

Past research on attitudes towards drought has been predominantly based on traditional survey instruments (Switzer and Vedlitz, 2017, Carlton et al., 2016, Stoutenborough and Vedlitz, 2014). While surveys often shed light on individuals' stated preferences, deeper analyses are required to understand behavioral adjustments to mitigate droughts over time (Lazrus, 2016). Further, conventional surveys focus on the variations among individuals while public policies are designed and implemented at an aggregate level. This mismatch on geographic scales naturally creates a gap between research results obtained from surveys and actionable policy items (Shao et al., 2018). Due to the time consuming and costly procedures, traditional surveys are conducted only at discrete points of time with a coarse temporal scale. This static nature in survey data limits our understanding of the dynamics of drought awareness (Nguyen et al., 2015).

In recent years, search engine query data and social media data have become popular alternative data sources to advance our limited understanding of the potential triggers and dynamics of social responses to natural disasters (Wang and Zhuang, 2017, Kam et al., 2019). These social monitoring data are classified into two main types: active (e.g., Twitter) and passive (e.g., Google Trends). While active social monitoring systems are often used as a tool of risk communication during the disaster emergence (Wang and Zhuang, 2017), they can be subject to misinformation (Wang and Zhuang, 2018). Whereas passive social monitoring systems provide observational data that can more accurately depict the growth and decay patterns of awareness of free agents during disaster emergency (Kam et al., 2019).

Google Trends provides passive social monitoring data based on online search actions that are non-communicative activities and presumably only motived by the user's interest. These search activities can therefore reveal the users' true interests and attention more accurately. Google Trends generates their data from search activities of over two billion users of Google Product around the world (Popper, 2017). These abundant Google Trends data can certainly provide an exhaustive overview of dynamical aggregate behaviors based on individual search activities. They have been used in previous studies to predict stock prices (Preis et al., 2013), to monitor disease outbreaks (Carneiro and Mylonakis, 2009), and to understand public perception of climate change (Anderegg and Goldsmith, 2014). In recent years, Google Trends data have been used to validate simulated drought awareness (Gonzales and Ajami, 2017) and to investigate the potential triggers and dynamic patterns of drought awareness during the 2011–2017 California drought (Kam et al., 2019).

Recently, the limitations of Google Trends data have been discussed (Nuti et al., 2014). One limitation is related to noise search activities that are defined as search activities using search terms including a common word (herein, drought) but not relevant to actual drought events. For example, Beyoncé released "Love Drought" on 24 April 2016 and it drove additional public search activities using the common term, "drought", during the on-going California drought (Kam et al., 2019). However, actual severe drought conditions increase the public interest or awareness about drought dramatically and consequently reduces the relative importance of these noise search activities irrelevant to the 2011–2017 California drought (Kam et al., 2019).

Drought risk can be quantified as the spatial extent of an area under a certain level of drought severity (Sheffield et al., 2009b). With the spatial extent reaching its maximum, the broad socioeconomic impact of drought becomes tangible, which leads to surges in awareness of and interest in the drought among a large group of people. This rapidly increased public awareness of drought drives search activities for drought information. Recent studies have used Google Trends data to explore the propagation of local drought awareness and interactions with other natural and human factors. For instance, Gonzales and Ajami (2017) investigated how water use patterns changed along with changes in public awareness of drought at the city level during the 2011–2017 California drought. Other studies investigated potential key drivers of the peaks of public awareness during the 2011–2017 California drought using Google Trends (Quesnel and Ajami, 2017, Kam et al., 2019). Different from all these previous studies in terms of scale, data source, and method, we conduct the first national-scale study to reveal the spatiotemporal patterns of drought awareness and their associations with drought risk using a passive social monitoring dataset (Google Trends) and a popular data compression technique (Principal Component Analysis).

Data and methods

Google trends and US drought monitoring data. Google Trends provides the information of public awareness about the specific search interest with a range of spatiotemporal scales through Google Search (http://trends.google.com). Google Trends data are available from 2004. In this study, state-level drought awareness (DA_i; i is an indicator of a US state (i = AL, …, WY)) is defined as the search activity using the specific term, "Drought" within the state of interest. The monthly relative search activity values of the 49 states of the contiguous US (CONUS) were retrieved on November 19th, 2018. Google Trends shows an index of relative search activity during a specific period that the user can determine since 2004. Depending on the chosen search period, the interval of relative search activities is automatically determined (from daily through monthly time steps). This relative search activity index is computed by dividing the total numbers of research activities within an interval by the maximum number of research activities during the chosen period.

The US Drought Monitor (USDM) provides the drought information by classifying the following four major drought intensities: abnormal dry (D0), moderate drought (D1), severe drought (D2), extreme drought (D3), and exceptional drought (D4). In this study, drought risk is defined by the total areal extent under abnormally dry condition (D0) and four types of drought severity (moderate (D1), severe (D2), extreme (D3), and exceptional (D4)). This study focuses on state-level drought risk (DR_i) over the 49 states of the CONUS during 2004–2017 to correspond to the temporal coverage of Google Trends. At the USDM website (https://droughtmonitor.unl.edu/), weekly percentages of the area under drought per state from 2000 are available. In this study, monthly state-level drought risk (DR_i) is defined as the percentage of the areal extents under drought condition reported in the middle week of each month. The
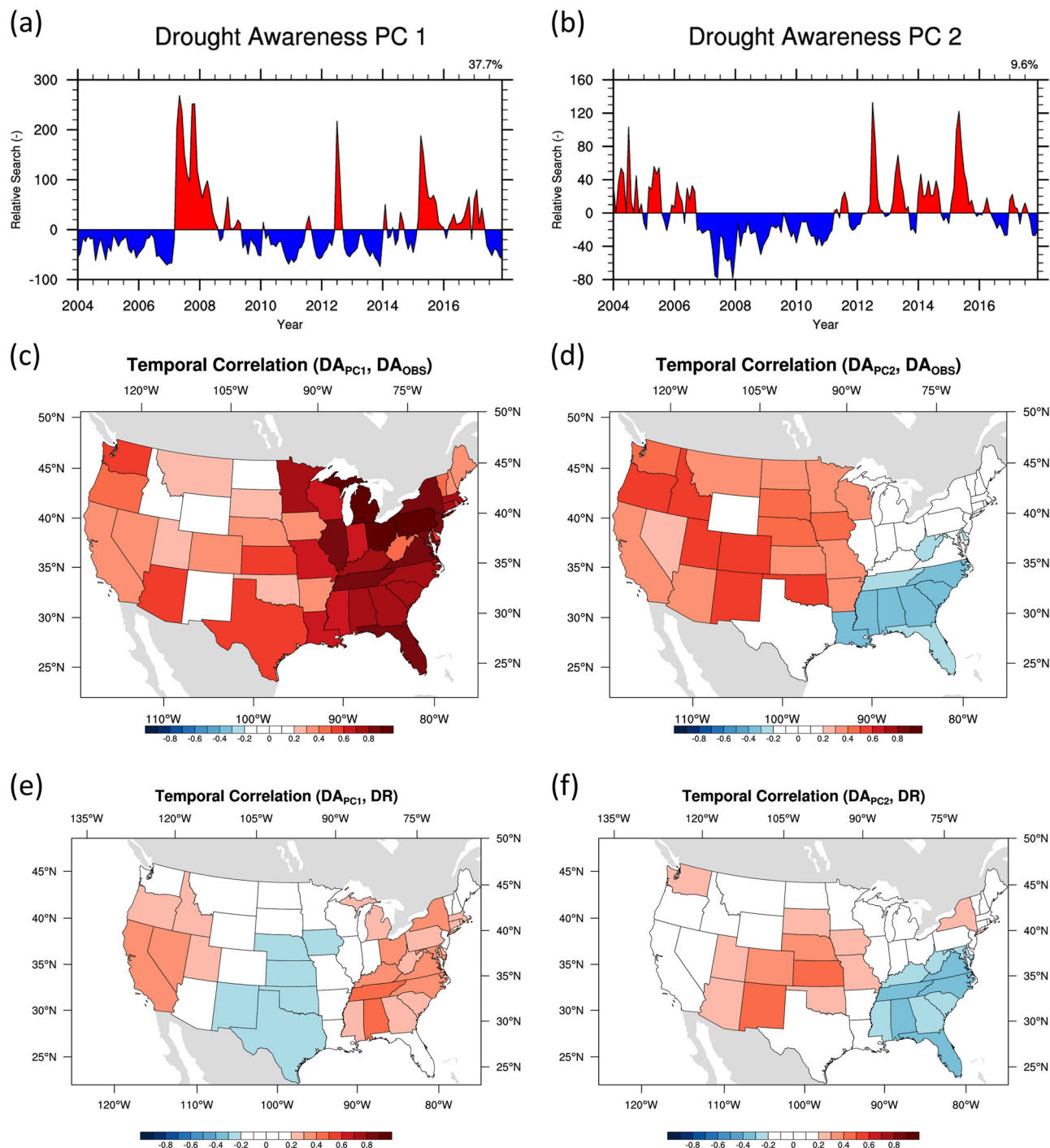
Fig. 1 Major principal components of state-level drought awareness and state-level drought risk over the United States. Scores of the first and second principal component modes of drought awareness (DA_{PC1} and DA_{PC2}) are shown in (**a**) and (**b**), respectively. Temporal correlations of DA_{PC1} with DA_i and DR_i are shown in (**c**) and (**e**), respectively. Temporal correlations of DA_{PC2} with DA_i and DR_i are shown in (**d**) and (**f**), respectively

monthly averages of drought awareness are consistent across different averaging methods (not shown) because drought risk (herein, the areal extent of drought) has long-term persistency, which means that the spatial extent under drought does not rapidly change within one month.

Principal component analysis: a tool for spatiotemporal patterns of drought awareness. Principle Component Analysis (PCA) is a statistical method to sort the maximum variance of the

data in a projection using an orthogonal linear transformation (Wilks, 2011). The loadings of PC modes (a 49-element vector) are computed from eigenvectors of the covariance matrix, which reveal the spatial patterns for the corresponding modes. The scores of the PC modes (a 168-element vector) are computed by projecting the monthly Google Trends data (the 49 by 168 matrix) on the loadings of the corresponding PC mode.

Using the PC1 and PC2 mode scores (Fig. 1(a, b), respectively), we create the maps of temporal correlations between each of the first two PC modes (DA_{PC1} and DA_{PC2}) on the one hand, and
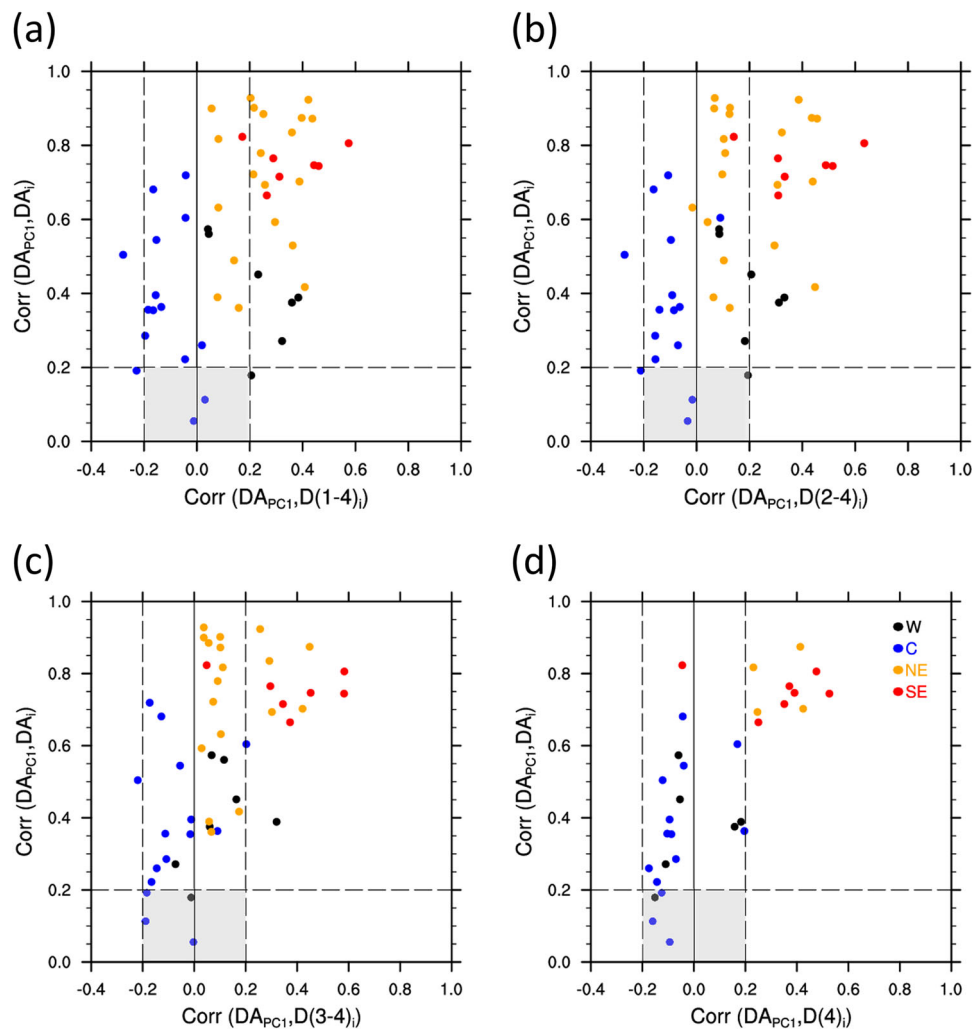
Fig. 2 Impact of drought severity on nation-wide drought awareness. Solid dots depict temporal correlation coefficients between DA_{PC1} and DA_i (y-axis) and DA_{PC1} and DR_i (x-axis; **a** D(1–4), **b** D(2–4), **c** D(3–4), and **d** D(4)). Black, blue, orange, and red indicate Western (W), Central (C), Northeastern (NE), and Southeastern (SE) region, respectively. Dashed lines and the shaded gray box denote the areas with insignificant temporal correlation coefficient at the 99% confident level

DA_i and DR_i, respectively, on the other hand. By doing so, we aim to understand the associations of the major PC modes with state-level drought awareness or state-level drought risk. For example, the map of the temporal correlations between DA_{PC1} and DA_i show the states in which the local drought awareness is temporally significantly correlated with DA_{PC1}. The map of the temporal correlations between DA_{PC1} and DR_i show the states in which the actual condition is positively correlated with DA_{PC1} (drought) and the actual condition is negatively correlated with DA_{PC1} (no drought). A pluvial (a period with increased rainfall) and associated flood often reduce significantly the spatial extent of a drought condition. These temporal correlation maps are known in the atmospheric science field as homogeneous (DA_{PC1} and DA_i) and heterogeneous (DA_{PC1} and DR_i) correlation maps for the PC1 mode. Here, we compute the 49 Pearson sample linear correlations of DA_{PC1} with DA_i and DR_i, respectively. Since the sampling distribution of the correlations is not normally distributed, the Pearson sample linear correlations are converted to Fisher's z-statistic and then are tested for significance at the 95 or 99% confidence level. The homogenous correlation map enables us to quantify the strength of the correlations between DA_{PC1} and DA_i via temporal correlation coefficients that range between −1 and 1, while their key spatial features are similar to those of the corresponding PC loadings. Comparison between the

homogeneous and heterogeneous correlation maps for the first PC mode informs which and how significantly actual drought and flood occurrences (the states with positive and negative correlation coefficients, respectively) contribute to the associated spatial patterns of DA_i with the first PC mode.

In this study, we find that Google Trend data are unstable over certain states due to a lack of search activities in rural areas until the advent of 3G wireless mobile telecommunications technology in 2008 (Lewis, 2018). Therefore, we use PCA to identify these noise states. The scores of three PC modes (PC7, PC10, and PC15) show rapid temporal fluctuations from 0 to 100 over 2004–2008. Their scores are consistent with the temporal patterns of instability of Google Trends data between 2004 and 2008. We superimpose these PC modes (a 168-element vector; the sums of these PC loadings at time t ($t = 1, …, 168$)) and compute their temporal correlation coefficients with state-level drought awareness (DA_i). These superimposed PC scores show statistically significant temporal correlation coefficients at the significance level, 0.05, with DA_i in nine states including Idaho, Wyoming, North Dakota, Colorado, Vermont, New Hampshire, Maine, Rhode Island, and Mississippi. These nine states were therefore defined as the noise states of Google Trends for drought awareness analysis (see Section 1 in Supplementary Information).
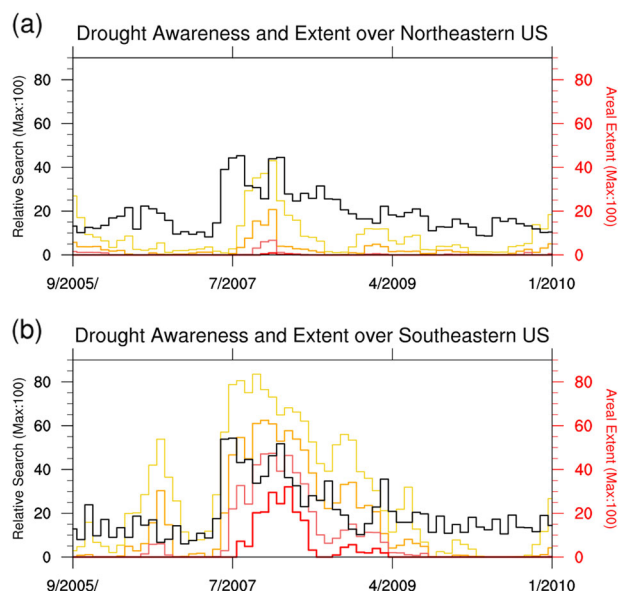
Fig. 3 Temporal patterns of drought awareness and drought risk over the Northeastern and Southeastern US states ((a) and (b), respectively). In **a** and **b**, dark red, light red, orange, gold, and khaki lines depict monthly time series of regional averages of the areal extents under exceptional drought (D4), extreme drought (D3), severe drought (D2), moderate drought (D1), and abnormally dry (D0). In **a** and **b**, black lines depict monthly time series of regional averages of drought awareness

Results

Spatiotemporal patterns of US drought awareness from the PCA. The first two principal component modes of drought awareness (DA_{PC1} and DA_{PC2}) can explain about half of the total variance of drought awareness in the contiguous United States (CONUS) (38 and 10% of the variance, respectively). The peaks in the scores of DA_{PC1} are well correlated with the occurrences of the three major US droughts, namely the 2007–2008 Southeastern and Western US drought, 2012 Midwestern US drought, and 2011–2017 California drought (Fig. 1(a)). The map of the temporal correlation between DA_{PC1} and DA_i shows positive correlations across the states (the national drought awareness; Fig. 1 (c)). This national drought awareness is driven by drought occurrences over the Eastern and Western US regions, namely the 2007–8 Southeastern US drought and California droughts during 2007–2008 and 2011–2017 (Fig. 1(e)). Negative correlations between DA_{PC1} and DR_i over the Southern Great Plains states are driven by the 2007 flood extremes (Dong et al., 2011). Historically, co-occurrence of a flood over the Southern Great Plains and a drought over the Southeastern US originate from one mechanism related to moisture transport from the Gulf of Mexico. Weak southerly meridional (in the south-north direction) fluxes and westerly zonal (in the west-east direction) moisture fluxes at low level cause anomalous moisture transport from the Gulf of Mexico and thus precipitation surplus over the Southern Great Plains and precipitation deficit over the Southeastern US region (Kam et al., 2014).

The map of temporal correlation between DA_{PC2} and DA_i identifies the Central Great Plains and the western part of the Southwest US region as the regions associated with the PC2 mode (Fig. 1(b)). The peaks in the scores of DA_{PC2} match well with the occurrence of the 2012 Midwestern US drought (Fig. 1(d)). The DA_{PC2} scores show negative anomalies during 2007–2010 when flood extremes occurred over the Southern Great Plains. These negative anomalies drive negative correlations between DA_{PC2} and DA_i over Southeastern US states. It indicates that DA_{PC2}

captures regional-scale search activities among the drought-affected states. The third principal component (DA_{PC3}; 8% of the variance) is related to the 2011–2017 California and 2016–2017 Northeastern US drought (see Section 2 in Supplementary Information) at a smaller regional scale than the two major PC modes.

Sensitivity test of drought severity to drought awareness. To understand the impact of drought severity on correlations between DA_{PC1} and DR_i, we compute temporal correlations between DA_{PC1} and DR_i with the four different levels of severity (x-axis in Fig. 2). Here, the four levels are classified as the cumulative areal extents under moderate through exceptional drought (D(1–4)), severe through exceptional drought (D(2–4)), extreme and exceptional drought (D(3–4)), and only exceptional drought (D4). As the severity of the drought increases, the areal extent decreases. Results show that a majority of the Northeastern and Southeastern US states show statistically significant temporal correlations between DA_{PC1} and DA_i (the y-axis in Fig. 2), regardless of the severity. Most of Northeastern US states (16 out of 20 states) have not experienced drought with severity of D4 over 2004–2017 however drought awareness of these Northeastern US states show strong temporal correlations with DA_{PC1} (Fig. 2(d); see Section 3 in Supplementary Information).

To further investigate the timing of the peak of search activities over the Northeastern and Southeastern US, we compute monthly time series of the regional averages of drought awareness during the 2007–2008 Southeastern US drought. Results show that the timing of the first peak of search activities over these US regions happened in June 2007 when the Northeastern US experienced no local drought condition but search activities in the Northeastern US grew at the same rate as search activities in the Southeastern US did (Fig. 3). It indicates that the Northeastern US is very sensitive to drought emergency, regardless of the geographical location of the occurrence.

Most of the Western and Central US states show insignificant temporal correlation coefficients between DA_{PC1} and DR_i (D(4); Fig. 2(d)). Most of these regions except for New Mexico, Idaho, Wyoming, and North Dakota show relatively weaker, but still significant, temporal correlations between DA_{PC1} and DA_i than the Southeastern and Northeastern states. The results imply that the Western and Central states are also aware of drought regardless of whether they experience local drought or not.

Impact of lead and lagged times on drought awareness and risk. To further investigate how timely each state responds to actual local droughts and how fast local drought awareness spreads across the nation, we computed temporal correlations of DA_{PCA1} and DA_i from three lead to three lagged months (−3 through +3 month time step; Fig. 4). Correlations at one month lead and lag are computed by temporal correlations between DA_i ($t−1$) and DA_{PCA1} (t) ($t = 2, …, 168$) and DA_i ($t+1$) and DA_{PCA1} (t) ($t = 1, …, 167$), respectively. All four CONUS regions show the strongest temporal correlations of DA_{PCA1} and DA_i at zero lead month, indicating that awareness of a major drought can reach to the national-scale from the local drought occurrence within a month. The eight Northeastern states including Maryland, Michigan, New Jersey, New York, Ohio, Pennsylvania, Virginia, and Washington DC, show stronger temporal correlation coefficients of DA_{PC1} and DA_i than other Northeastern US states. The results suggest that a high level of public awareness of drought over these Northeastern states might be due to their strong socioeconomic background, such as high GDP per capita, and high education level among others. Earlier research has shown the strong predictive power of educational attainment in
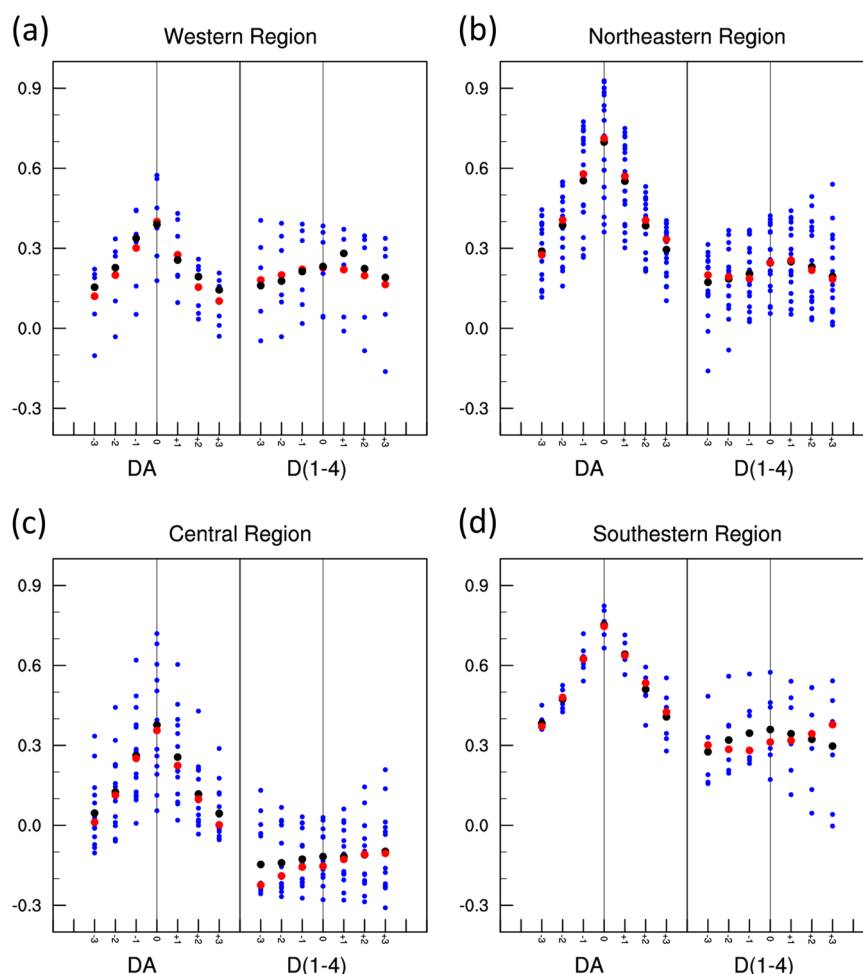
Fig. 4 Lead and lagged correlations of the nation-wide drought awareness (DA$_{PC1}$) with the state-level drought awareness (DA$_i$) or risk (DR$_i$) over the four US regions (**a** Western Region, **b** Northeastern Region, **c** Central Region, and **d** Southeastern Region). Blue dots depict the temporal correlation coefficients of DA$_{PC1}$ and DA$_i$ (left in each panel)/DR$_i$ (right) from −3 months through +3 months. Black (red) dots depict the means (medians) of lead and lagged temporal correlation coefficients of the corresponding states to each region

climate change awareness worldwide (Lee et al., 2015) and income in local drought awareness in the U.S. (Switzer and Vedlitz, 2017).

Moreover, temporal correlations between DA$_{PC1}$ and DR$_i$ at the severity level, D(1–4), are higher at the lagged months in the four US regions than those at the lead months. The Western and Northeastern regions have the strongest temporal correlation coefficients of DA$_{PC1}$ and DR$_i$ at one lagged month (+1 month) and the Southeastern states show the strongest temporal correlations at three lagged months (+3 month). Positive correlations between DA$_{PC1}$ and DA$_i$ of the Central states indicate that their residents who experienced the 2007–2008 floods (negative temporal correlation between DA$_{PC1}$ and DR$_i$) are willing to search the information for droughts that occurred remotely like the 2007–2008 Southeastern drought.

In the Southeastern US states, correlations between DA$_{PC1}$ and DR$_i$ for D(3 − 4) are the strongest at the three lagged months when the monthly drought areal extent values (DR$_i$ (t + 3)) lead the monthly values of nation-wide drought awareness (DA$_{PC1}$ (t)) (Fig. 5). North Carolina and Florida have different regional climate systems that are influenced by not only the Gulf of Mexico but also the Atlantic Ocean, leading to insignificant correlations of DA$_{PC1}$ and DR$_i$. The results suggest that people in this region are quick to be aware of the development of local drought condition (D1 or D2) and their awareness reaches its

peak even before the actual condition arrives at its climax. For example, the Southeastern states showed timely public awareness of drought even before the peak of drought such as the 2007–2008 drought (see Fig. 3). The public was certainly aware of this ongoing drought. This suggests that the substantial economic losses during the 2007–2008 Southeastern US drought might be due to other factors such as not having effective and efficient drought policy and mitigation plans in place. A further study obviously needs to be conducted to investigate the effectiveness and efficiency of the current drought policy and mitigation plans in the Southeastern US to validate this speculation.

State-level correlation analysis of drought awareness and drought risk. To investigate linkages of drought risk among states, we compute temporal correlations of DR$_i$, among the 49 US states (left triangle in Fig. 6). Within a region, temporal correlations of DR$_i$ among the states are significantly positively correlated, indicating that the areal extent of severe US droughts tends to be at least the regional-scale (Sheffield et al., 2009a). The Southeastern states and the adjacent Northeastern states such as Delaware, Kentucky, Maryland, Virginia, and Washington DC show significantly positive correlations. The Southeastern region and the Northwestern region show negative correlations of DR$_i$ due to the 2006 pluvial floods in Idaho, Oregon, and Washington.

These extreme events of the Northwestern regions was caused by landfalling intense atmospheric rivers (Neiman et al., 2008). We also compute temporal correlations of DA_i among the 49 US states to investigate connections of drought awareness. The temporal correlations of DA_i show that the Northeast US states are significantly positively correlated with all the other US states, except for Montana, Wyoming, North Dakota, Oklahoma, and

South Dakota. It indicates that a high level of drought awareness in the Northeastern US states play a role in reaching national drought awareness over the US (DA_{PC1}).

Linear regression analysis of socioeconomic factors. The strength of temporal correlations between DA_{PC1} and DA_i might be determined by different natural conditions and socioeconomic structures (see Section 3 in Supplementary Information). Simple linear regression analyses are conducted to assess the effects of three socioeconomic factors, including the ranks of each state by population, GDP per capita, and education, (the x-axis in Fig. 7 (a–c), respectively) on correlations between DA_{PC1} and DA_i among all 49 states (the y-axis in Fig. 7). In this study, the percentage of the adult resident with bachelor's degree or higher[1] is used for an indicator of education level (see Section 4 in Supplementary Information). A state with a lower rank show a more population, a higher GDP per capita, and a high percentage of the adult residents with bachelor's degree or higher. We find that population is the highest influential socio-economic factor on correlations between DA_{PC1} and DA_i explaining 40% of the total variance ($R^2 = 0.41$). The ranks of GDP per capita and education explain less than 5% of the total variance. As Fig. 6(a) demonstrates, three outliers exist. California, the most populated state, shows a relatively low temporal correlation between DA_{PC1} and DR_{CA} possibly due to the significant increase of recent search activities in recent years (Gonzales and Ajami, 2017). Despite having the second most population, Texas shows a correlation between DA_{PC1} and DR_{TX}, 0.5, much lower than the expected correlation from the linear model (0.82; $-0.01 \times X + 0.84$ from Fig. 7(a), X is the rank of each state by population). This low correlation of Texas might be due to the 2007–8 flood events in the Southern Great Plains. The District of Columbia (DC), in spite of having a small population (the 47th ranked in population), shows a high temporal correlation between DA_{PC1} and DA_{DC}, above 0.8, which is much higher than the expected correlation, 0.37. These results indicate other factors including both socioeconomic and natural ones should be accounted for in order to identify powerful predictors of drought awareness.

Discussion
Large populations, high GDP per capita, and high percentage of the population with bachelor degree (i.e., the low ranks of

Fig. 5 Impact of drought severity on correlations between DA_{PC1} and DR_i over the Southeastern US states. Lead and lagged temporal correlation coefficients between DAPC1 and DAi of the Southeastern US states are shown in the y-axis with the total areal extents from abnormal dry through exceptional (D0–4), moderate through exceptional (D1–4), severe through exceptional (D2–4), extreme and exceptional (D3–4), and exception (D4) in the x-axis. Red and black solid circles depict the means and medians of the temporal correlations of DAPC1 and DRi over the Southeastern US states, respectively, from −3 months through +3 months. Blue cross, asterisk, and empty circles depict the temporal correlations of DAPC1 and DRi over the state of Florida, North Carolina, and the rest five states of the Southeastern US region
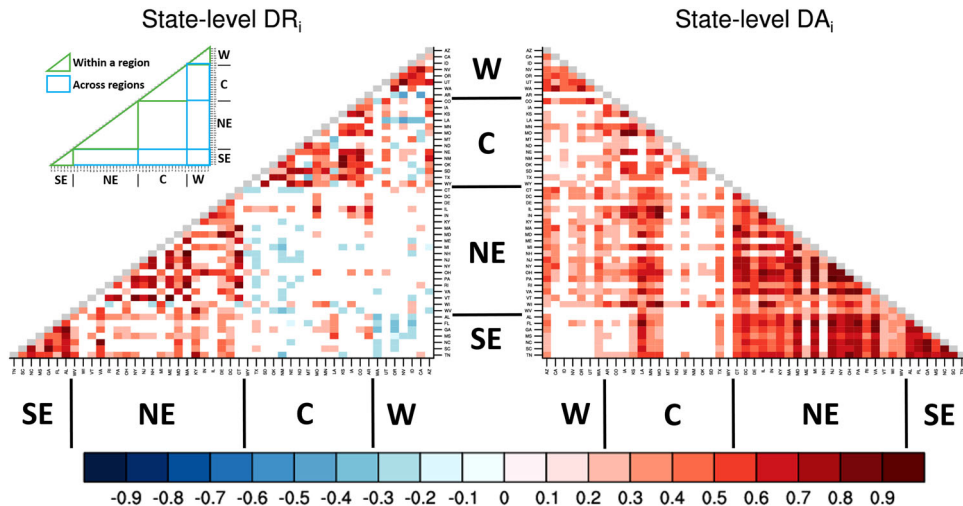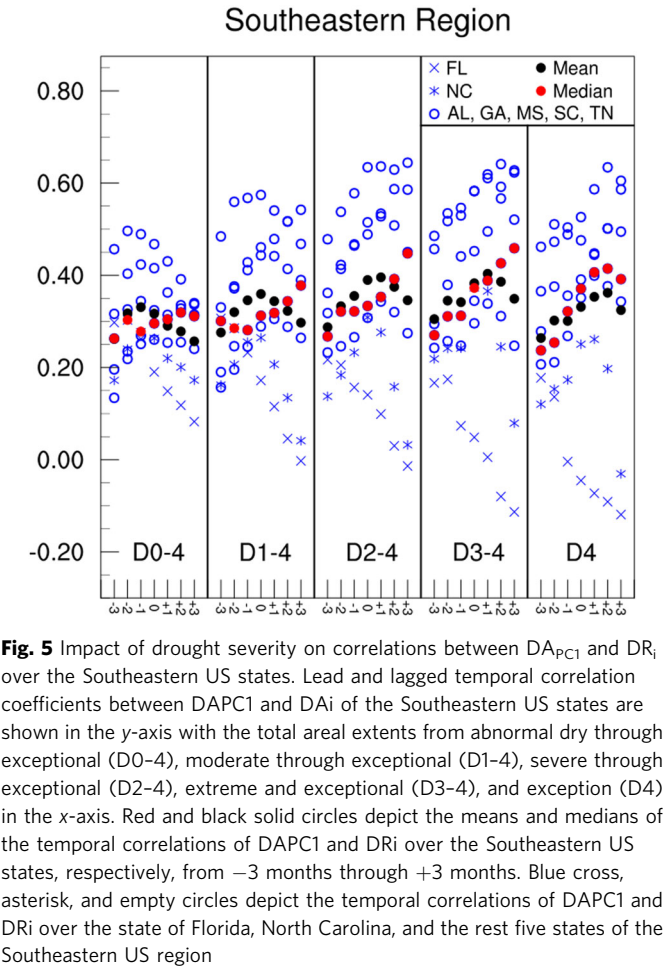

Fig. 6 State-level correlation analysis of drought awareness and drought risk. Colors in the grid cells depict temporal correlation coefficients of DAi (left triangle) and DRi (right triangle; D(1–4)) of one state drought awareness with those of the rest 48 states. To assist interpretation, areas that depict interstate relationships within the region (green colored boundaries) and across regions (blue colored boundaries) are shown in the legend box at upper-left corner
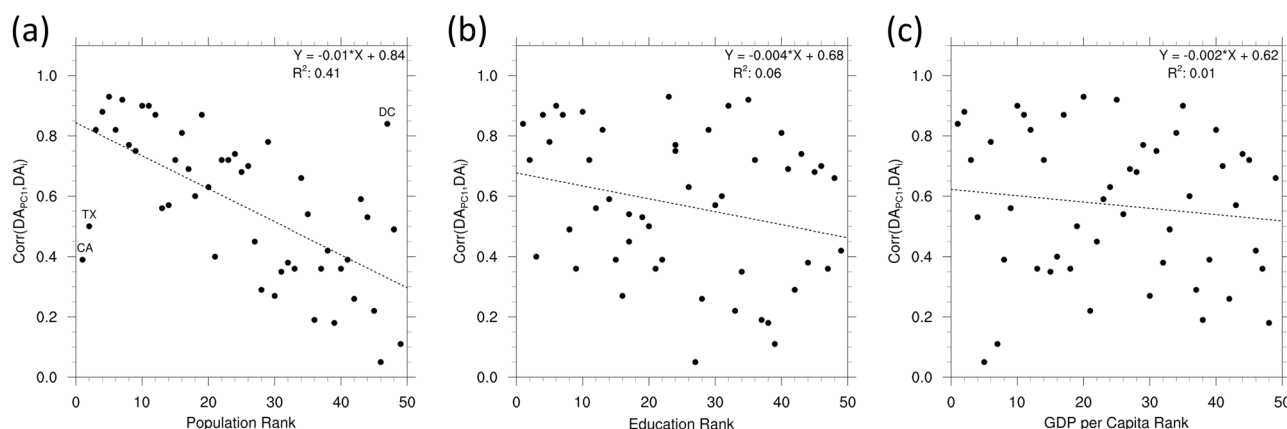
Fig. 7 Impact of socioeconomic factors on correlations between DA_{PC1} and DA_i. In the x axis, the states with lower ranks have more population (**a**), higher percentage of the residents with bachelor's degree or higher (education) (**b**), and more GDP per capita (**c**) (e.g., 1 is the state with the most population). Dotted line depict the regression lines between the ranks and the temporal correlation coefficients between DA_{PC1} and DA_i. The equation of the regression line and R^2 are shown in the top-right corner of the sub-figures

socioeconomic indicators) among the Northeastern and Western states still suggest that residents in these states would be willing to search information for not only local but also remote droughts and are more aware of issues beyond their geographic boundaries. This speculation is grounded in Maslow's theory of human motivation, suggesting that people tend to satisfy their physiological needs such as food and shelter before turning their attention to other needs such as self-actualization, aesthetics, and environmental values (Maslow, 1943). Simply put, people with more socio-economic resources can afford to show concern for environmental issues even when they occur elsewhere. Past severe drought experiences (Dustbowl drought in 1930s; Schubert et al., 2004) and economic interests (e.g., agriculture states) in the Central US states might condition these residents to be highly alert to droughts and thus propel drought awareness to reach a nation-wide level. A further analysis with other socio-economic factors from different data sources (active monitoring data and traditional survey data) are required to reveal the key socio-economic factors on the dynamics of drought awareness diffusion and how local drought awareness is interacted across states.

The observational results from Google Trends raise the following questions about a causal mechanism of drought awareness diffusion. What factors create this high level of drought awareness in the Northeast region? In addition to the aforementioned socio-economic status, may the concentration of federal agencies in this region be another factor? These agencies might be involved in drought monitoring and forecasting (e.g., Department of Agriculture, National Aeronautics and Space Administration, and National Oceanic and Atmospheric Administration) and policy-making (e.g., The Whitehouse and Congress) and thus the Northeastern US be highly alert to the emerging droughts in any US state. Is the Northeastern US a hub of the diffusion routes of drought information?

We propose a hypothesis about the US drought information diffusion route: socio-politically influential states, not the drought-affected states, serve as a key factor driving the national-scale drought awareness through an information diffusion route. The data and methods used in this study are limited in testing this hypothesis. First, Google Trends data do not provide an indication of how information is disseminated. Second, the temporal correlation analysis between DA_{PC1} and DA_i reflects joint awareness in some states, and an information diffusion route is far from being determined. To explore the US drought

information diffusion route, a network structure or model is required via national panel data and active social monitoring data such as Twitter (e.g., Fig. 5 of Musaev et al. (2018)) to investigate possible hubs and inter-state connections. However, this study provides insights that will generate further interests in identifying a drought information diffusion route and help policy makers design and implement more effective and efficient drought mitigation plans at the interstate level.

Conclusion

This study presents the observational evidence of the spatio-temporal pattern of the US drought awareness. The key findings are (1) Northeastern US is the most sensitive to the emergence of drought, regardless of the geographic location of the occurrence; (2) once a major drought's impact is felt, others who live nearby search for the information about drought almost simultaneously within one month. This study succeeds in expanding the utility of passive social monitoring data from billions of the users of Google Products to reveal the social dynamics of drought awareness during the propagation of drought at the national scale.

More in-depth research is required to identify a causal mechanism of drought awareness diffusion through an integrative qualitative and quantitative assessment. Social responses to droughts are multidimensional and consist of awareness, emotions, perceptions, and actions. Social awareness of drought can certainly interact with other types of relevant natural hazards (e.g., flood, wildfire, and heatwave) and it can be conducive to the formation of risk perception, which in turn can facilitate adaptation actions. Eventually, the advanced knowledge of social dynamics via a synthetic assessment of surveys and social monitoring data (passive and active) will be instrumental in the development of proactive drought mitigation programs, such as national water saving incentive programs, by utilizing the key players in the existing drought diffusion route.

Data availability

All datasets analyzed or generated are presented in the paper.

Notes

1 https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_17_5YR_S1501

References

Anderegg WRL, Goldsmith GR (2014) Public interest in climate change over the past decade and the effects of the 'climategate' media event. Environ Res Lett 9:054005

Carlton JS, Mase AS, Knutson CL, Lemos MC, Haigh T, Todey DP, Prokopy LS (2016) The effects of extreme drought on climate change beliefs, risk perceptions, and adaptation attitudes. Clim Change 135:211–26

Carneiro HA, Mylonakis E (2009) Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis 49:1557–64

Cook BI, Ault TR, Smerdon JE (2015) Unprecedented 21st century drought risk in the American Southwest and Central Plains. Sci Adv 1:e1400082

Cook BI, Miller RL, Seager R (2009) Amplification of the North American "Dust Bowl" drought through human-induced land degradation. PNAS 106:4997–5001

Dong X, Xi B, Kennedy A, Feng Z, Entin JK, Houser PR, Schiffer RA, L'Ecuyer T, Olson WS, Hsu K, Liu WT, Lin B, Deng Y, Jiang T(2011) Investigation of the 2006 drought and 2007 flood extremes at the Southern Great Plains through an integrative analysis of observations. J Geophys Res 116:D03204

Gonzales P, Ajami N (2017) Social and structural patterns of drought-related water conservation and rebound. Water Resour Res 53:10619–34

Hayes MJ, Wilhelmi OV, Knutson CL (2004) Reducing drought risk: bridging theory and practice. Nat Hazards Rev 5:106–13

Hoerling M, Eischeid J, Kumar A, Leung R, Mariotti A, Mo K, Schubert S, Seager R (2013) Causes and Predictability of the 2012 Great Plains Drought. Bull Am Meteor Soc 95:269–82

Kam J, Sheffield J, Wood EF (2014) A multiscale analysis of drought and pluvial mechanisms for the Southeastern United States. J Geophys Res Atmos 119:2014JD021453

Kam J, Stowers K, Kim S (2019) Monitoring of drought awareness from google trends: a case study of the 2011–17 California Drought. Wea Clim Soc 11:419–429

Lazrus H (2016) "Drought is a relative term:" drought risk perceptions and water management preferences among diverse community members in Oklahoma, USA. Hum Ecol 44:595–605

Lee TM, Markowitz EM, Howe PD, Ko C-Y, Leiserowitz AA (2015) Predictors of public climate change awareness and risk perception around the world. Nat Clim change 5:1014

Lewis M (2018) The fifth risk. WW Norton & Company, New York, NY

Maslow AH (1943) A theory of human motivation Psychological Rev 50:370–96

Musaev A, Kam J, Stowers K (2018) Harnessing data to create an effective drought management system. In: The 15th Internation Conference on Information Systems for Crisis Response and Management (ISCRAM). ISCRAM Association and Rochester, New York, NY

Neiman PJ, Ralph FM, Wick GA, Kuo Y-H, Wee T-K, Ma Z, Taylor GH, Dettinger MD (2008) Diagnosis of an Intense Atmospheric River Impacting the Pacific Northwest: Storm Summary and Offshore Vertical Structure Observed with COSMIC Satellite Retrievals. Mon Wea Rev 136:4398–420

Nguyen H-L, Woon Y-K, Ng W-K (2015) A survey on data stream clustering and classification. Knowl Inf Syst 45:535–69

Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, Murugiah K (2014) The use of google trends in health care research: a systematic review. PLoS ONE 9:e109583

Popper B (2017) Google announces over 2 billion monthly active devices on Android. The Verge Online: https://www.theverge.com/2017/5/17/15654454/android-reaches-2-billion-monthly-active-users

Preis T, Moat HS, Stanley HE (2013) Quantifying trading behavior in financial markets using Google Trends. Sci Rep 3:1684

Quesnel KJ, Ajami NK (2017) Changes in water consumption linked to heavy news media coverage of extreme climatic events. Sci Adv 3:e1700784

Schubert SD, Suarez MJ, Pegion PJ, Koster RD, Bacmeister JT (2004) On the cause of the 1930s Dust Bowl. Science 303:1855–9

Schubert SD, Suarez MJ, Pegion PJ, Koster RD, Bacmeister JT (2008) Potential predictability of long-term drought and pluvial conditions in the US Great Plains. J Clim 21:802–16

Seager R, Hoerling M, Schubert S, Wang H, Lyon B, Kumar A, Nakamura J, Henderson N (2015) Causes of the 2011–14 California Drought. J Clim 28:6997–7024

Seager R, Neelin D, Simpson I, Liu H, Henderson N, Shaw T, Kushnir Y, Ting M, Cook B (2014) Dynamical and thermodynamical causes of large-scale changes in the hydrological cycle over North America in Response to Global Warming. J Clim 27:7921–48

Seager R, Tzanova A, Nakamura J (2009) Drought in the southeastern United States: causes, variability over the last millennium, and the potential for future hydroclimate change. J Clim 22:5021–45

Shao W, Gardezi M, Xian S (2018) Examining the Effects of Objective Hurricane Risks and Community Resilience on Risk Perceptions of Hurricanes at the County Level in the U.S. An Innovative Approach Annals of the American Association of Geographers, Gulf Coast. https://www.tandfonline.com/doi/abs/10.1080/24694452.2018.1426436

Sheffield J, Andreadis KM, Wood EF, Lettenmaier DP (2009a) Global and continental drought in the second half of the twentieth century: severity–area–duration analysis and temporal variability of large-scale events. J Clim 22:1962–81

Sheffield J, Andreadis K, Wood E, Lettenmaier D (2009b) Global and continental drought in the second half of the twentieth century: Severity–area–duration analysis and temporal variability of large-scale events. J Clim 22:1962–81

Smith AB, Matthews JL (2015) Quantifying uncertainty and variable sensitivity within the US billion-dollar weather and climate disaster cost estimates. Nat Hazards 77:1829–51

Stoutenborough JW, Vedlitz A (2014) Public attitudes toward water management and drought in the United States. Water Resour Manag 28:697–714

Switzer D, Vedlitz A (2017) Investigating the determinants and effects of local drought awareness. Wea Clim Soc 9:641–57

Wang B, Zhuang J (2017) Crisis information distribution on Twitter: a content analysis of tweets during Hurricane Sandy. Nat Hazards 89:161–81

Wang B, Zhuang J (2018) Rumor response, debunking response, and decision makings of misinformed Twitter users during disasters. Nat Hazards 93:1145–62

Wilks DS (2011) Statistical methods in the atmospheric sciences. Vol. 100. Academic press, Cambridge, MA

Yuan X, Wood EF, Luo L, Pan M(2011) A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction. Geophys Res Lett 38: L13402

Acknowledgements

Author contributions

JK conceived the study with inspiration from his previous studies. SK performed the analyses and SK and JK prepared the figures. JK and WS mainly wrote the paper. All authors contributed to discussion and the paper.

Competing interests

The authors declare no competing interests.

Additional information