10-2020

Calibrating Human Attention as Indicator: Monitoring #drought in the Twittersphere

Kelly Smith

Andrew Tyre

Zhenghong Tang

Michael Hayes

Adnan Akyuz

BAMS Article

Calibrating Human Attention as Indicator

Monitoring #drought in the Twittersphere

Kelly Helm Smith, Andrew J. Tyre, Zhenghong Tang,
Michael J. Hayes, and F. Adnan Akyuz

ABSTRACT: State climatologists and other expert drought observers have speculated about the value of monitoring Twitter for #drought and related hashtags. This study statistically examines the relationships between the rate of tweeting using #drought and related hashtags, within states, accounting for drought status and news coverage of drought. We collected and geolocated tweets, 2017–18, and used regression analysis and a diversity statistic to explain expected and identify unexpected volumes of tweets. This provides a quantifiable means to detect state-weeks with a volume of tweets that exceeds the upper limit of the prediction interval. To filter out instances where a high volume of tweets is related to the activities of one person or very few people, a diversity statistic was used to eliminate anomalous state-weeks where the diversity statistic did not exceed the 75th percentile of the range for that state's diversity statistic. Anomalous state-weeks in a few cases preceded the onset of drought but more often coincided with or lagged increases in drought. Tweets are both a means of sharing original experience and a means of discussing news and other recent events, and anomalous weeks occurred throughout the course of a drought, not just at the beginning. A sum-to-zero contrast coefficient for each state revealed a difference in the propensity of different states to tweet about drought, apparently reflecting recent and long-term experience in those states, and suggesting locales that would be most predisposed to drought policy innovation.

https://doi.org/10.1175/BAMS-D-19-0342.1
Corresponding author: Kelly Helm Smith, ksmith2@unl.edu
In final form 24 April 2020

AFFILIATIONS: Smith—National Drought Mitigation Center, School of Natural Resources, University of Nebraska–Lincoln, Lincoln, Nebraska; **Tyre and Hayes**—School of Natural Resources, University of Nebraska–Lincoln, Lincoln, Nebraska; **Tang**—Community and Regional Planning Program, University of Nebraska–Lincoln, Lincoln, Nebraska; **Akyuz**—North Dakota State Climate Office, North Dakota State University, Fargo, North Dakota

The U.S. Drought Monitor, a weekly map showing the location and intensity of drought conditions, is assembled each week by a rotating team of authors assisted by a nationwide listserv of 450 expert interpreters of state and local climate conditions. The U.S. Drought Monitor is based on a "convergence of evidence," incorporating many streams of objective data, reconciled by expertise, including input and interpretation from state and local professionals, and observations about local conditions. Subjective inputs include text descriptions, sometimes with photos, sent via the nationwide listserv; observations from the highly successful citizen science project, Collaborative Community Rain Hail and Snow Network (CoCoRaHS); and observations submitted to the National Drought Mitigation Center's (NDMC) Drought Impact Reporter (DIR). Motivating volunteer observers has been an ongoing question: CoCoRaHS observers are consistent, but they depend on an intensive education and support process and do not have the capability to submit photos, while DIR observations tend to come in surges that appear to be associated with triggering assistance under the Livestock Forage Disaster Program (Lackstrom et al. 2013, 2017; Meadow et al. 2013; Smith et al. 2014). Comparison of producers' recollections of a fast-emerging drought in 2016 with objective drought indicators found that qualitative reports from a written survey could help assess the accuracy of high-resolution drought monitoring datasets (Otkin et al. 2018). Farmers, ranchers, agricultural advisors, and others sometimes share observations about field conditions and their experiences via Twitter, yielding data on planting progress that correlated with National Agricultural Statistics Service (Zipper 2018). Hence, contributors to the U.S. Drought Monitor process, some of whom interact with stakeholders via Twitter, have asked whether monitoring tweets could provide additional useful information, in part as an alternative to information without the bias introduced by the fact that the DIR is perceived by many as a means of providing input to the U.S. Drought Monitor map, and without the time investment required to generate and sustain CoCoRaHS observations. More broadly, drought researchers cite the need to calibrate hydrometeorological drought indicators for relevance by comparing them with impact data (Hayes et al. 2011; Kallis 2008; Van Loon et al. 2016), as well as the difficulty in assembling fully representative quantitative drought impact data, because impact data either tend to be narrowly focused on single-sector results such as crop yield or exist across collections of painstakingly assembled, qualitative, text-based reports that lend themselves to statistical analysis as presence–absence data (Stahl et al. 2016).

Literature review

Several academic and applied disciplines, including natural hazards, political communication, media ecology, epidemiology, and data science, are experimenting with using social media, particularly Twitter, to monitor and detect events of interest. Reflecting growing interest exploring uses of social media in resource management, a survey of environmental research using social media as data found the number of studies increased from 1 in 2011 to 61 in 2017, and divided them into studies of people, including attitudes toward the natural world; of nature, such as observations about land conditions; and related to planning and policy, such as disaster response (Ghermandi and Sinclair 2019). Goodchild (2007) described a sensor network that consists of humans themselves, each equipped with some working subset of the five senses and with the intelligence to compile and interpret what they sense, and each free to rove the surface of the planet. This network of human sensors has over 6 billion components, each an intelligent synthesizer and interpreter of local information (p. 218).

Sakaki et al. (2010) observed that their Twitter-based system for detecting earthquakes was faster than the Japanese Meteorological Agency, and that each tweet represented sensory information. In Bangkok in 2011, information on social media reduced flood loss by an average of 37% by giving enough warning time to move belongings to higher ground; this warning information was not available through other sources (Allaire 2016). Near-real-time flood maps for Jakarta could be created from tweets, where a high proportion of the population uses Twitter (Eilander et al. 2016). In the Philippines and Pakistan, tweets and the Global Flood Detection Satellite System provided information about flooding to humanitarian organizations from 1 to 7 days sooner than normal channels (Jongman et al. 2015). In addition to looking for terms such as "earthquake," Australia's Emergency Situation Awareness–Automated Web Text Mining (ESA–AWTM) system specifically looked for reports mentioning infrastructure damage (Cameron et al. 2012) and is available to hazard managers in Australia and New Zealand from Australia's Commonwealth Scientific and Industrial Research Organisation (https://esa.csiro.au/ausnz/about-public.html). Semantic analysis of social media before and after Typhoon Haiyan, which struck the Philippines in 2013, found that microblog content such as tweets could serve as a useful index for damage assessment (Deng et al. 2016). Tweets could be used at least in part to predict distribution of damage from Hurricane Sandy (Guan and Chen 2014; Kryvasheyeu et al. 2016; Zou et al. 2018). Tweets about road damage could provide more timely and accurate information than what was available from official sources alone (Kumar et al. 2014). Twitter can help detect the outbreak of postdisaster diseases (Chen and Xiao 2016; Chunara et al. 2012; Kryvasheyeu et al. 2016).

A key means of detecting a variation from normal, such as a hazard event, is by establishing a baseline that can be associated with normal, and by measuring or detecting variations from normal, so systems that depend on humans as sensors monitor baseline chatter, and detect anomalies, such as sudden increases in uses of certain words, known as "bursts" (Abdelhaq et al. 2013; Cameron et al. 2012; Fitzhugh 2015; Mathioudakis and Koudas 2010), or spikes (Sakaki et al. 2010). However, drought is a slow-moving disaster (Svoboda et al. 2002), so monitoring drought via an increase in the social media conversation is a longer, slower process, comparable to epidemiological surveillance. Epidemiologists establish baseline levels of diseases that they are tracking, so that they can identify or anticipate higher levels of activity (Hess et al. 2014). If we consider humans to be part of an Earth system, tweets about drought can be considered a symptom of a water-short area, or sensory information, in the terminology of Sakaki et al. (2010).

For both weather hazards and epidemiologic surveillance, it is necessary to distinguish chatter prompted by actual events from chatter prompted by discussion of events: the number of tweets on climate extremes or weather events could be predicted by media coverage of climate extremes, along with the extremes themselves (Kirilenko et al. 2015; Ripberger et al. 2014). Efforts to detect the flu via social media found that media reports about the flu increased tweets but confounded the detection process "because media attention increases 'chatter'—messages that are about influenza but that do not pertain to an actual infection" (Broniatowski et al. 2013, p. 1).

Analysis of Google Trends searches for drought as a measure of awareness during California's 2011–17 drought found that both social triggers such as official responses to drought and natural triggers such as drought itself contributed to sustained awareness of drought (Kam et al. 2019). Surveying Texans, political scientists found that the level of drought severity is the strongest predictor of drought awareness, along with ideology and demographics (Switzer and Vedlitz 2017). They gauged drought awareness by whether or not survey respondents could correctly state whether they had been in drought in the past year, according to the U.S. Drought Monitor.

In the realms of both hazards and politics, awareness, sometimes known as "situational awareness" or understood as a degree of attention, is seen as a precursor to response or to mitigative policy action (Ripberger et al. 2014; Tang et al. 2015). Synthesizing theoretical and empirical work across several fields, Silver (2019, p. 301) defines attention as "the process of noticing, selecting, and focusing on one or more external stimuli (e.g., hazardous event or event-related information) to which people are exposed." Scholars of political communications commonly identify news coverage as serving an agenda-setting function, telling people what to think about (McCombs and Shaw 1972). Although social media complicates the picture, providing a medium for individuals to communicate back to media (Searles and Smith 2016), national newscasts and major newspapers still tend to lead with the same stories, and studies consistently find that media attention to an issue predicts citizens' attention (Gruszczynski and Wagner 2017). Compared day to day, issues related to public order, including natural disasters, were more likely to receive attention first from social media and then from traditional news media (Neuman et al. 2014). A study of tweets about a storm and tornado warning in Ontario found that citizens were more likely than weather professionals to share personal observations of the event (Silver and Andrey 2019).

Within the agricultural sector, some farmers and ranchers tweet as a form of "agvocacy" (Burgess et al. 2015). The Ag Chat Foundation, established to promote agvocacy, defines it as "ag proactively telling our story" (AgChat Foundation 2019) and organizes a weekly Twitter event using the hashtag #agchat.

Methods

We hypothesize that the number of tweets about #drought are explained by news about drought (Broniatowski et al. 2013; Gruszczynski and Wagner 2017; Kirilenko et al. 2015; McCombs and Shaw 1972; Ripberger et al. 2014; Searles and Smith 2016) and by drought itself (Broniatowski et al. 2013; Kirilenko et al. 2015; Switzer and Vedlitz 2017). Then we quantitatively and qualitatively investigate tweet volumes that are higher than predicted to see whether they reflect personal experience of emerging drought (Neuman et al. 2014), sustained awareness from long-term exposure (Kam et al. 2019), or other influences.

We use regression analysis to predict the number of #drought tweets (the dependent variable) for each state-week (the unit of analysis) using four independent variables: drought status on the U.S. Drought Monitor, news about drought, and population, as well as an estimated variable, states' propensity to tweet about drought. We identify higher-than-expected number of tweets that are not accounted for by either drought status or news stories about drought. These unaccounted-for volumes of tweets, by definition, may reflect experiences that are not already depicted on the U.S. Drought Monitor or picked up in news. A diversity statistic screens out state-weeks with a high volume of tweets from one person or a very small number of users. We then analyze the resulting anomalous state-weeks to see whether they provide early warning of emerging conditions or other potentially useful information.

Data

Geolocating tweets. One of the first hurdles in using Twitter to detect geographically specific events is settling on a method for associating tweets with locations. Only a miniscule portion of people who tweet about drought have enabled geotagging on their mobile devices, which provides latitude–longitude coordinates for the origin of each tweet. For personal safety reasons, the default geotagging setting is "off." But text-based location information is much more common, with 71.4% of Twitter users in 2013 filling in the user location field in their

Twitter user profiles, which is generally interpreted as where people live (Leetaru et al. 2013). Geocoding services can translate this information into latitude and longitude coordinates, with varying degrees of accuracy, depending in part on the precision and clarity of the text the user entered. "Earth," for example, is not particularly informative and results in a disproportionate number of tweets for an arbitrary location assigned by the geocoding service, and geocoding sometimes assigns coordinates to users' whimsical entries, such as "the void" or "middle earth" if it is the name of an establishment anywhere on the planet. Thus, quite a bit of care must be exercised in using geocoded locations (Hecht et al. 2011). Despite these caveats, carefully filtered location text can be used to associate the content of a tweet with location with a reasonable degree of accuracy (Jung and Uejio 2017; Sakaki et al. 2010), particularly when using only geotagged tweets would result in having essentially no data.

Collecting tweets. Our analysis used tweets with user-provided locations at city or county scale. We read the user location field to screen for U.S. cities, counties, metro areas, or small regions. We used Twitter Archiving Google Sheet (Hawksey 2014) for our weekly searches, in part due to the ease of geocoding associated with this method, and the Rtweet package (Kearney 2019) to retrieve user profile information. Search terms were #drought; #drought plus the two-letter postal abbreviation for each state, i.e., #droughtTX; #drought17, #drought2017, #drought18; and #drought2018. A total of 18,914 tweets from 2017 and 2018 made it through our filtering process. We omitted retweets, as we were interested in place-based observations and information, not message amplification. We only used tweets that could be associated with a municipality or county and omitted those with no geographic location or those that only provided a state. We filtered out tweets that specifically mentioned other states or countries, as well as those using drought metaphors for sports and relationships. We excluded tweets from lists of bots and spammers that we identified over time. For this preliminary study at a smaller scale, we also excluded a few professional outliers, because the extreme frequency of their tweets (@EdJoyce) or their larger-than-state focus (@DroughtCenter) would skew results. To arrive at our response variable, we then counted the number of tweets for each state-week, with weeks starting on Monday.

User profile data. We used the Rtweet package to add user profile information to tweets, and then read for themes to devise word sets to group people, and fine-tuned experientially. Main groups and some of the associated terms included agricultural producers or those closely associated with agriculture ("farm," "corn," "calf," "grower," "organic"), media ("news," "journalist," "radar"), and scientists ("PhD," "university," "research," "climatologist").

Privacy. Ensuring ethical use of individuals' information shared via social media is an evolving consideration for researchers contending with both a move toward more open data and recognition that using publicly shared data in ways not originally intended may be objectionable to Twitter users (Fiesler and Proferes 2018; Ghermandi and Sinclair 2019; Zipper et al. 2019). Using hashtagged search terms and filters tended to limit the tweets we collected to a well-defined conversation that was often dominated by media and climate or agricultural professionals. The #drought17 hashtag was also publicized as a means for agricultural producers and others to share their experiences with drought, contributing to a collective understanding and implying that the tweets would receive official consideration. We shared our search results within the network of professional drought observers but not publicly. As a practical matter, Twitter's Terms of Service include the requirement that developers sharing tweets make an effort to learn whether users have removed or modified their tweets (https://developer.twitter.com/en/developer-terms/agreement-and-policy.html), and follow their wishes. Implementing such a process would increase the investment required beyond the scale of this preliminary investigation.

Number of news stories. We used state-specific news stories about drought that were published in the United States, collected via the Meltwater media search service. Meltwater is a company that markets artificial intelligence (AI)-based issue tracking services to public relations professionals. We used a Boolean query to search Meltwater's comprehensive database of news stories to identify drought-related stories. In keeping with U.S. copyright law, search results do not include full text, but fields with headlines, the sentence including the search term, and AI-derived key phrases provide a good sense of the content. We filtered out stories that ran nationwide and stories for each state that had out-of-state content so that we had a good idea of how many media stories about drought within a given state appeared in that state. After filtering, we had a total of 15,640 news stories for the 2-yr period; aggregating them for each state and week created a variable called "newscount."

Drought data. We used the Drought Severity and Coverage Index (DSCI) statistic for each state, each week. The DSCI is a weighted sum of the proportion of an area in each category of drought. It converts the weekly U.S. Drought Monitor categories of intensity into a single value that takes areal coverage and magnitude of the drought into account for an area (county, climate division, climate region, National Weather Service Regions, River Forecast Center Regions, urban areas, as well as USDA Climate Hub regions) (Akyuz 2017). These weekly values can be accumulated throughout a drought period for a given location for comparison with other drought periods. This index, containing not only drought intensity and coverage but also the duration when accumulated over a period of time, can provide a single summary statistic representative of an entire drought.

Population data. We used "Annual Estimates of the Resident Population for the United States, Regions, States and Puerto Rico: 1 April 2010 to 1 July 2018 (NST-EST2018-01)" from the U.S. Census Bureau (U.S. Census Bureau 2019).

Analysis

Statistical modeling. Using state-weeks as the unit of analysis, we modeled the relationship between tweets, news stories, and drought status, accounting for population and other differences between states. The response variable was the number of tweets collected in each state, each week, from 4 January 2017 through 31 December 2018. We fit global models using two main model types, negative binomial, a frequent choice for modeling overdispersed count data, and Poisson inverse Gaussian (PIG), a less common choice for modeling more overdispersed count data (Hilbe 2014). For negative binomial models, we used the gam() function, family = nb() from the "mgcv" R package (Wood 2017). For PIG models, we used the gamlss() function, family = PIG, from the "gamlss" R package (Rigby and Stasinopoulos 2005). We used state population as the log-offset link variable and used sum-to-zero contrast coefficients for each state, so population and unique location factors were taken into account. We expressed newscount and DSCI as standard deviation from the mean, making it easier to interpret the relative influence of each. The global model was

$$\begin{aligned}\ln\left(\lambda_{i,t}\right) = {} & \beta_0 + \beta_1\left(\text{sdnewscount}_{i,t}\right) + \beta_2\left(\text{sdDSCI}_{i,t}\right) \\ & + \beta_{3,i}\left(\text{sdDSCI}_{i,t}\right)\left(\text{sdnewscount}_{i,t}\right) + \beta_{4,i} \\ & + \ln\left(\text{population}_{i,t}\right)\end{aligned} \tag{1}$$

$$y_{i,t} \sim \text{Poisson Inverse Gaussian}\left(\lambda_{i,t}, \lambda + \alpha\lambda^3\right),$$

where i is the state, t is time, β_0 is the coefficient for the intercept, β_1 is the coefficient for standardized news count, β_2 is the coefficient for standardized DSCI, β_3 is the coefficient for the interaction of standardized DSCI and standardized newscount, β_4 is the sum-to-zero contrast coefficient for state, λ is the mean of the distribution, and α is the model dispersion.

Because one would reasonably expect the extent and severity of drought to be correlated with the volume of news coverage, we also explored the relationship between those two variables to make sure each was independently contributing to the model, and that they were not covariates. We compared prediction intervals with actual numbers of tweets, to see how well the model worked and to see whether differences in predicted and actual numbers of tweets reflected unusual events or individual experience.

Applying a diversity statistic. For less populous states, to screen out undue influence by a small number of users, such as professionals promoting a workshop or a service, we considered diversity of users, with tweets from several different users being more indicative of grassroots interest than several tweets from a single user. To create a diversity statistic, we used the "vegan" R package (Oksanen et al. 2010). First, we tabulated the number of users whose tweets appeared in each state-week, and then computed the Shannon–Wiener diversity index, which is one of several diversity indices used in ecological analyses to describe proportional abundance of species (Morris et al. 2014) and actually derives from communication theory (Spellerberg and Fedor 2003). To account for the differences between states, we used the top quarter of the diversity statistic range for each state to identify state-weeks when a surge represented more tweeters than usual. We defined anomalous weeks as those with a higher number of tweets than the upper limit of the prediction interval with a diversity statistic in the top quarter of the state's range.

Quantitative and qualitative analysis of anomalous state-weeks. We compared anomalous weeks with change in drought status and with the proportion of tweets from agricultural producers, based on words in user descriptions, which are brief profiles that Twitter invites users to provide. To compute change in drought status, we calculated and then summed the change in DSCI over 1, 2, 3, and 4 weeks, both lagging and leading. We also performed content analysis, reading the tweets from the anomalous weeks in closer detail to determine whether consistent themes emerged. We compared content of #drought17 tweets, the most distinctive subset of tweets, with the general collection of tweets via the tidytext R package (Silge and Robinson 2016).

As a service to the drought monitoring community and to help foster discussion, we began producing interactive maps of the filtered tweets we collected, and sharing them with the expert observers on the U.S. Drought Monitor listserv (Fig. 1). We used a different color icon to code tweets using the #drought18 hashtag, because those tweets were more likely (though by no means guaranteed) to be from agricultural producers sharing original observations. Clicking on the icons enabled drought observers to read tweets and access URLs for associated photos.

Results

Main themes and temporal patterns. Simply looking at raw numbers over time of all the #drought tweets we collected, and not accounting for population, California interests tended to dominate, given that it is a large, populous state, much of which is semiarid (Fig. 2). California alone accounted for more than 40% of the tweets we collected.

California's multiyear drought ended in early 2017, but for the first few months of 2017, the nation's #drought tweeting was still mostly retrospective, and was about California's experience. The peak the week of 9 January 2017, was in response to heavy precipitation there,

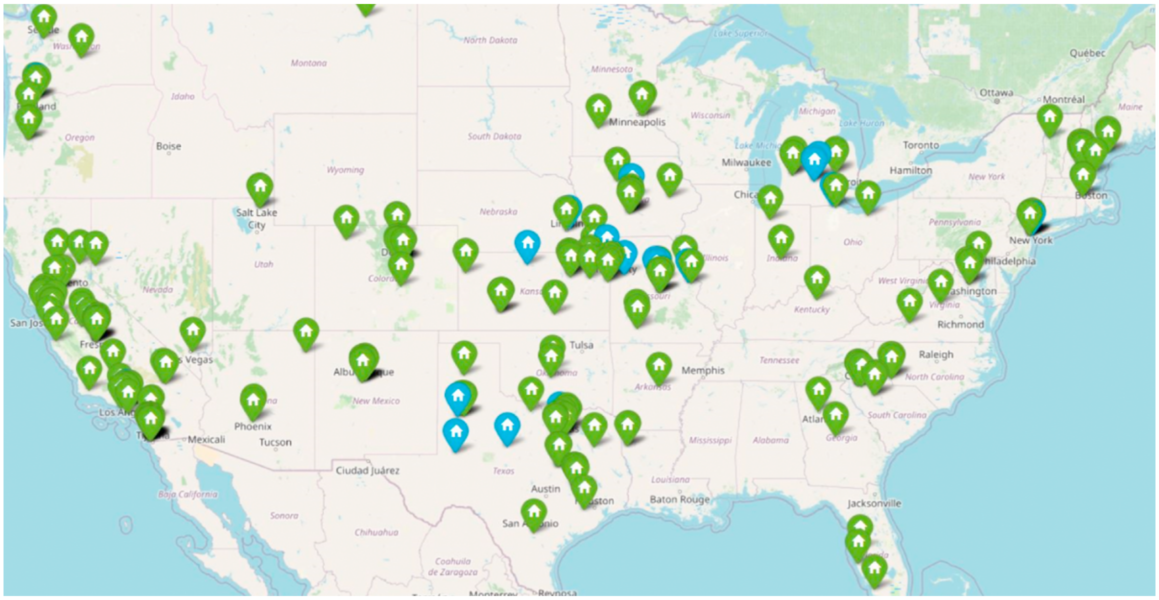

Fig. 1. Map of tweets from a week in late July 2018. This is a screen capture of one of the weekly maps of search results for #drought and related hashtags, distributed to the nation's drought monitoring community. Blue markers are tweets that used #drought18, a hashtag used by a higher proportion of agricultural producers than the rest of the tweets.

speculating about whether it meant drought was over. The spike in early April was the official declaration of the end of the drought in California. The peak the week of 29 January 2018 was driven by drought reemerging in central California, and the peak the week of 7 August 2018, by California's Ferguson wildfire. The midsummer swells in both years were more broadly representative, following a general pattern of more interest in drought during the growing season, boosted slightly in 2017 by agricultural producers, state climatologists, and others in the northern plains tweeting about #drought17. The #drought17 hashtag appeared to provide an outlet for farmers and ranchers concerned about emerging drought, and resulted in a distinct group of tweets that included a higher proportion of grassroots observations from agricultural producers, including photos of field conditions. In 2018, the #drought18 hashtag was less concentrated in space and time than #drought17, although as in the preceding year, the proportion of agricultural producers was higher for #drought18 tweets than for the search as a whole. For our entire collection of 18,926 tweets across both years, media accounted for 40%, and agricultural producers, about 13%. But of the 464 tweets that used the #drought17 hashtag, 61% identified themselves with agricultural production and 20%

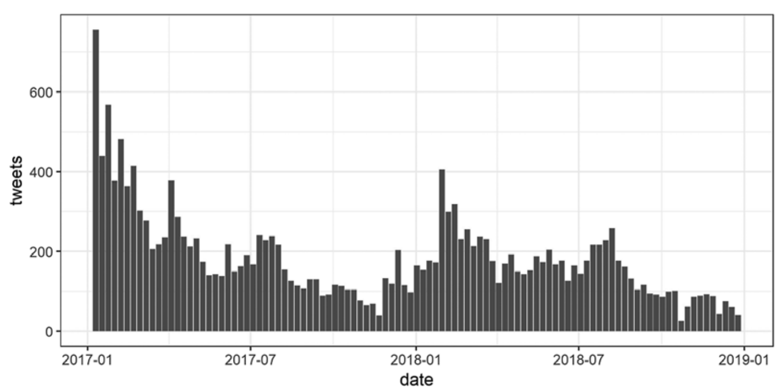

Fig. 2. The raw numbers of #drought tweets collected each week, 2017–18, for the nation as a whole. Tweets from California made up 41% of the total. The high weekly numbers early in 2017 reflected interest in California's ebbing drought and debate over how long it takes heavy rains to make a dent in long-term drought. In early January, a California user tweeted "If only there was a way to capture water, when we have too much, and save it until we have too little… #drought." In February, another said, "Would put Jerry Brown out to pasture, but it is under two feet of water. #CAdrought #jerrysdrought #fakegovernor #senilehappens #overthehill." The governor lifted the drought emergency in April, but the conversation on managing water scarcity continued: "@JerryBrownGov lifts #CAdrought emergency, retains prohibition on wasteful practices."

with media, and of the 258 that used the #drought18 hashtag, 48% identified with agricultural production and 15% with media. The #drought17 tweets comprised 5% of the total in 2017, and the #drought18 tweets, 3% in 2018.

The volume of tweets was highest on Thursdays, the day that the U.S. Drought Monitor is released and disseminated (Fig. 3), indicative of "top down" information, that is, dissemination of an official assessment.

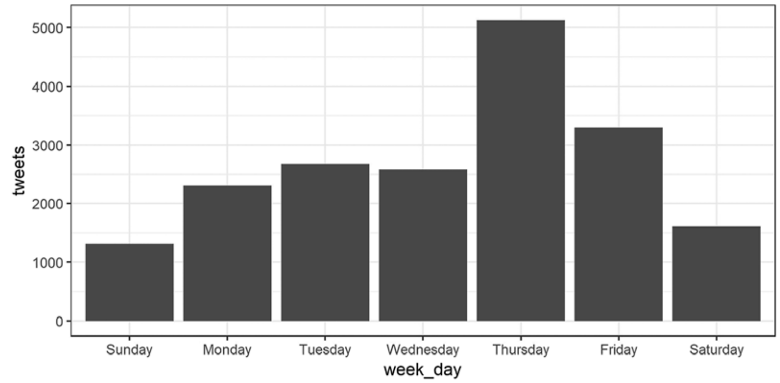

Fig. 3. Tweet frequency by day of week. More people tweeted about drought on Thursdays, the day that the U.S. Drought Monitor is released. This Thursday tweet from a TV meteorologist was typical: "Recent snow in extreme northwest Kansas has helped a little bit but drought status across the rest of the state continues."

Relationship between DSCI and newscount. Checking to see whether DSCI and newscount covaried during the study period, we found that the Pearson's *R* correlation between standardized DSCI and standardized newscount was overall 0.28, but it varied greatly by state, from a high of 0.72 for North Dakota to a low of −0.22 for Nevada. California, where much of the drought discussion was retrospective, had a −0.09 correlation between drought status and news coverage (Fig. 4). The inverse relationship for 2017–18 in California was almost certainly related to discussion of the just-ended drought. A similar comparison of standardized DSCI and standardized newscount for a longer period of time, 2011–18, found higher overall correlation, with a mean of 0.47, and individual state values ranging from zero in Alaska and West Virginia to 0.84 in Nebraska, with 0.73 in California.

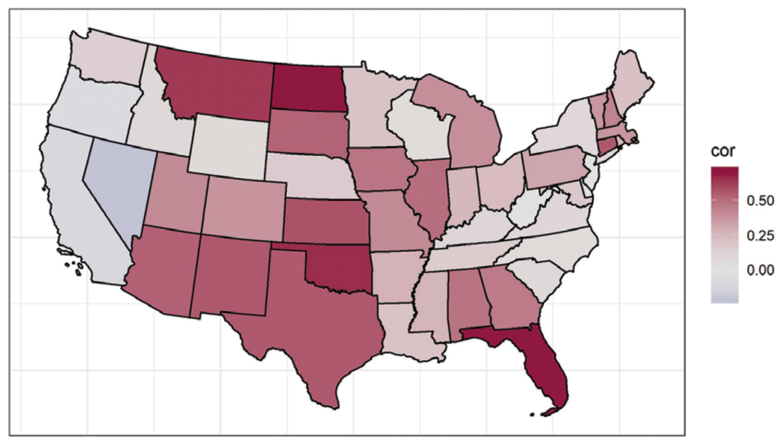

Fig. 4. Pearson's *R* correlation between DSCI and newscount. We compared DSCI, a measure of the intensity of U.S. Drought Monitor coverage for each state, with the number of news stories about drought in that state, each week during 2017–18, to see whether they covaried. We found sufficiently small covariance to include them both as predictors in our model. Negative values in California were related to ongoing discussion of the just-ended drought. Preliminary analysis over a longer time found greater correlation between drought and news coverage in many states, and inverse relationships disappeared.

Model fit. The PIG model was the better fit, with a dispersion statistic closer to 1, a lower Akaike's information criterion (AIC) score, and lower sum of squared residuals (Table 1). Comparing prediction intervals with

Table 1. Negative binomial and Poisson inverse Gaussian (PIG) model statistics. The PIG model was a better fit than the negative binomial model. The PIG model's dispersion statistic was closer to 1, its AIC statistic was lower, and the sum of squared residuals was lower. Its generalized *R* squared was 0.59, slightly better than the 56% of deviance explained by the negative binomial model.

Distribution	Dispersion statistic	AIC	Sum of squared residuals	How much it explains
Negative binomial	1.14	16,763.57	5939.69	Deviance explained = 56%
Poisson inverse Gaussian	1.04	16,682.79	5404.83	*R* squared = 0.59
Formula for both models: count ~ DSCI × newscount + state + offset[log(pop)]				

actual tweet counts, out of 5,253 state-week observations, 4,771 fell within the prediction interval, 79 had fewer tweets than predicted, and 403 had more tweets than predicted (Fig. 5).

Influence of state, news coverage, and drought status. Newscount and DSCI were each associated with an increase in the number of tweets and were statistically significant predictors (see Table 2). Their effect sizes were nearly identical. The sum-to-zero contrast coefficients also revealed that location mattered, with a regional pattern emerging in states' propensity to tweet about drought (see Fig. 6). Holding either DSCI or newscount constant revealed that the location coefficient had a bigger effect than either DSCI or newscount. Figure 7 shows the effect of newscount with DSCI held constant for select states.

Using the diversity statistic. The diversity statistic behaved differently in different states (Fig. 8). Alaska and Delaware each only had a handful of #drought tweets during our study period, so they never achieved any level of diversity. At the other end of the spectrum, California's diversity statistic was never lower than 2.3, the only state to have a minimum weekly diversity statistic greater than zero.

Exploring anomalous state-weeks. Of the 79 state-weeks with overpredictions, 39 were for California, 37 of

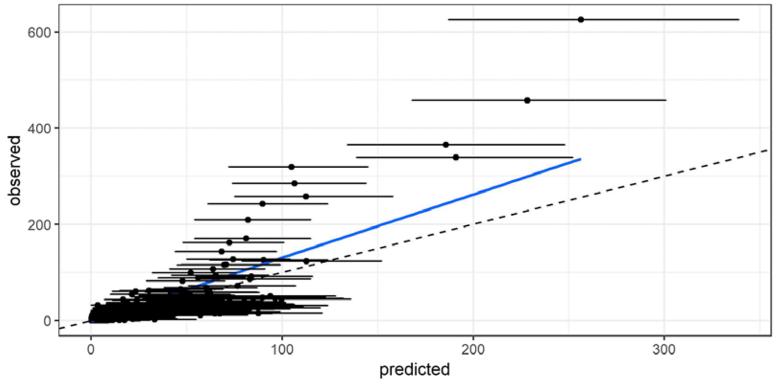

Fig. 5. Comparison of observed to predicted numbers of tweets. The best-fit line comparing the relationship of actual counts to counts predicted by our model had an intercept of −1.02, a slope of 1.32, and a *t* statistic of 111, with 5,251 degrees of freedom, for which the probability was one, finding there was not a significant difference between predicted and observed values, indicating a well-fit model. The blue line is the ratio of actual observed tweets to the number of tweets predicted by our model, and the dotted line is where it would be if the ratio were 1:1.

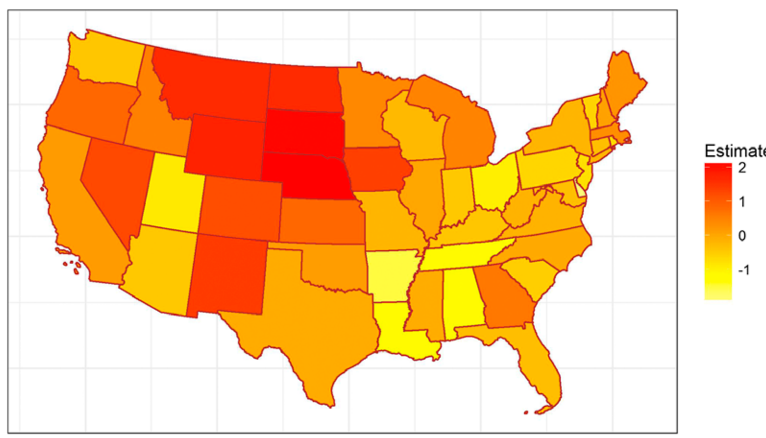

Fig. 6. State contrast coefficients showing propensity to tweet. The sum-to-zero contrast coefficients in our model revealed that some states had a higher propensity to tweet about drought. Most, though not all, state coefficients were significant at the 0.0001 level. New Mexico's coefficient was highest, at 2.07, followed by plains states and California.

Table 2. Coefficients of best-fit model. Newscount and DSCI (both standardized) were each associated with an increase in the number of tweets, and were statistically significant predictors, with probability < 0.001. Their effect sizes, in the "Estimate" column, were nearly identical. An interaction between newscount and DSCI was small, negative, and statistically significant, with a coefficient of −0.03 and probability < 0.05.

Coefficient	Estimate	Standard error	*t* value	Probability
(Intercept)	−16.6268	0.454221	−36.606	$<2 \times 10^{-16}$
Standardized DSCI	0.324263	0.015275	20.617	$<2 \times 10^{-16}$
Standardized newscount	0.293248	0.016	18.342	$<2 \times 10^{-16}$
Interaction between standardized DSCI and standardized newscount	−0.026221	0.011237	−2.305	0.0212208

which were in 2018. Of the 403 underpredictions, 30 were in California, 18 in Texas, and 15 in Florida. We identified 324 anomalous state-weeks with higher-than-expected volume of tweets that also had a diversity statistic above the 75th quantile. For the anomalous state-weeks, the mean proportion of tweets from agricultural producers was 12%, in contrast with 5% from the nonanomalous weeks.

To see how changes in the depiction on the U.S. Drought Monitor related to tweet counts and anomalous weeks, we summed the differences over the past 4 weeks (lagged) and over the next 4 weeks (leading). A Kolmogorov–Smirnov test determined that the difference between anomalous and nonanomalous weeks was statistically significant, more so with lagged than with leading changes (Table 3).

Anomalous weeks, those with bars taller than the prediction interval, with blue indicating they were over the diversity threshold, were more strongly associated with changes in the past four weeks of DSCI than in the upcoming four weeks, although Montana was a visible exception, with a surge in 2017 preceding an increase in the number of predicted tweets on the time series plot (Fig. 9). Tweets from summer 2017 in Montana focused on conditions affecting crop and cattle producers, and used the #drought17 hashtag. Of the 8 weeks from 20 June to 8 August 2017, in Montana, 6 met our definition of anomalous, and the proportion of agricultural producers ranged from none to 30%. For those 8 weeks, the mean change in lagged DSCI was 189.7, and in leading DSCI was 204.4. As one rancher

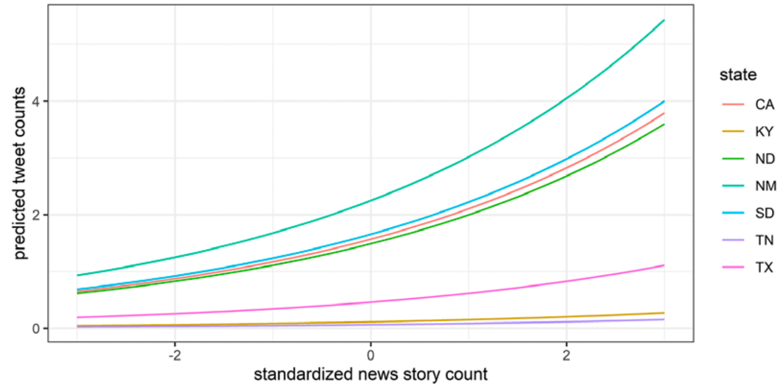

Fig. 7. Effect of news coverage on number of tweets for select states. Holding DSCI (U.S. Drought Monitor coverage) constant and plotting the effect of newscount, the variable reflecting news coverage, reveals that the state coefficient (see Fig. 6) is more influential than newscount. In other words, states' propensity to tweet about drought mattered more than how much news media were reporting on drought. New Mexico's state coefficient was highest, and Tennessee's is low.

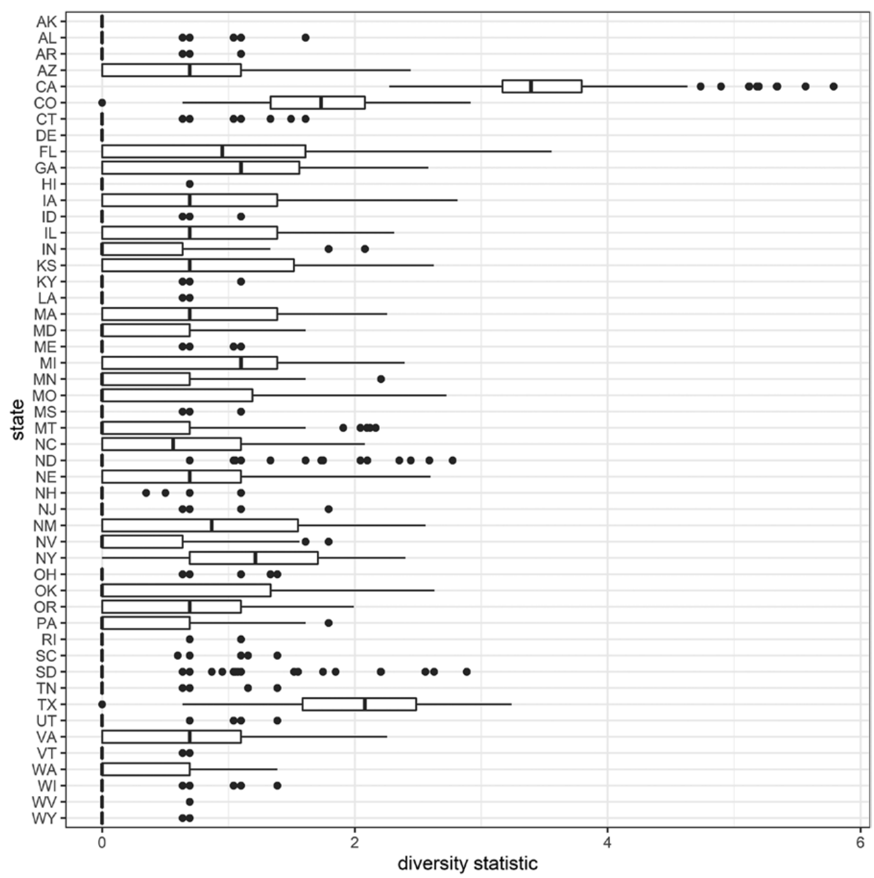

Fig. 8. Shannon–Weiner diversity statistic applied to numbers of tweeters by state. We applied the Shannon–Weiner diversity statistic as a way to help reduce the number of anomalous weeks. Particularly in less populous states, a single user's promotion of an idea or event could skew the tweet count higher. This figure contrasts the diversity statistics for each state. We considered state-weeks anomalous if the tweet count was higher than the top of our model's prediction interval and if the diversity statistic was in the highest quartile of diversity statistics for the state.

Table 3. Kolmogorov–Smirnov test applied to lagging and leading DSCI change. To see how changes in the depiction on the U.S. Drought Monitor related to anomalous weeks, we summed the differences over the past 4 weeks (lagged) and over the next 4 weeks (leading) for each state-week, and grouped them by whether or not they were anomalous. A Kolmogorov–Smirnov test determined that the differences in the distributions of DSCI changes between anomalous and nonanomalous weeks were statistically significant, more so with lagged ($D = 0.2225$, $p = 1.03 \times 10^{-12}$) than leading ($D = 0.100262$, $p = 0.004$) changes.

Anomalous state-week	Mean DSCI lead change	Max DSCI lead change	Min DSCI lead change	Mean DSCI lag change	Max DSCI lag change	Min DSCI lag change	Mean proportion producers
No	−5.227203	524.68	−700	−6.465264	423.97	−700	0.0492096297
Yes	−14.3777	342.53	−758.48	16.500207	514.9	−631.13	0.1202150223

K–S test	DSCI lead	DSCI lag
Difference	0.100262	0.222255
Probability	0.004	1.03×10^{-12}

tweeted, "I looked at my neighbor's winter wheat today. None of it filled. Making hay out of it as we speak. #drought17."

Instances of states with several consecutive or nearly consecutive weeks of higher-than-expected tweets revealed consistent themes across time from different users. They described conditions and key concerns.

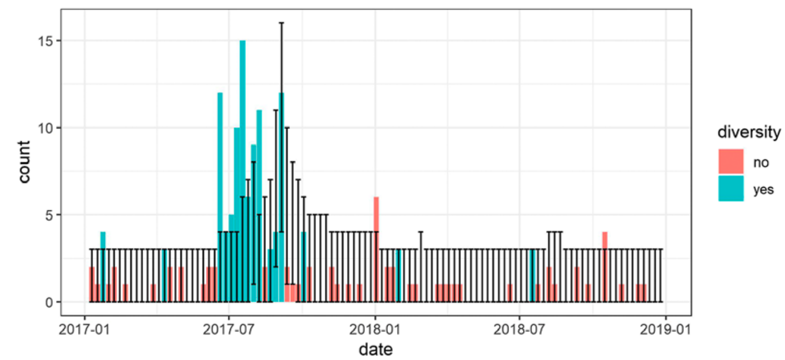

Fig. 9. Montana #drought tweets, 2017–18. Tweets from summer 2017 in Montana focused on conditions affecting crop and cattle producers, and used the #drought17 hashtag. For example, "Barely made it to "stubble-high by the 4th of July" ugh... #mtag #drought17 #gluten" and "This summer is going to test my mental and emotional strength. I love farming and ranching but damn this #drought17 is making it tough."

California tweets from anomalous weeks early in 2017 mentioned policy and sustainability, as well as immediate conditions (Fig. 10). Going into 2017, California was emerging from a multiyear drought. Tweets collected from the week beginning 10 January through the week beginning 2 May mentioned heavy rains, snowfall, flooding, mudslides, reservoirs filling, the Oroville Dam spillway collapse, and the recovery of hydropower production. Impacts mentioned included groundwater depletion; land subsidence; equity issues (rural, disadvantaged, tribal); tree die-off; crops not planted; dry wells; West Nile Virus; Valley Fever; and ecosystem and habitat damage. The discrepancy between heavy precipitation and ongoing official drought status and conservation requirements prompted discussion about when and whether water conservation should end. The governor officially declared an end to drought in April. Tweets mentioned water rates, the need to manage water sustainably, age of and investment in infrastructure, and desalination. Workshops were held for ranchers. As one agency noted, "CA had a great winter but the drought has left an indelible mark on our water use psyche." As time went on, California tweets questioned whether it was appropriate to stop talking about drought in a state that needs to reconcile its long-term water use with the possibility of a hotter, drier future, and some tweets that mentioned drought did so in context of recent wildfire or flood events.

The number of California drought tweets again exceeded predictions in early 2018, when drought reemerged, at one point affecting nearly half of the state. But late winter storms—a "March miracle"—brought heavy snows and eased concerns. Perhaps reflecting the highly urban population, the proportion of agriculture-oriented users only reached 11% for any of the underpredicted weeks. The lower number of tweets in 2018 was consistent with the issue having lost urgency, or simmering down, though it was still a much higher volume of tweets than from the rest of the country.

Floridians (Fig. 11) were more focused on a range of impacts, with many tweets mentioning fire and one mentioning the prospect of increased human–alligator contact. New Mexico in early 2018 was preoccupied with a very dry start to the water year and high fire danger, with a high proportion of tweets from media and no tweets in those weeks from agricultural producers, based on our coding of user profiles.

In contrast, agricultural producers in the plains employed the #drought17 and to some extent #drought18 hashtag to share accounts of their experiences. Higher proportions of tweets using variations of the year-specific hashtags were from agricultural producers, compared with all of the #drought tweets collected. For example, in the consecutive anomalous weeks from 27 June through 18 August 2017, the proportion of tweets from North Dakota producers ranged from 22% to 67%. Word use analysis found that tweets using #drought17 or #drought18 were more likely than the general body of tweets to refer to specific crops, such as corn, wheat, oats, cotton, or beans; to "cows;" to "rain" rather than "water" or "rainfall;" and to "day" rather than "week." For example, a South Dakota producer tweeted "Early listing of bred cows and pairs in salebarn this week in the drought area of South Dakota? Over 4000 #drought17." The #drought17 and #drought18 tweets were also more likely to use the words "bad," "burning," "hard," "toll," and "shit." Another South Dakota producer tweeted, "#drought17 #rayofhope Planted April 29 full cover wheat. Looks crappy but holding on made a

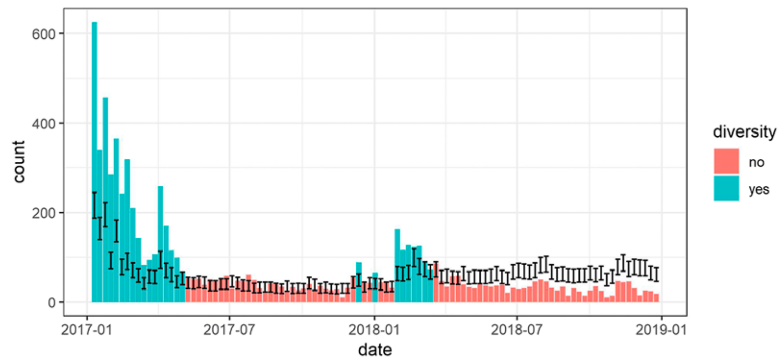

Fig. 10. California #drought tweets, 2017–18. Going into 2017, California was emerging from a multiyear drought, and Twitter reflected more interest in drought than our model predicted. Tweets collected from the week beginning 10 Jan through the week beginning 2 May mentioned heavy rains, snowfall, flooding, mudslides, reservoirs filling, the Oroville Dam spillway collapse, and the recovery of hydropower production. Impacts mentioned included groundwater depletion, land subsidence, equity issues (rural, disadvantaged, tribal), tree die-off, crops not planted, dry wells, West Nile Virus, Valley Fever, and ecosystem and habitat damage. The discrepancy between heavy precipitation and ongoing official drought status and conservation requirements prompted discussion about when and whether water conservation should end. The governor officially declared an end to drought in April. Tweets mentioned water rates, the need to manage water sustainably, age of and investment in infrastructure, and desalination. Workshops were held for ranchers. As one agency noted, "CA had a great winter but the drought has left an indelible mark on our water use psyche." The number of drought tweets again exceeded predictions in early 2018, when drought reemerged, at one point affecting nearly half of the state. But late winter storms—a "March miracle"—brought heavy snows and eased concerns. Perhaps reflecting the highly urban population, the proportion of agriculture-oriented users only reached 11% for any of the underpredicted weeks.

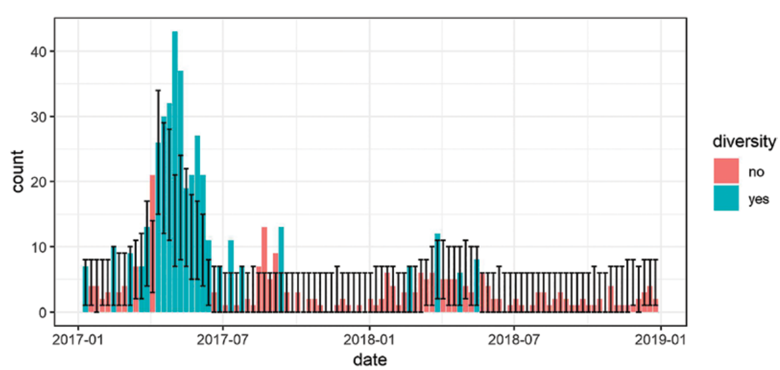

Fig. 11. Florida #drought tweets, 2017–18. Tweets collected in Florida from mid-April through mid-June 2017 reflected a variety of drought concerns in Florida, until heavy rains eased conditions. Tweets mentioned fire, water shortages, conservation, a drier cypress swamp, low lake levels, higher salinity, lawn maintenance tips, and more work for cattle producers, as well as a uniquely Floridian concern: "#drought and #matingseason bring alligators into close proximity with humans." The weekly proportion of agricultural tweets topped out at 11%.

dirt ball at 4in depth." The larger collection of tweets was more likely to include "climate," "California," and "conditions" (Fig. 12).

Isolated anomalous weeks reflected less organized sets of concerns. In early 2017, Nevada #drought Twitter celebrated an unusually abundant start to the water year ("Starting to run out of superlatives. Record setting start for water year and January 2017 #cawx #nvwx #drought"). In some instances, investigating anomalous state-weeks turned up a few hitherto-undiscovered spammers or commercially driven streams (such as content aggregation bots, or people selling diet aids, rain barrels, and efficient appliances or chronically prolific professionals (usually journalists or meteorologists) that were skewing results. Perhaps reflecting its cosmopolitan nature or simply the volume of users, anomalous weeks in New York also had a variety of tweets that made it through filters to reflect a variety of national and global concerns, rather than local drought ("Severe #wildfires spread in western states during unprecedented #drought" or "#drought in #Tuscany").

Discussion

We built on and confirmed previous research that both natural hazard events and news about natural hazard events would explain some but not all of the number of tweets about the hazard. Consistent with other researchers' findings, news coverage (Broniatowski et al. 2013; Gruszczynski and Wagner 2017; Kirilenko et al. 2015; McCombs and Shaw 1972; Ripberger et al. 2014; Searles and Smith 2016), and drought status (Broniatowski et al. 2013; Kirilenko et al. 2015; Switzer and Vedlitz 2017) were each statistically significant predictors of #drought tweets. We also determined that anomalous weeks were of interest, reflecting heightened interest due to emerging or recent drought, and with a higher proportion of tweets from agricultural producers, who tended to share more original content based on personal experience. This is consistent with the finding that natural disasters or issues related to public order tended to appear more first in social media and then be taken up by traditional media (Neuman et al. 2014). Besides

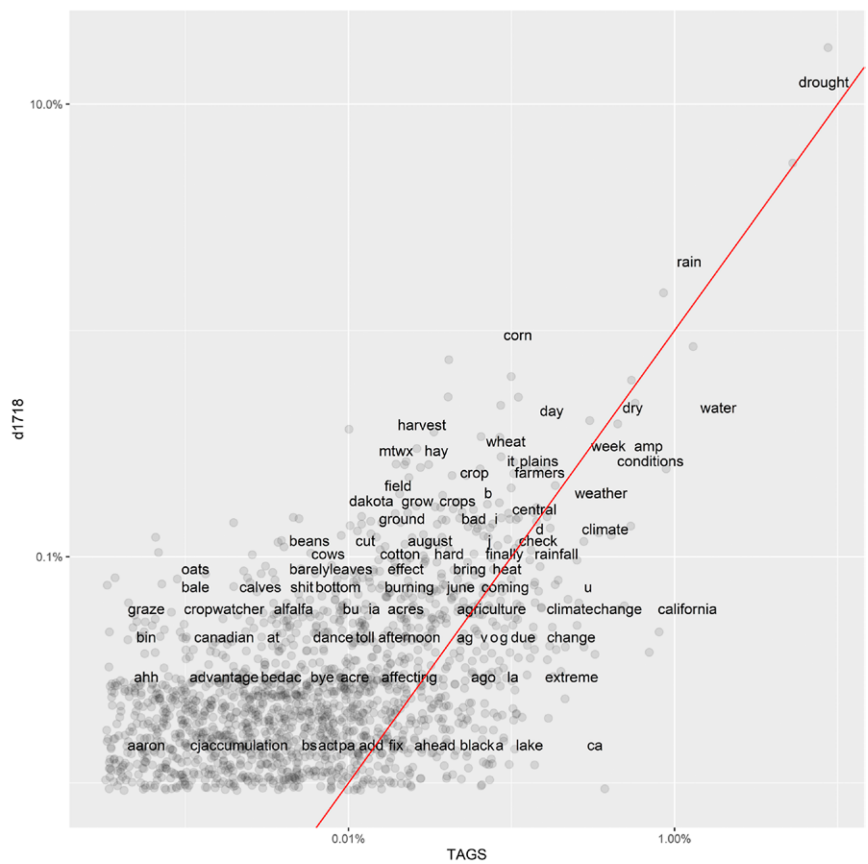

Fig. 12. Most characteristic words from #drought17 tweets vs all tweets. This figure, generated via the tidytext R package, contrasts word frequencies from the subset of tweets using variations of #drought17 or #drought18 hashtags with the rest of the tweets we collected with the Twitter Archiving Google Spreadsheet (TAGS). Farmers and ranchers, particularly in the plains, adopted #drought17 as a way to describe their experiences and field conditions, for example: "I think you guys are getting the exact same #drought17 we are. Big rain promises but nothing happens." Many incorporated humor, such as "Thankful for grape farmers, they help me forget about the impending #drought17 #itsonlyiowa." This image contrasts frequencies of words that appear in both sets of tweets, displaying words that appear in a similar proportion of tweets near the red 1:1 line, and words that appear proportionately more frequently in one set or the other at greater distance from the red line.

being characterized by higher numbers of tweets, many of these weeks included higher proportions of tweets from agricultural producers, and were part of time periods undergoing intensification, as measured by the Kolmogorov–Smirnov comparison of DSCI distributions.

The state and regional variation we found in propensity to tweet, shown in the consistently higher coefficients for plains states, New Mexico, and California, was an unexpected and informative finding. It appears to reflect real differences in awareness of drought, likely related to recent and historic experience, which would be consistent with the development of long-term awareness that Kam et al. (2019) identified. The predicted range of tweet volume, or the baseline for each state, was analogous to epidemiologists' baseline understanding of disease rates (Hess et al. 2014). Observed tweet volumes higher than this baseline, particularly those higher than would be predicted including news and official drought status, were analogous to the surges or spikes in attention that emergency managers use to identify events of interest (Abdelhaq et al. 2013; Cameron et al. 2012; Fitzhugh 2015; Mathioudakis and Koudas 2010; Sakaki et al. 2010).

Our use of a diversity statistic to filter surges of interest is an addition to methods used in the literature surveyed above. It is a means of distinguishing signal from noise, reducing the time and effort needed to investigate surges of interest by distinguishing surges based on widespread interest rather than on a single publicity campaign.

Isolated anomalous weeks were difficult to interpret, and taken in aggregate, anomalous weeks did not anticipate emerging drought. However, paying attention to anomalous weeks in real time, catching the leading edge by identifying two or three anomalous weeks in a row when they first appear within a state, can help identify or confirm an emerging drought, as in Missouri in 2018 (Fig. 13) or Montana or North Dakota in 2017 (Fig. 14). On the qualitative side, those familiar with local conditions, such as state climatologists, may be able to glean new information from the content of tweets by agricultural producers or others who are sharing personal experiences. We have not done a formal evaluation of use of the maps shared with the U.S. Drought Monitor listserv each week (Fig. 1), but experience suggests that the maps are of interest to state climatologists, weather service employees, and others who gather information about conditions within states, and that more ways to distinguish original, grassroots observations from top-down sharing of official assessments would make the maps more useful.

Our use of "#drought" and related hashtags successfully limited our results to a manageable size and tuned in to a structured, somewhat official conversation. Although drought is quite real to farmers eyeing dusty fields, it is still an abstraction, referring to the difference between expectation and reality. Many tweets may relate to dry conditions, on farms and ranches and in other contexts, without using "drought" or "#drought." To hone in on emerging conditions that people may not be hashtagging, i.e., to get past searches for "#drought" or "#drought19," a next step would to explore less visible conversations, possibly starting with search terms such as the production-oriented words that our analysis identified as being distinctive in the #drought17 subset of tweets, or to look for a drought signal in tweets related to planting progress (Zipper 2018). Another possibility would be to filter larger searches based on user profile information, either with comparatively simple filters such as use of the word "ranch," or using natural language processing to develop more refined indicators of agricultural production. The greater volume of tweets that could result from searching beyond hashtags could also allow for publicly shareable data visualizations, with aggregation protecting individuals' privacy and end users' sensibilities.

The state and regional variation we found in propensity to tweet, shown in the consistently higher coefficients for plains states, New Mexico, and California, was an unexpected and informative finding. It appears to reflect real differences in awareness of drought, likely related to recent and historic experience. California, coming off a multiyear drought, was still having a largely retrospective conversation, some of which included speculation and debate about whether it was appropriate to stop talking about drought, when the need to manage water supply is constant. Drought and the resulting well-documented declines in groundwater appear to have tipped the balance in favor of passage of California's Sustainable Groundwater Management Act in 2014 (Leahy 2015). New Mexico has been in some degree of drought on the U.S. Drought Monitor nearly continuously for the past 20 years, so it could also be expected to be highly drought-aware. The plains states experienced growing season drought in 2017, and are part of an arid region where agriculture is economically dominant. States in the lower Mississippi and Ohio River valleys were least likely to tweet about drought, and have in many cases experienced less drought in the past 20 years than states farther west. Based on the idea that awareness is a precursor for action, one could infer that the plains states, along with New Mexico, may be favorable locales for innovations or advances in drought planning and policy. It is already occurring in California.

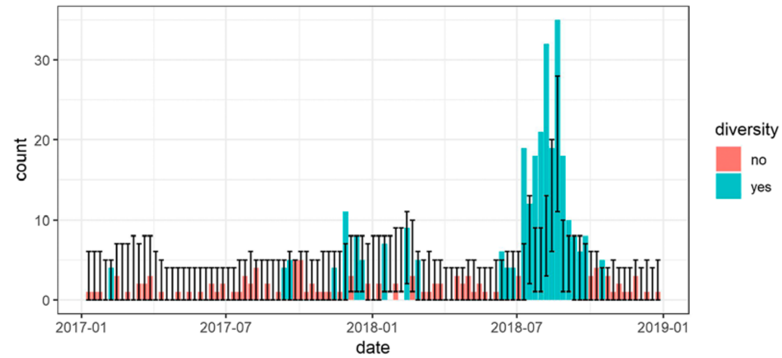

Fig. 13. Missouri #drought tweets, 2017–18. Of the 9 weeks from 10 Jul through 4 Sep 2018, in Missouri, 7 were anomalous. The mean change in 4-week DSCI lags for those weeks was 87.44, and the mean change in 4-week DSCI leading values was −11.3. As many as a quarter of the tweets in some weeks were identified with agricultural production. Drought affected corn and other crops, and dried ponds and pastures. The state made hay and water available from state parks. Tweet content focused on the hardships that farmers and ranchers were enduring ("None, zero, zip - that is how much rain we had at our house last night") as well as official response ("Breaking news: @GovParsonMO hosts drought press conference to announce water hauling & haying on state lands"). Rains began in August, and by October, the state had seen substantial improvement.

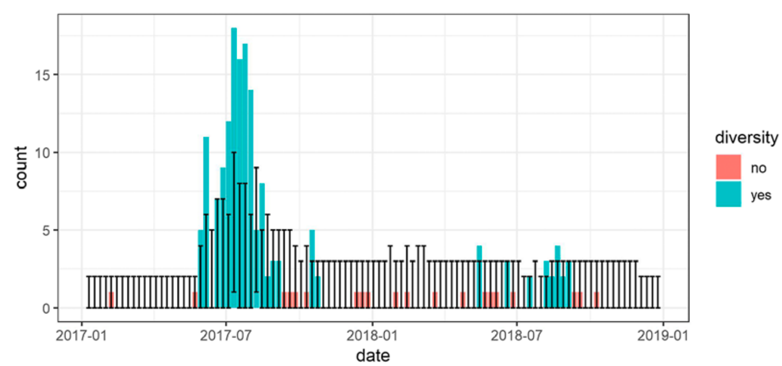

Fig. 14. North Dakota #drought tweets, 2017–18. North Dakota was one of the plains states in 2017 where farmers and ranchers used the #drought17 hashtag to describe their experiences. North Dakota tweets mentioned drought status, and responses to drought such as emergency grazing, hotlines, and fireworks bans. Impacts mentioned included water quality for ranchers, early weaning calves, crunchy hay fields, dust pneumonia, and concerns about progress or yield of several crops: sugar beets, soybeans, corn, milo, and barley. As one farmer said, "These soybeans so thirsty they text me 'you up?' at 2:30 a.m. every night." The proportion of tweets from agricultural producers in anomalous weeks from 27 Jun through 15 Aug 2017, ranged from 22% to 67%, and the mean change in lagged DSCI for those 8 weeks was 89.5, and in leading DSCI, −6.8.

Our research applies previously documented relationships between disasters, news media, and social media to drought, determines normal or baseline rates of tweeting for each state, and uses the understanding of normal to detect unexpectedly high rates of tweets, which can then be further described statistically, looking at considerations such as the proportion of users who describe themselves as agricultural producers, and analyzed qualitatively or quantitatively for themes and issues. Tweets may also serve as a cross-sector metric of drought impacts, serving as a quantitative scan for intensity of interest, and including qualitative information about specific experiences. Real-time implementation of this method of analysis would contribute a cross-sector, quantifiable, impact-based metric to drought monitoring. We anticipate that the NDMC will implement this process in real-time in 2020, with anomalous state-weeks identified for further investigation accompanying the interactive tweet maps that the drought monitoring community already receives every week.

Conclusions

The #drought search and analysis method we used suggests that tweets can be another source of data used to detect or confirm human experience of drought. Just as no single hydrometeorological indicator is considered sufficient to capture all aspects of drought, #drought tweets are one more metric to consider, and they represent a real addition to quantifiable drought impact data. Drought tweets reflect needs and interests identified by agencies and organizations involved in water and drought management, as well as on-the-ground experiences of agricultural producers and others whose lives and livelihoods are affected by drought. Tweets are a measurement of drought impact, even when the impact is primarily an awareness of a problem that may require attention. Our findings suggest that expanded consideration of social media and big data as a source of meaningful data for comparison with hydrometeorological drought indices would be fruitful. Content analysis of tweets, which could be an initial statistical scan of words used, provides insight on what type of impacts people are experiencing and can help identify new impact experiences.

Acknowledgments. This project was partially supported by 2017–18 and 2018–19 cooperative agreements between the National Drought Mitigation Center and the U.S. Department of Agriculture. None of the authors has a financial conflict of interest with this project.

References

Abdelhaq, H., C. Sengstock, and M. Gertz, 2013: EvenTweet. *Proc. VLDB Endowment*, **6**, 1326–1329, https://doi.org/10.14778/2536274.2536307.

AgChat Foundation, 2019: Foundation Q&A. Accessed 15 August 2019, https://agchat.org/who-we-are/foundation-qa/.

Akyuz, F. A., 2017: Drought severity and coverage index. U.S. Drought Monitor, National Drought Mitigation Center, accessed 22 November 2019, https://droughtmonitor.unl.edu/About/AbouttheData/DSCI.aspx.

Allaire, M. C., 2016: Disaster loss and social media: Can online information increase flood resilience? *Water Resour. Res.*, **52**, 7408–7423, https://doi.org/10.1002/2016WR019243.

Broniatowski, D. A., M. J. Paul, and M. Dredze, 2013: National and local influenza surveillance through twitter: An analysis of the 2012-2013 influenza epidemic. *PLOS ONE*, **8**, e83672, https://doi.org/10.1371/journal.pone.0083672.

Burgess, J., A. Galloway, and T. Sauter, 2015: Hashtag as hybrid forum: The case of #agchatoz. *Hashtag Publics: The Power and Politics of Networked Discourse*, N. Rambukkana, Ed., Peter Lang, 61–76.

Cameron, M. A., R. Power, B. Robinson, and J. Yin, 2012: Emergency situation awareness from twitter for crisis management. *Proc. 21st Int. Conf. on World Wide Web*, Lyon, France, ACM, 695–698.

Chen, G., and L. Xiao, 2016: Selecting publication keywords for domain analysis in bibliometrics: A comparison of three methods. *J. Informetrics*, **10**, 212–223, https://doi.org/10.1016/j.joi.2016.01.006.

Chunara, R., J. R. Andrews, and J. S. Brownstein, 2012: Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian Cholera outbreak. *Amer. J. Trop. Med. Hyg.*, **86**, 39–45, https://doi.org/10.4269/ajtmh.2012.11-0597.

Deng, Q., Y. Liu, H. Zhang, X. Deng, and Y. Ma, 2016: A new crowdsourcing model to assess disaster using microblog data in Typhoon Haiyan. *Nat. Hazards*, **84**, 1241–1256, https://doi.org/10.1007/s11069-016-2484-9.

Eilander, D., P. Trambauer, J. Wagemaker, and A. van Loenen, 2016: Harvesting social media for generation of near real-time flood maps. *Procedia Eng.*, **154**, 176–183, https://doi.org/10.1016/j.proeng.2016.07.441.

Fiesler, C., and N. Proferes, 2018: "Participant" perceptions of twitter research ethics. *Soc. Media Soc.*, **4**, 1–14, https://doi.org/10.1177/2056305118763366.

Fitzhugh, S., 2015: The spacetime continuum: Using spatio-temporal filtering techniques to revisit rumor theory. Ph.D. dissertation, University of California-Irvine, 109 pp.

Ghermandi, A., and M. Sinclair, 2019: Passive crowdsourcing of social media in environmental research: A systematic map. *Global Environ. Change*, **55**, 36–47, https://doi.org/10.1016/j.gloenvcha.2019.02.003.

Goodchild, M. F., 2007: Citizens as sensors: The world of volunteered geography. *GeoJournal*, **69**, 211–221, https://doi.org/10.1007/s10708-007-9111-y.

Gruszczynski, M., and M. W. Wagner, 2017: Information flow in the 21st century: The dynamics of agenda-uptake. *Mass Commun. Soc.*, **20**, 378–402, https://doi.org/10.1080/15205436.2016.1255757.

Guan, X., and C. Chen, 2014: Using social media data to understand and assess disasters. *Nat. Hazards*, **74**, 837–850, https://doi.org/10.1007/s11069-014-1217-1.

Hawksey, M., 2014: Need a better Twitter Archiving Google Sheet? TAGS v6.0 is here! MASHe, accessed 29 December 2016, https://mashe.hawksey.info/2014/10/need-a-better-twitter-archiving-google-sheet-tags-v6-0-is-here/.

Hayes, M., M. Svoboda, N. Wall, and M. Widhalm, 2011: The Lincoln declaration on drought indices: Universal meteorological drought index recommended. *Bull. Amer. Meteor. Soc.*, **92**, 485–488, https://doi.org/10.1175/2010BAMS3103.1.

Hecht, B., L. Hong, B. Suh, and E. H. Chi, 2011: Tweets from Justin Bieber's heart: The dynamics of the location field in user profiles. *Proc. 2011 Annual Conf. on Human Factors in Computing Systems*, New York, New York, ACM, 237–246.

Hess, J. J., S. Saha, P. J. Schramm, K. C. Conlon, C. K. Uejio, and G. Luber, 2014: Projecting climate-related disease burden: A guide for health departments. National Center for Environmental Health, Centers for Disease Control and Prevention, www.cdc.gov/climateandhealth/pubs/projectingclimaterelateddiseaseburden1_508.pdf.

Hilbe, J. M., 2014: *Modeling Count Data*. Cambridge University Press, 283 pp.

Jongman, B., J. Wagemaker, B. Romero, and E. de Perez, 2015: Early flood detection for rapid humanitarian response: Harnessing near real-time satellite and twitter signals. *ISPRS Int. J. Geo-Inf.*, **4**, 2246–2266, https://doi.org/10.3390/ijgi4042246.

Jung, J., and C. K. Uejio, 2017: Social media responses to heat waves. *Int. J. Biometeorol.*, **61**, 1247–1260, https://doi.org/10.1007/s00484-016-1302-0.

Kallis, G., 2008: Droughts. *Annu. Rev. Environ. Resour.*, **33**, 85–118, https://doi.org/10.1146/annurev.environ.33.081307.123117.

Kam, J., K. Stowers, and S. Kim, 2019: Monitoring of drought awareness from Google Trends: A case study of the 2011–17 California drought. *Wea. Climate Soc.*, **11**, 419–429, https://doi.org/10.1175/WCAS-D-18-0085.1.

Kearney, M. W., 2019: rtweet: Collecting Twitter data. R Package, accessed 22 September 2019, https://cran.r-project.org/package=rtweet.

Kirilenko, A. P., T. Molodtsova, and S. O. Stepchenkova, 2015: People as sensors: Mass media and local temperature influence climate change discussion on Twitter. *Global Environ. Change*, **30**, 92–100, https://doi.org/10.1016/j.gloenvcha.2014.11.003.

Kryvasheyeu, Y., H. Chen, N. Obradovich, E. Moro, P. Van Hentenryck, J. Fowler, and M. Cebrian, 2016: Rapid assessment of disaster damage using social media activity. *Sci. Adv.*, **2**, e1500779, https://doi.org/10.1126/sciadv.1500779.

Kumar, A., M. Jiang, and Y. Fang, 2014: Where not to go? Detecting road hazards using twitter. *Proc. 37th Int. ACM SIGIR Conf. on Research & Development in Information Retrieval*, Gold Coast, QLD, Australia, ACM, 1223–1226.

Lackstrom, K., and Coauthors, 2013: The missing piece: Drought impacts monitoring. Rep. from a workshop in Tucson, AZ, 23 pp., http://cisa.sc.edu/PDFs/Drought_Impacts_Report_June2013_final.pdf.

——, A. Farris, D. Eckhardt, N. Doesken, H. Reges, J. Turner, K. H. Smith, and R. Ward, 2017: CoCoRaHS observers contribute to "condition monitoring" in the carolinas: A new initiative addresses needs for drought impacts information. *Bull. Amer. Meteor. Soc.*, **98**, 2527–2531, https://doi.org/10.1175/BAMS-D-16-0306.1.

Leahy, T. C., 2015: Desperate times call for sensible measures: The making of the California Sustainable Groundwater Management Act. *Golden Gate Univ. Environ. Law J.*, **9**, 5–40.

Leetaru, K., S. Wang, A. Padmanabhan, and E. Shook, 2013: Mapping the global Twitter heartbeat: The geography of Twitter. *First Monday*, **18**, https://doi.org/10.5210/fm.v18i5.4366.

Mathioudakis, M., and N. Koudas, 2010: TwitterMonitor: Trend detection over the twitter stream. *Proc. 2010 ACM SIGMOD Int. Conf. on Management of Data*, Indianapolis, IN, ACM, 1155–1158.

McCombs, M. E., and D. L. Shaw, 1972: The agenda-setting function of mass media. *Public Opin. Quart.*, **36**, 176–187, https://doi.org/10.1086/267990.

Meadow, A. M., M. A. Crimmins, and D. B. Ferguson, 2013: Field of dreams or dream team? Assessing two models for drought impact reporting in the semiarid southwest. *Bull. Amer. Meteor. Soc.*, **94**, 1507–1517, https://doi.org/10.1175/BAMS-D-11-00168.1.

Morris, E. K., and Coauthors, 2014: Choosing and using diversity indices: Insights for ecological applications from the German biodiversity exploratories. *Ecol. Evol.*, **4**, 3514–3524, https://doi.org/10.1002/ece3.1155.

Neuman, W. R., L. Guggenheim, S. M. Jang, and S. Y. Bae, 2014: The dynamics of public attention: Agenda-setting theory meets big data. *J. Commun.*, **64**, 193–214, https://doi.org/10.1111/jcom.12088.

Oksanen, J., and Coauthors, 2010: Vegan: Community ecology package. R package, accessed 13 August 2019, https://cran.r-project.org/web/packages/vegan/index.html.

Otkin, J. A., T. Haigh, A. Mucia, M. C. Anderson, and C. Hain, 2018: Comparison of agricultural stakeholder survey results and drought monitoring datasets during the 2016 U.S. Northern Plains flash drought. *Wea. Climate Soc.*, **10**, 867–883, https://doi.org/10.1175/WCAS-D-18-0051.1.

Rigby, R. A., and D. M. Stasinopoulos, 2005: Generalized additive models for location, scale and shape. *J. Roy. Stat. Soc.*, **54**, 507–554, https://doi.org/10.1111/j.1467-9876.2005.00510.x.

Ripberger, J. T., H. C. Jenkins-Smith, C. L. Silva, D. E. Carlson, and M. Henderson, 2014: Social media and severe weather: Do tweets provide a valid indicator of public attention to severe weather risk communication? *Wea. Climate Soc.*, **6**, 520–530, https://doi.org/10.1175/WCAS-D-13-00028.1.

Sakaki, T., M. Okazaki, and Y. Matsuo, 2010: Earthquake shakes Twitter users: real-time event detection by social sensors. *Proc. 19th Int. Conf. on World Wide Web*, Raleigh, NC, ACM, 851–860.

Searles, K., and G. Smith, 2016: Who's the boss? Setting the agenda in a fragmented media environment. *Int. J. Commun.*, **10**, 2074–2095.

Silge, J., and D. Robinson, 2016: tidytext: Text mining and analysis using tidy data principles in R. *J. Open Source Software*, **1**, 37, https://doi.org/10.21105/joss.00037.

Silver, A., 2019: Public attention to risks, hazards, and disasters: A retrospective review and proposed conceptual model. *Risks Hazards Crisis Public Policy*, **10**, 294–310, https://doi.org/10.1002/rhc3.12165.

——, and J. Andrey, 2019: Public attention to extreme weather as reflected by social media activity. *J. Contingencies Crisis Manage.*, **27**, 346–358, https://doi.org/10.1111/1468-5973.12265.

Smith, K. H., M. Svoboda, M. Hayes, H. Reges, N. Doesken, K. Lackstrom, K. Dow, and A. Brennan, 2014: Local observers fill in the details on drought impact reporter maps. *Bull. Amer. Meteor. Soc.*, **95**, 1659–1662, https://doi.org/10.1175/1520-0477-95.11.1659.

Spellerberg, I. F., and P. J. Fedor, 2003: A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon-Wiener' index. *Global Ecol. Biogeogr.*, **12**, 177–179, https://doi.org/10.1046/j.1466-822X.2003.00015.x.

Stahl, K., and Coauthors, 2016: Impacts of European drought events: Insights from an international database of text-based reports. *Nat. Hazards Earth Syst. Sci.*, **16**, 801–819, https://doi.org/10.5194/nhessd-16-801-2016.

Svoboda, M., and Coauthors, 2002: The Drought Monitor. *Bull. Amer. Meteor. Soc.*, **83**, 1181–1190, https://doi.org/10.1175/1520-0477-83.8.1181.

Switzer, D., and A. Vedlitz, 2017: Investigating the determinants and effects of local drought awareness. *Wea. Climate Soc.*, **9**, 641–657, https://doi.org/10.1175/WCAS-D-16-0052.1.

Tang, Z., L. Zhang, F. Xu, and H. Vo, 2015: Examining the role of social media in California's drought risk management in 2014. *Nat. Hazards*, **79**, 171–193, https://doi.org/10.1007/s11069-015-1835-2.

U.S. Census Bureau, 2019: Annual estimates of the resident population for the United States, regions, states and Puerto Rico: April 1, 2010 to July 1, 2018 (NST-EST2018-01). www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html.

Van Loon, A. F., and Coauthors, 2016: Drought in a human-modified world: Reframing drought definitions, understanding, and analysis approaches. *Hydrol. Earth Syst. Sci.*, **20**, 3631–3650, https://doi.org/10.5194/hess-20-3631-2016.

Wood, S. N., 2017: *Generalized Additive Models: An Introduction With R*. 2nd ed. CRC Press, 476 pp.

Zipper, S. C., 2018: Agricultural research using social media data. *Agron. J.*, **110**, 349–358, https://doi.org/10.2134/agronj2017.08.0495.

——, and Coauthors, 2019: Balancing open science and data privacy in the water sciences. *Water Resour. Res.*, **55**, 5202–5211, https://doi.org/10.1029/2019WR025080.

Zou, L., N. S. N. Lam, H. Cai, and Y. Qiang, 2018: Mining Twitter data for improved understanding of disaster resilience. *Ann. Amer. Assoc. Geogr.*, **108**, 1422–1441, https://doi.org/10.1080/24694452.2017.1421897.