



Droughts and Media: when and what do the newspapers talk about the droughts in England?

Inhye Kong¹, Jan Seibert¹, Ross S. Purves¹

¹Department of Geography, University of Zurich, Zurich 8057, Switzerland

5 *Correspondence to:* Inhye Kong (inhye.kong@geo.uzh.ch)

Abstract. The UK is traditionally known for its wet climate, but droughts are also relatively commonplace. Using newspapers as a medium of public communication, this study explores the timing and content of media coverage of droughts in England. We constructed a corpus from newspaper articles related to droughts in England to analyse the temporal alignment of media coverage with meteorological anomalies and, using topic modelling, the emerging topics covered by drought-related articles in newspapers. Our findings reveal that media coverage generally coincides with meteorological drought, but the inverse case is not always the case, suggesting additional conditions to generate media coverage (e.g., seasonality, precedent condition of long-term precipitation shortage). Dominant topics include the status of water deficiency and weather forecasts, the mismanagement of water companies and the enforcement of hosepipe bans, highlighting current challenges in water management practices in England.

15 1 Introduction

The UK is widely known for its wet climate and frequent flooding, but droughts have been common throughout history with a major drought every 5-10 years on average (Dayrell et al., 2022; Marsh et al., 2007; Murphy et al., 2006). Within the last two decades, The National Hydrological Monitoring Programme of the UK has released major drought reports for 2003, 2004-2006, 2010-12, and 2018-19 as a part of unusual hydrological events (UKCEH, n.d.). Within the UK, England is more susceptible to droughts due to both less precipitation and greater water demand from human influences than Scotland, Wales or Northern Ireland (Tjideman et al., 2018). This pattern is especially distinctive around southern England, including the London metropolitan area (Dessai and Sims, 2010; Folland et al., 2015). Climate change is expected to heighten drought-prone meteorological conditions in Europe (Guerreiro et al., 2018; Spinoni et al., 2018), and England is no exception (Dobson et al., 2020; Kay et al., 2023).

25 Droughts, an example of extreme weather events, are periods with a long-term imbalance between rainfall and evapotranspiration (Wilhite & Glantz, 1985). Droughts can be further distinguished into detailed categories: while meteorological droughts are defined by a deficiency in precipitation, hydrological droughts are characterised by periods of low streamflow (i.e., shortfalls in runoff and aquifer recharge) and agricultural droughts indicate the period when the availability of soil water is limited especially during the growing season (Van Loon, 2015). Aside from hydroclimatic viewpoints of



30 droughts, socioeconomic droughts redirect attention to the impact of drought on society, including agricultural and economic loss and failure to meet water demands for people and the environment (Van Loon, 2015; Wilhite & Glantz, 1985).

Compared to other natural hazards, droughts receive less attention from the media and public (Brimicombe, 2022). This has been ascribed to the characteristics of droughts: compared to floods and wildfires that entail rapid onsets, droughts develop slowly, often spanning over long periods for months, years, and even decades as a ‘creeping phenomenon’. There are, therefore, 35 often temporal mismatches between meteorological, hydrological, agricultural, and socioeconomic droughts. For example, meteorological droughts do not always coincide with hydrological and agricultural droughts and only in a few cases lead to socioeconomic droughts. This is mainly due to the time lag effects on rainfall, streamflow, and groundwater recharge (Van Loon, 2015), and many cases of meteorological and hydrological drought do not impact the public. Still, media representation of droughts is mixed with different categories of droughts, and little is known about how the media responds to them, more 40 precisely when the media features the droughts. More critically, media coverage is not solely driven by events. Newspapers choose topics likely to interest or impact their readers directly, and other external factors may also influence the reporting of extreme weather, confirming the role of media in agenda setting and framing. Understanding media attention is important as media representation can affect public behaviours or alter public attitudes (Quesnel & Ajami, 2017).

To explore the extent of media coverage in response to extreme weather events, including droughts, previous studies assumed 45 the number of article publications to signal the degree of media and public attention (Brimicombe, 2022; Kim et al., 2019; Quesnel & Ajami, 2017). Recent studies have upscaled the volume and range of data from digital text archives encompassing historical and contemporary corpora and analysing links between media coverage and meteorological conditions. Dayrell et al. (2022) examined UK newspaper corpora over the last 200 years to explore the correlation between a drought index (i.e., SPI) and media coverage. O’Connor et al. (2022) examined the relationship between drought indices (i.e., SPI and SSI) and 50 historical newspaper articles (1900-2016) in Ireland and found the best predictor of drought-related articles to be SPI-3.

In addition to the volume of media coverage, media content has been widely explored in hazard studies to identify dominant topics, main actors and stakeholders, and emerging concerns. Again, newspapers are typically at the centre of analytic materials, having a large circulation size reaching different population strata (Boykoff, 2008). Moreover, newspapers are 55 penned by professional journalists and reporters, and are seen as a trustworthy medium for information dissemination, science communication, and legitimization of societal issues (Schmidt et al., 2013). Newspapers are thus typically seen as more trustworthy and representative than social media sources (Antwi et al., 2022). Media content can thus serve as evidence for policymakers and practitioners to (re)define policy priority and seek feasible relief measures.

Regarding the analysis of drought-related texts, there has been long-standing research with different analytic approaches and core domains of interest. Dow (2010) utilized qualitative content analysis (NVivo) to analyse decades-long news coverage 60 (1998-2007) in the Carolinas US identifying the impacts of droughts (e.g., agriculture, fires, lawns, recreation, etc.). Wei et al. (2015) also employed qualitative content analysis to explore historical news media reporting of water issues in Australia (1843-2011), revealing institutions and policy initiatives related to different types of water-related events. Osaka et al. (2020) examined content bias between nationwide and regional newspapers to analyse media attribution (i.e., likelihood and severity)



in descriptions of Californian droughts between 2011-2017. Painter et al. (2021) collected online news reports of the 2019
65 summer heatwave across France, Germany, the Netherlands, and the UK, and analysed how climate change was represented
in the content. Dayrell et al. (2022) took a corpus linguistics approach to filter the drought-related content from historical
archives and then closely read articles at the peak times of media coverage in history. While most works heavily relied on a
qualitative approach, natural language processing (NLP) methods enable a computational approach to read text; for example,
Sodoge et al. (2023) categorised types of drought impacts (i.e., agriculture, fires, forestry, livestock) in German newspapers
70 (2000-2021) using a machine learning text classifier.

There is a need to understand public perception of drought events, which can be framed with the analysis of media
representation in terms of media coverage and content. The first quarter of the 21st century was the period when broad scientific
consensus on anthropogenic climate change was reached, and also a period with more frequent, intense, and widespread
drought impacts. UK government policies have clarified a need to introduce programs for climate change adaptation, including
75 relevant programs for droughts (e.g., The Third National Adaptation Programme (NAP3) and the Fourth Strategy for Climate
Adaptation Reporting). Still, there is limited evidence to understand these drought events and how people perceive the hazard
in drought-prone areas, especially in England.

In this study, we analyse the public perception of droughts in England by using newspaper articles, focusing on 1) media
coverage related to meteorological conditions and 2) media content about droughts. We hypothesise that media coverage is
80 driven by the condition of hydrological droughts (i.e., higher temperatures and lower precipitation) with a stronger bias in the
summer season due to a confirmation bias – that is to say, an expectation that drought occurs in warm periods. In our content
analysis, we expect to find topics related to the direct impact on water consumers, and to see these transforms as drought
impact persists. Therefore, our research questions are summarised as follows:

RQ1. How does the media coverage relate to meteorological conditions and seasonality in England?

85 RQ2. What are typical topics in media content, and how do they differ for different major drought events?

2 Methodology

To answer the research questions, we build a newspaper corpus about drought events in England, and analyse the corpora a)
to compare the frequency of media coverage to drought-related meteorological anomalies and seasonality (RQ1), and to
explore the media content using natural language processing (RQ2). All the code used in our analysis is available as R-
90 Markdown (see Supplementary Material 1).



2.1 Data collection

2.1.1 Corpus building

The first step in any text-related project is to create a corpus composed of relevant documents for analysis. Among many approaches to building a corpus, we chose a keyword-based approach to collect newspaper articles about drought-related events in England. We chose only England, as opposed to the whole UK, since water management is delegated to the devolved governments of Northern Ireland, Wales and Scotland (Robins et al. 2017). Moreover, England's drier climate and higher population density leave it more susceptible to droughts (Folland et al., 2015), and we expected to find more news articles referring specifically to England.

To collect newspaper articles, we used the media hub platform Nexis Uni (www.advance.lexis.com). Initially, we retrieved articles that contained the term 'drought' in their headline field and were published between January 2000 and August 2023. However, since drought is an example of a term that may be used metaphorically (e.g., 'goal drought' in sporting terms), we further refined the search results to retain those containing the terms 'dry' and 'England'. We aimed to maximise both precision (i.e., the proportion of articles retrieved whose topic was water shortage in some form) and recall (i.e., the proportion of all articles describing water shortage in some form that we retrieved).

In making a query, we also defined the source of newspapers. We considered both broadsheet and tabloid newspapers to include a wide-spanning political spectrum (i.e., right- and left-wing stances) as well as readership (i.e., both wide-read tabloid journalism and broadsheets, referred to as the quality press) (Boykoff, 2008; Norton & Hulme, 2019). The broadsheet sources in our corpora were five newspapers (The Guardian, The Independent, The Times, Financial Times, and The Daily Telegraph) and tabloid sources were three newspapers (The Sun, Daily Mirror, Daily Mail). When the source of media was not identified, the articles were removed. In the end, we obtained 1,361 drought-related articles.

Having built the corpus, we carried out data-cleaning steps to increase precision (i.e., the proportion of relevant articles from the queried results). We first removed duplicate articles with the same title and then further removed articles that have their primary entries for the metadata to be geographic and thematic fields were irrelevant (e.g., geographic metadata referring to Wales, Scotland, Northern Ireland, and abroad; subject metadata referring to sport). Nonetheless, some very similar articles remained in the corpus, where, for example, online articles were produced as slightly edited versions of printed articles or vice-versa. After these steps, 836 articles (61%) remained for the analysis.

2.1.2 Meteorological data

We collected meteorological data from the web archive of the UK Meteorological Office (MET), the United Kingdom's national weather service. We focused on two meteorological variables as explanatory indicators of drought reports in media: monthly temperature (°C) and monthly precipitation (mm). We used two indices that capture long-term variation of climate in England, Mean Central England Temperature (°C) (hereafter CET; Parker et al. 1992, https://www.metoffice.gov.uk/hadobs/hadcet/data/meantemp_monthly_totals.txt) and the Monthly England and Wales



125 precipitation (mm) (hereafter EWP; Alexander and Jones, 2000,
https://www.metoffice.gov.uk/hadobs/hadukp/data/monthly/HadEWP_monthly_totals.txt). Both of these are well-established
regional indicators, and we chose temperature and precipitation rather than a direct indicator of drought (e.g., the Palmer
Drought Severity Index) since we wished to test the correlation between media coverage and instant meteorological events.
We first calculated long-term averages of CET and EWP for 1991-2020 (30 years), which served as baselines. We then
130 computed anomalies, i.e., deviations of CET and EWP from the long-term baseline. Additionally, we calculated the anomalies
for the temporal windows of 3, 6, and 12 months (e.g., for an event in January, the window of 3-month includes January and
the preceding December and November). This is to consider the lagging effect in physical processes, in the form of water
scarcity, which we hypothesised would align with the case for media coverage in that prolonged periods of higher temperatures
or low precipitation were likely to increase reporting of droughts. As a result, we prepared the average temperature anomalies
135 over 1, 3, 6, and 12-month windows, and the summed precipitation anomalies over 1, 3, 6, and 12-month windows.
We defined seasons as autumn SON (September-October-November), winter DJF (December-January-February), spring
MAM (March-April-May), and summer JJA (June-July-August), since September (the onset of autumn) is the beginning of
the hydrological year.

140 2.2 Media coverage in relation to meteorological conditions

2.2.1 Finding the optimal anomaly window

We used counts of monthly newspaper articles as a proxy for media coverage. Then, we aimed to find the best-fitting window
of meteorological anomalies (i.e., temperature and precipitation) for explaining the function of media coverage. For temporal
windows of 1, 3, 6, 12 months, we created scatter plots of temperature anomaly (x-axis) and precipitation anomaly (y-axis)
145 with the size of points indicating articles counts. We expected that drought-related articles would be most common in the 4th
quadrant – positive in temperature anomalies and negative in precipitation anomalies. To test for this, we used the chi-square
test for the total count of newspaper articles at each quadrant that explained the article distributions regarding each quadrant.

2.2.2 Seasonality test

150 Once we defined the optimal window, we proceeded with testing the bias of media coverage for seasonality. This is to test the
confirmation bias, i.e., the tendency to report droughts in summer seasons because of hot and dry weather. We generated a
correlation plot of the count of newspaper articles and meteorological anomalies by each season (i.e., SON, DJF, MAM, JJA),
and proceeded with the chi-square test.



155 2.2.3 Time-series analysis

We created time-series plots to show how the meteorological anomalies (i.e., temperature and precipitation) for the optimal window were correlated to media coverage. The first plot is a combination of a bar plot for monthly article counts and a line plot for anomaly fluctuations. The second plot delivers more in-depth information with a heatmap showing the monthly article counts and meteorological anomalies (i.e., temperature and precipitation), along with text labels and colour gradients. To
160 reduce skewed colour gradients due to outliers, we applied a winsorisation normalization method to replace the colours for extreme outliers with the colours of statistical endpoints of 1% and 99%.

2.3 Media topic: topic modelling

Topic modelling is an unsupervised computational method to identify and group topics through natural language processing.
165 The algorithm groups words found in documents in a corpus according to their probability of belonging to a topic (Blei et al., 2010). Topics are reported as groups of words associated with probabilities, and researchers can label them in a post-hoc process. Topic modelling is a common unsupervised approach to exploring the content in a corpus through so-called distant reading, without recourse to reading individual articles. It is often effective in giving an overview of different themes found in text in case of natural hazards (Zhang et al., 2021).

170 We carried out topic modelling using nouns in English, chosen for their important role in delivering key semantics about subjects (e.g. who is talking) and objects (e.g. what do they talk about). We used the pre-trained model in spaCy library (Benoit & Matsuo, 2020) to process our texts and extract lemmatized nouns as inputs for topic modelling. Lemmatising reduces words to a common root, allowing us to treat singular and plural forms as the same term. We then ran structural topic modelling (STM) using the r package ‘stm’ (Roberts et al., 2019), which employs latent Dirichlet allocation (LDA) and additional
175 metadata to perform topic clustering (Roberts et al., 2014). Since topic modelling is an unsupervised method, it is necessary to experiment to find the best-fitting number of topics, k , and we used two measures: coherence, which measures how well the terms in a topic are grouped in a document, and exclusivity, which captures how well topics split documents into different groups.

After performing topic modelling, we arranged the topics according to the descending order of topic proportions. For each
180 topic, we retrieved the ten keywords with the highest probabilities and ten keywords for the highest frex scoring algorithm (i.e., keyword extraction methods considering both frequent in and exclusive to a topic of interest) to understand the context of topics. Using these nouns, we assigned names to the topics.

Since STM is a mixed membership model, one article can contain more than one topic. For analytic convenience, we assigned one prime topic for each article based on the topical probability of exceeding a certain threshold (i.e., $3/k$). Assigning one crisp
185 topic to each article is useful for estimating the count and ratio of topics over time. Extracting entity names, such as frequently mentioned organization names within the sub-corpora, was also performed.



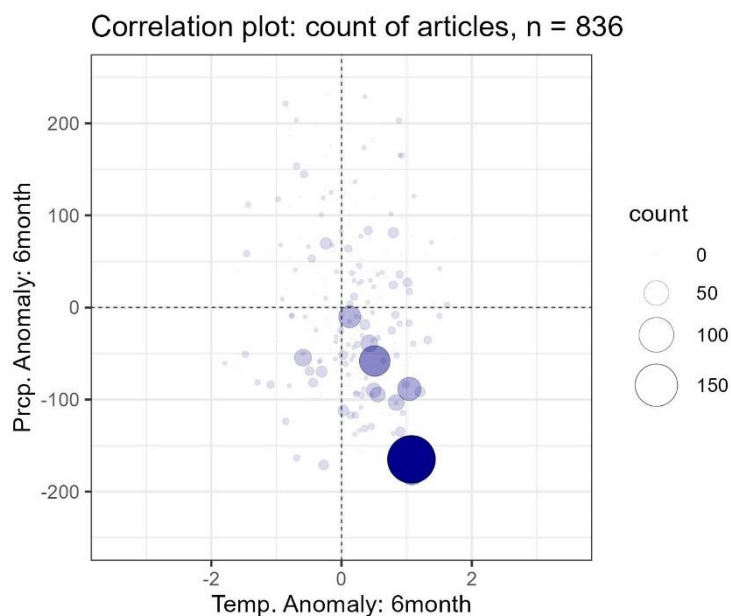
To take a closer look at the media content, we chose the drought in the first half of 2012 (Drought in Spring 2012 hereafter) and the drought in mid-2022 (Drought in Summer 2022 hereafter), which were reported as extreme droughts on the one hand by the official reports of the Centre for Ecology & Hydrology (i.e., Major Hydrological Events) and research papers (e.g.,
190 Barker et al., 2024), and on the other by looking at the times in which massive media coverage was accumulated in the heatmap above. At this point, we looked into the headlines and body content corresponding to the topics to further understand the context in which these droughts were represented. Additionally, we visualized the topic compositions using a treemap visualization (Tennekes, 2023).

195 **3 Results**

3.1 Media coverage

3.1.1 Lagging effect of meteorological conditions on media coverage

We tested the lagging effect of meteorological anomalies on media coverage with 1, 3, 6, and 12-month windows to confirm the 6-month window to be the most optimal to explain media coverage (Figure 1; results for other windows in Supplementary
200 Material 2). Qualitatively, it is clear that most articles (643) are found in the 4th quadrant, representing higher-than-normal temperature and lower-than-normal precipitation – conditions we might associate with a potential confirmation bias. The first, second, and third quadrants had 69, 42, and 82 counts, respectively, and the distribution of article counts differs significantly from what we would expect by chance (χ -test, $p < 0.001$).



205

Figure 1 Correlation plot between temperature anomaly (6-month average) and precipitation anomaly (6-month sum) to explain the number of news articles (i.e., the size and colour of points). The anomaly is calculated based on the average of 1991-2020.

3.1.2 Seasonality bias on media coverage

210 To explore in more detail whether media coverage is subject to seasonality, in addition to meteorological anomalies, we broke
down the plot into seasons (Figure 2). Quantitatively, there is a noticeable concentration of reporting of drought in spring
(MAM) and summer (JJA) with the distribution of articles differing from the overall distribution of climatic anomalies (χ -test,
 $p < 0.001$). This suggests that media coverage increased when the meteorological conditions were favourable to droughts (4th
quadrant). Yet, the similar meteorological conditions in seasons other than spring or summer did not result in as much media
215 coverage.



Correlation plot: count of articles by seasonality

SON: 55, DJF: 115, MAM: 295, JJA: 371

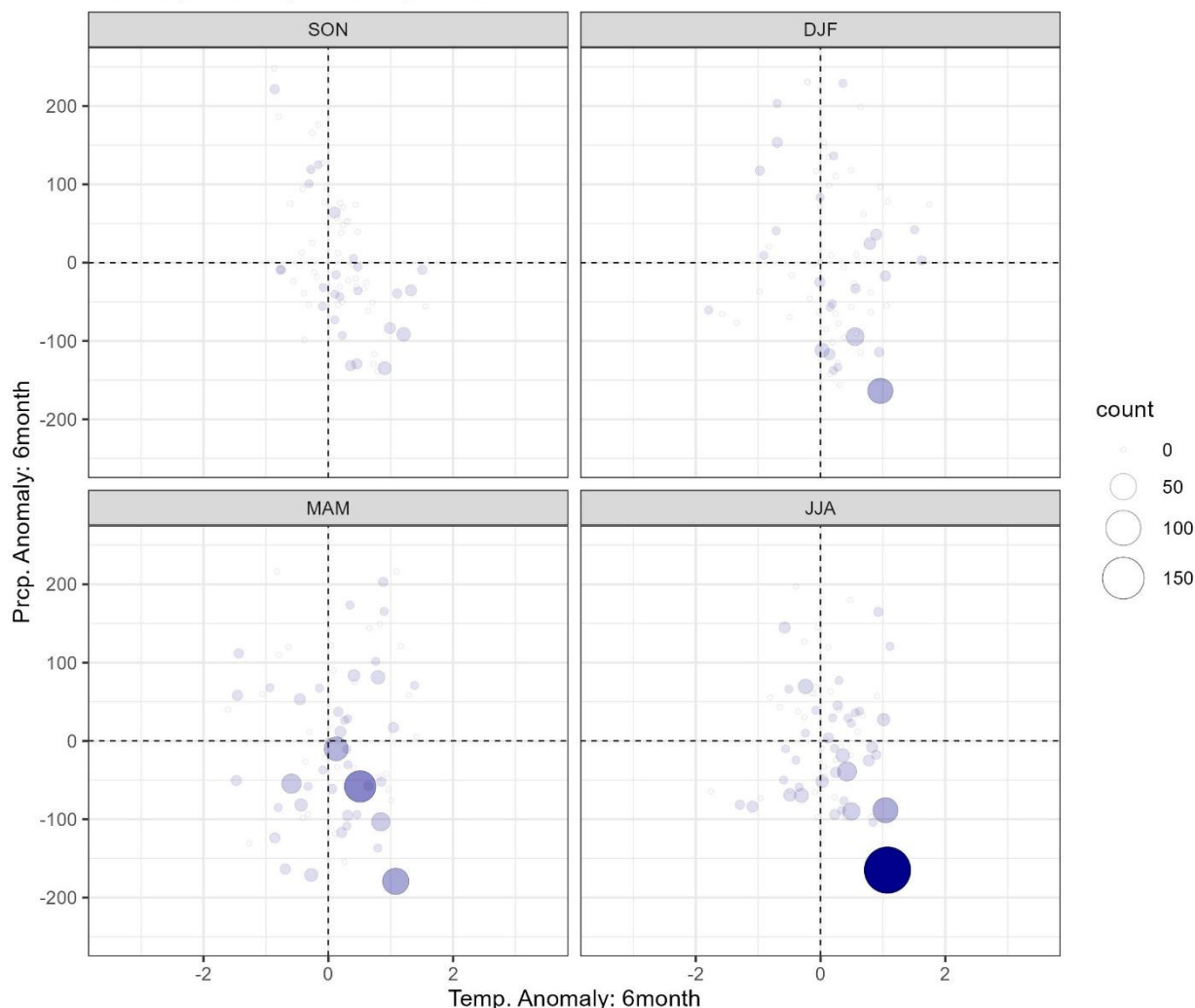


Figure 2 Correlation plot between temperature anomaly (6-month average) and precipitation anomaly (6-month sum) to explain the number of news articles (i.e., the size of points) depending on seasonality. The anomaly is calculated based on the average of 1991-2020.

220

3.1.3 Interplay between meteorological conditions on media coverage

The combined plot of media coverage and meteorological anomalies (Figure 3) shows how media coverage co-developed with drought-prone conditions, which are shaded in yellow for higher temperature and lower precipitation anomalies. From this plot, we found multiple drought events gathered substantial media coverage, aligning with major drought events declared in governmental reports and scientific articles, e.g., in 2005-2006; 2011; 2017 and 2018/19 (Turner et al., 2021) and most

225



markedly in 2022 (Barker et al., 2024). The plot also suggested that most media coverage for major droughts coincided when the meteorological condition consisted of both positive temperature and negative precipitation anomalies (see the yellow shades). However, not all drought-prone conditions resulted in media coverage. For further analyses, we chose the two droughts with the most pronounced media responses, spring 2012 and summer 2022.

230

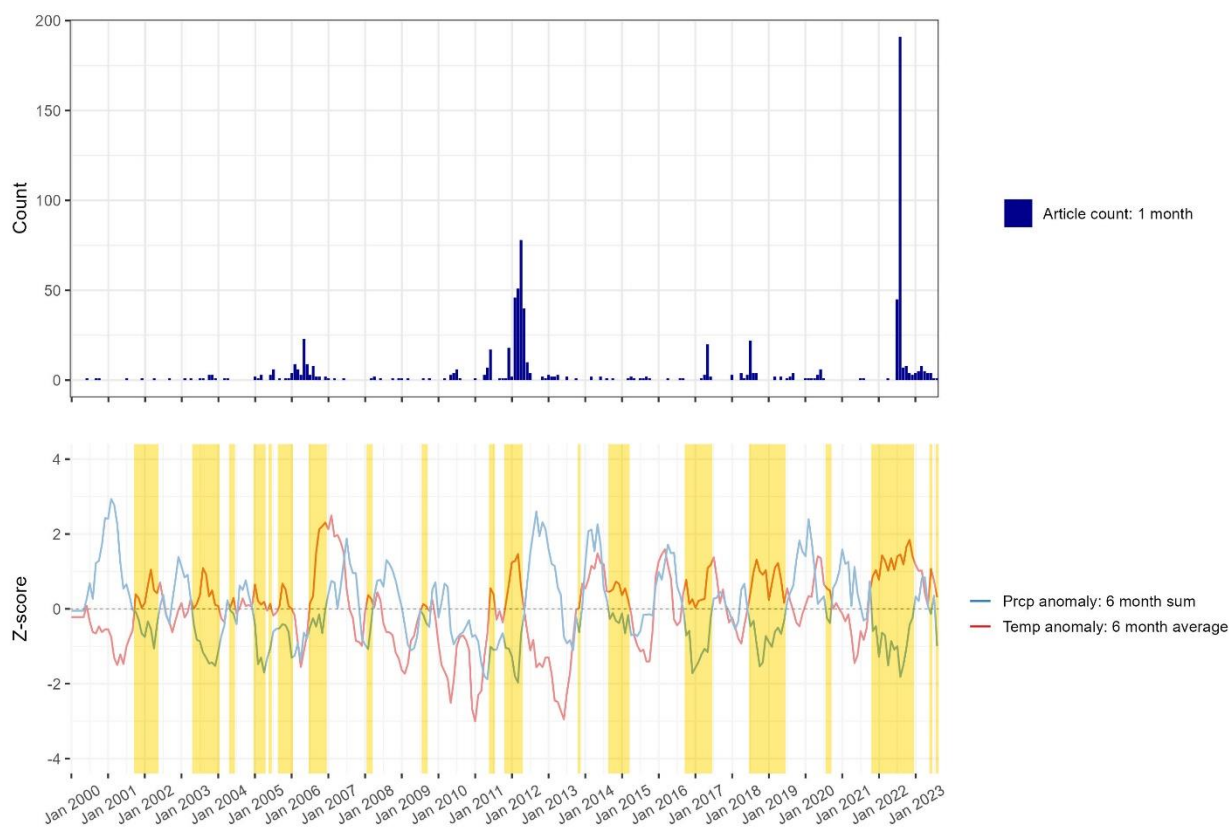


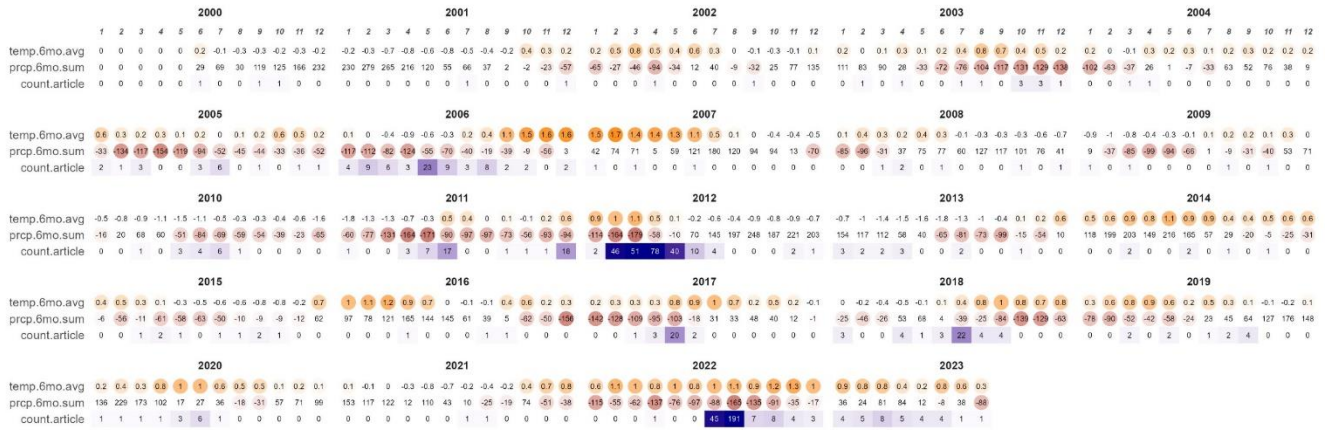
Figure 3 Time-series plot to show the media coverage (i.e., monthly count of newspaper articles) (top) and the meteorological anomalies (i.e., 6-month average temperature and 6-month sum precipitation) along with the drought-prone conditions in yellow shades (bottom).

235

To explore these patterns in more detail, we also plotted a heatmap (Figure 4). Drought-related media coverage was the greatest when both temperature and precipitation deviated from long-term average baselines at the same time, confirming our interpretation of Figure 3. For the most pronounced media coverage, including drought in spring 2012 and summer 2022, the drought-prone conditions were compounded by longer preceding periods of positive temperature and negative precipitation

240

anomalies, i.e., periods of six months or more of rainfall deficit, accompanied by anomalously high temperatures. Furthermore, the heatmap shows how media coverage of drought persisted after the drought-prone anomalies were resolved. To find out more about what was discussed in media coverage, we turn to topic modelling.



245 **Figure 4 Heatmap to show temporal trends of 1) temperature anomaly (6-month average), 2) precipitation anomaly (6-month sum), 3) number of news articles. Colour schemes were normalised to minimise the extreme effects of outliers.**

3.2 Media topics about drought

3.2.1 Topic labels

250 According to the coherence and exclusivity tests, we confirmed the optimal number of topics, k , to be 10 (see Supplementary material 3). Then we summarized the frequent keywords, frex keywords, and organization names for the topics in Table 1, following the order of topical proportions in the corpus (Supplementary Material 4a). Topics were then labelled as a post-hoc process to find the best title representing the keywords and organization names.

255 The most common topic in the corpus was Topic #1, was labelled “Drought status” as it described the severity and expanse of drought events, focusing on hydrological drought, mentioning ‘water’ ‘level’ and hydrological terms such as ‘river’ and ‘groundwater’. Common organizations in this topic included ‘The Environment Agency’ (a government body responsible for water related legislation and flood warnings) and ‘Thames Water’ (the largest water supply company in the UK, providing water to Greater London and beyond).

260 The second and fourth most popular topics were “Rain forecast” (Topic #10) and “Flooding” (Topic #6). These topics appear to relate to the (potential) end of drought periods and likely reflect articles describing events associated with the end of long, dry periods (though not necessarily indicating the end of hydrological drought). These topics are again related to ‘The Environment Agency’ and ‘The Met Office’ (which provide flood and weather warnings respectively in England).

The third most popular topic was “Water company” (Topic #5), which included the discussion of supply failure and incompetent managerial practices through terms such as ‘leak’. Common organization entities included major water suppliers, such as ‘Thames Water’, ‘Southern Water’, ‘Ofwat’, the Water Service Regulation Authority, also ranked high in organization



265 entities associated with this topic. This topic is clearly related to water as a resource, which has become increasingly controversial after water privatization in England (Bayliss et al., 2021).

The topic “Hosepipe ban” (Topic #8) followed, capturing a commonly used restriction during droughts in the UK, namely banning water usage for outdoor pools, gardening, and car washing. Accordingly, many organizations in this topic included major water suppliers, such as ‘Thames Water’ and ‘Southern Water’. Hosepipe bans are important since they are a visible
270 measure directly impacting on individual consumers and their behaviour.

Other topics which were less common included “Aquatic ecosystem” (Topic #4), “Agriculture” (Topic #2), and “Gardening” (Topic #3). These topics exemplified public attention paid to practical impacts to the environment and everyday life (i.e., groceries). Organizations related to these topics included the RSPB (The Royal Society for the Protection of Birds) and Natural England for “Aquatic ecosystem”, NFU (National Farmers’ Union) for “Agriculture”, and RHS (Royal Horticultural Society)
275 for “Gardening”.

Topic #7 consists of heterogeneous keywords that made it difficult to interpret their context in which these topics appear. These topics may reflect metaphorical use of drought, despite our data cleaning processes, since many keywords seem related to sport.

Finally, Topic #9 seems to include more general keywords related to water-based hazards in the UK, such as sea-level rise and
280 coastal erosion and challenges with respect to insurance for homeowners. As such, it expands beyond drought to other hazards, often reported as general consequences of climate change.

Table 1 Description of 10 topics with their label names, probability keywords, frex keywords, and key organization names. The order of topics follows the expected topic proportions in the corpus.

#	Topic labels	Ten keywords based on probability (prob), frex estimation (frex) and ten frequent organization entities in descending order of frequency
1	Drought status	<ul style="list-style-type: none"> • (prob) water, drought, river, level, rainfall, company, area, ban, hosepipe, year • (frex) agency, river, flow, status, groundwater, environment, restriction, permit, level, drought • (org) The Environment Agency (including ‘the –’, ‘EA’) (274+122+84+62), Thames Water (119), Southern Water (73), South East Water (59), Anglian Water (56), Government (48), The Met Office (44)
10	Rain forecast	<ul style="list-style-type: none"> • (prob) rain, year, rainfall, month, weather, drought, per, cent, drought, record • (frex) mm, rain, average, snow, rainfall, inch, month, pollen, aquifer, figure • (org) The Environment Agency (47+35), The Met Office (47+35+19), Thames Water (26), Southern Water (19), United Utilities (13), The Centre for Ecology and Hydrology (12)



5	Water companies	<ul style="list-style-type: none"> • (prob) water, company, year, drought, supply, day, litre, leak, order, use • (frex) leak, meter, target, order, desalination, bill, litre, metering, pipe, leakage • (org) Thames Water (138), Ofwat (116), The Environment Agency (53+36), Southern Water (44), South East Water (38), Government (30)
6	Flooding	<ul style="list-style-type: none"> • (prob) rain, temperature, week, weather, day, flood, water, drought, flooding, area • (frex) hail, flash, thunderstorm, tornado, wind, downpour, flooding, alert, lightning, blaze • (org) The Environment Agency (79+50+32+18), The Met Office (73+64+37), Thames Water (35), Yorkshire Water (15)
8	Hosepipe bans	<ul style="list-style-type: none"> • (prob) water, ban, hosepipe, customer, drought, company, temperature, day, area, week • (frex) customer, pool, health, fire, hosepipe, ban, bottle, wildfire, heatwave, fine • (org) Thames Water (154), Southern Water (82), The Environment Agency (69), South East Water (63), Government (45), The Met Office (44), Yorkshire Water (39), Welsh Water (38)
4	Aquatic ecosystems	<ul style="list-style-type: none"> • (prob) water, year, climate, reservoir, river, change, drought, government, wildlife, plan • (frex) beaver, wetland, vole, toad, lapwing, habitat, bird, habitat, population, abstraction, specie(s) • (org) The Environment Agency (41+11), Thames Water (38), Southern Water (18), RSPB (17), Natural England (17), Ofwat (13), WWF (11)
2	Agriculture (farmers and crops)	<ul style="list-style-type: none"> • (prob) farmer, crop, year, food, price, drought, weather, condition, potato, vegetable • (frex) barley, potato, wheat, harvest, livestock, crop, grower, yield, farmer, carrot • (org) NFU (The National Farmers' Union) (38+10), The Environment Agency (19), EU (12), Government (7), Lidl (7)
3	Gardening (terrestrial ecosystems)	<ul style="list-style-type: none"> • (prob) water, plant, garden, tree, drought, lawn, summer, ban, hosepipe, butt • (frex) butterfly, robinia, leaf, bluebell, leave, pseudoacacia, berry, ivy, lawn, pot • (org) Thames Water (11), Southern Water (10), The Environment Agency (10), South East Water (8), The Met Office (8), RHS (6)
7	N/A (sport and hobby)	<ul style="list-style-type: none"> • (prob) year, drought, time, thing, day, school, people, rain, goal, garden • (frex) lock, cave, prairie, trophy, goal, show, school, game, defender, housewife • (org) Chelsea (16), RHS (15), Arsenal (12), Government (10), Thames Water (10)
9	N/A (general water hazards)	<ul style="list-style-type: none"> • (prob) year, cent, per, home, drought, flood, property, village, flooding, climate • (frex) subsidence, claim, insurer, council, barge, village, beach, defence, ship, map • (org) The Environment Agency (11+8), The Independent (10), Thames Water (9), BGS (8)



285

3.2.2 Close reading

We then explored how drought in spring 2012 and summer 2022 were presented in media contents performing close reading of newspaper headlines (c.f. Dayrell et al. 2022). We also visualized the topic compositions for each month for these two events, to show the prevalence of topics. Topic compositions for the entire study periods can be found in Supplementary
290 Material 4b.

Unsurprisingly, “Drought status” (Topic #1) was prevalent throughout the droughts in spring 2012 and summer 2022, which implied the onset of drought-prone conditions, e.g. “Crippling drought hits south and east of England [Article 5, February 2012]” and “Drought fears after England suffers driest spell since 1976 [Article 255, July 2022]”.

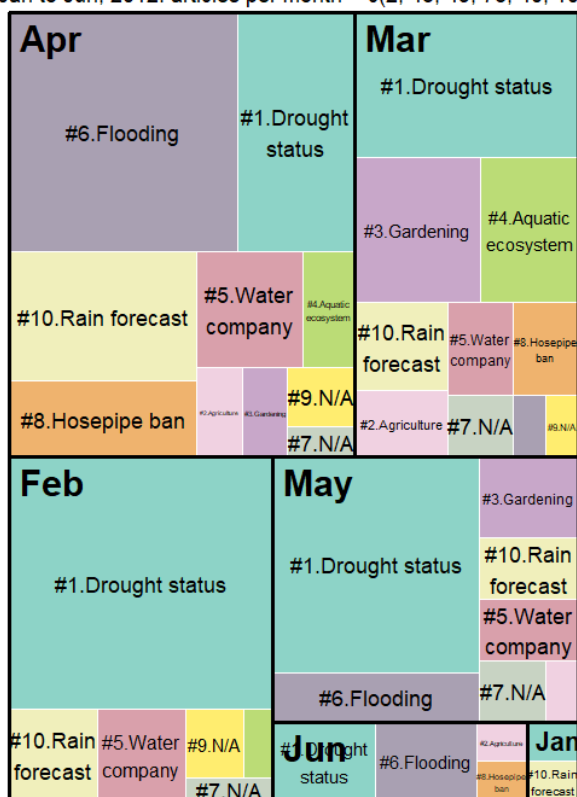
We found the counter-intuitive topic of “Flooding” (Topic #6) to be the most dominant topic for April during the drought in
295 spring 2012. The topic also appeared in August for drought in summer 2022, although less prominently. The articles in this topic positioned the flooding in contrast to the precedent droughts, e.g. “From drought to flood warnings in a week [Article 475, 2012.4]” or “Gales and floods but, yes, it's still a drought [Article 486, 2012.2]”. Such rhetoric was repeated in the “Rain forecast” (Topic #10), “Thunderstorms and heavy rain to hit England... but it won't ease the drought [Article 827, August 2022]”.

300 One noteworthy topic related to drought in summer 2022 was the “Hosepipe ban” (Topic #8). Although it is a common measure in the UK to mitigate water demand during the drought-prone conditions, the topic did not command much attention during the drought in spring 2012. In contrast, in summer 2022 it accrued substantial media coverage, e.g., “Southern Water announces hosepipe ban over drought fears [Article 764, July 2022]”.

An interesting finding for the drought in spring 2012 included the “Gardening” (Topic #3) in March 2012, e.g., “Drought: a
305 gardening survival guide for a dry season [Article 357, March 2012]”. This topic was less salient in other months of 2012 and during the drought in summer 2022. Media reporting about gardening may be related to the seasonality with water shortages an especially important issue during the early spring growing season.



Jan to Jun, 2012. articles per month = c(2, 45, 49, 76, 40, 10)



July to Dec, 2022. articles per month = c(45, 189, 7, 8, 4, 3)

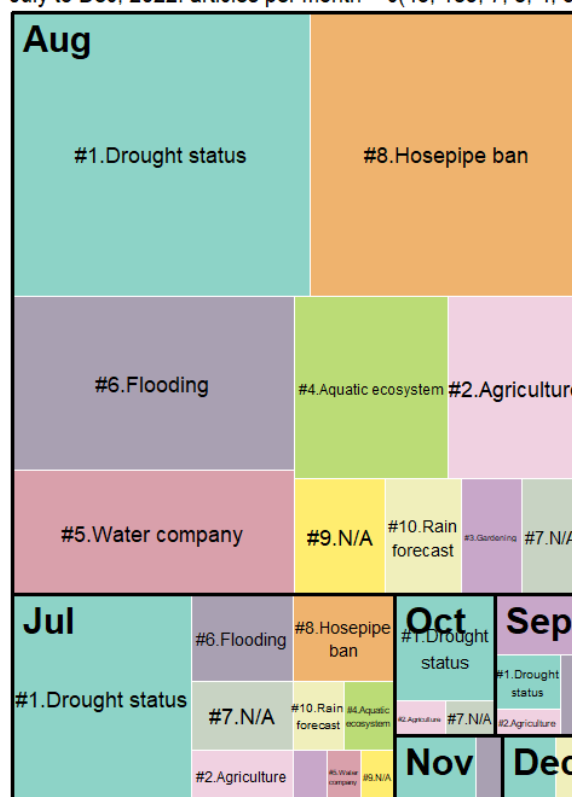


Figure 5 Treemaps comparing topic compositions for droughts in spring 2012 and summer 2022.

4 Discussion

310 4.1 How do meteorological conditions relate to media coverage?

Our findings confirmed that substantial media coverage was supported by drought-prone meteorological conditions, with distinctive examples of droughts in spring 2012 and summer 2022 (Figure 1). However, drought-prone conditions with a 6-month window did not always generate noticeable media coverage as shown in 2003, 2005, 2017, and 2018-2019. Such an absence of media coverage, despite climatic anomalies favourable for drought conditions in our analysis, is interesting because

315 the UKCEH also confirmed the drought-prone conditions in these years with reports of unusual hydrological events published for the years of 2003, 2004-2006, 2010-2012, and 2018-2019 (UKCEH, n.d.). A similar mismatch between media coverage and drought-prone meteorological conditions was also identified by Dayrell et al. (2022) who explored the text corpora about droughts over the last 200 years in the UK. Dayrell et al. (2022) suggested that such mismatched derived from the newsworthiness of events: even if the meteorological conditions are favourable for droughts, they may not generate media

320 coverage due to the lagging effect of rainfall shortage to newsworthy drought events. In addition, incipient droughts rarely get



media coverage. While the findings from Dayrell et al. (2022) are applicable to our analysis, we explored these mismatches with two potential further explanations.

The first is a seasonality bias. We found that drought-prone conditions in summer generated more media coverage than similar anomalies in other seasons. This could indicate a confirmation bias driving media coverage –drought is more likely to be reported in warmer conditions since it confirms a simplistic narrative relating drought to warm weather.

Second, preferential reporting of drought conditions in summer may relate to the hydrological cycle in the UK, which begins with a wet winter (SON) and ends with the summer season (JJA). Higher temperature anomalies in the summer are thus, given a long-term precipitation deficit, more likely to lead to drought conditions. This increased importance of temperature anomalies in summer concerning drought in England was explained by Folland through summer temperature leading to high rates of evapotranspiration.

Another potential cause of mismatches are precedent conditions, especially the precipitation. Even when we applied a 6-month window to capture the lagging effect of meteorological conditions on hydrological processes, we found that the preceding precipitation conditions, for example the intensity and duration of low precipitation anomalies were more critical to media coverage. For example, 2003 was remembered as a severe drought crisis that swept the entire European continent, and the UK also experienced anomalies in temperature and precipitation, marking “the second lowest February-October rainfall in 83 years”. Nonetheless, media coverage in this period was limited, likely because heavy rainfalls in November and December of 2002 filled up reservoirs and recharged groundwater resources (Marsh 2014; Marsh et al., 2014). Similar patterns can be also found in 2005 and 2019, which resulted in limited media coverage of drought.

The important role of precedent conditions, especially the long-term anomalies of low precipitation, in driving media attention allowed us to revisit the droughts in spring 2012 and summer 2022, and examine their antecedent conditions. The peak in media coverage during the drought in spring 2012 coincided with the drought-prone conditions of anomalies. Still, a more critical factor may lie in the precedent lack of precipitation (i.e., low precipitation anomalies) that had persisted since mid-2010 and dried up the groundwater table, soil moisture, and major reservoirs (Marsh et al., 2013). A similar pattern of long-term low precipitation anomalies was correlates with the extensive media coverage of the drought in summer 2022, which had persisted since early 2022. These long-term low precipitation anomalies both overlapped with the winter season, which is critical for water cycles in England. Dayrell et al. (2022) also suggested that newsworthy attention for droughts was associated with dry winters, with the impacts manifested in the following summer.

One of the key limitations of our study is its spatial granularity. We aggregated meteorological variables nationally (i.e., England), obscuring regional variations. In reality, UKCEH reported the drought in 2012 as beginning with rainfall deficiencies in Wales and Scotland in the first half of 2010. In 2011 and 2012, the drought was tracked in southern England due to the water shortfalls and higher temperatures (Marsh et al., 2013, p.7). By aggregating climatic data to England, we could not distinguish between geographic transitions in media coverage in reporting drought events.

Another limitation relates to the demand side of droughts, which exacerbates the impact on people and society. We chose England as a study site for being vulnerable to drought events (i.e., low precipitation in general, drier winters than other regions



355 of the UK) as well as high water demand, such as dense population and industrial infrastructures (Folland et al., 2015). Still, we did not link increased water demand to media coverage.

4.2 Which topics are represented in media?

360 Topic modelling allowed us to explore media representation of droughts and related narratives. We found eight meaningful topics dominant in drought-related articles in England from 2012 and 2022, including “Drought status” (Topic #1) which represented the role of media in describing weather conditions, as also found in “Rain forecast” (Topic #10) and “Flooding” (Topic #6).

Other than weather reports, a key topic we found in the media content was “Hosepipe ban” (Topic #8), a common measure of discouraging water usage in the context of England. During drought-prone conditions in England, private water companies, who manage water in England, discourage or ban people from using water for tasks such as gardening or car washing. 365 Connected to these issues, we identified the topic of “Water company” (Topic #5), which was often related to criticism of water supply companies unable to handle metering or leakage issues.

The topic of water supply in drought-related articles of England resonates with a study by Stahl et al. (2016), who also revealed that major concerns related to droughts in the UK revolve around water supply issues, in contrast to other countries where 370 different concerns emerged. Detailed concerns related to the public water supply in the UK included ‘local and regional water supply shortage and problems (drying up of springs/wells, reservoirs, streams)’ and ‘bans on domestic and public water use (e.g., car washing, watering the lawn/garden, irrigation of sports fields, filling of swimming pools)’ (Stahl et al., 2016). This result contrasted with other European countries where primary concerns with drought conditions were more likely to related to ‘agriculture and livestock farming’.

375 The following concerns in the UK were ranked with ‘water quality’ and ‘freshwater ecosystems’, along with ‘water quality deterioration of surface waters (significant change of physio-chemical indicator)’ the most, followed by ‘increased temperature in surface waters’ (Stahl et al., 2016). Such a finding corresponded with the topics we found in this study, such as “Aquatic ecosystem” (Topic #4). Both Stahl et al. (2016) and our work did not find ‘drinking water quality’ a major concern in the UK, at least not in drought related news articles in our case.

380 Still, a further limitation of our study relates to the nature of our corpus, which consists of a single genre, i.e., newspapers, which have also lost readership coincident with the emergence of online news platforms (Kim et al., 2019), social media platforms (e.g., Twitter) and non-text content platforms (e.g., YouTube). Although we attempted to diversify and maximize the readership by analysing both broadsheet and tabloid newspapers, this limitation could potentially be alleviated by data triangulation which collecting and comparing voices from different data sources and genres of text (Antwi et al., 2022).

385 We applied an unsupervised approach - topic modelling algorithm - to summarise and classify the content of newspapers. The advantages of this method include its transparency and ability to analyse large volumes of data, but it is important to note that the approach is stochastic and sensitive to how the corpus is built. Probabilities of keyword distribution were calculated using



390 a bag-of-words model, which does not consider the context in which the keywords appear beyond the documents in which they are found. Alternative approaches could, for example, predefine topics from the literature and use machine learning to extract these topics from a corpus (e.g., Sodge et al., 2023).

5 Conclusion

In this study, we built a newspaper corpus about droughts in England to analyse how the media coverage is related to meteorological conditions and what topics emerge in the media during drought periods. We found that ‘meteorological droughts’ –can partially explain the media coverage about droughts in England: media coverage coincided with both low-precipitation and high-temperature anomalies, compounded by prolonged negative anomalies in precipitation. We also found a seasonality bias in media coverage, with more reporting of droughts in warmer periods and summer. Key topics with topic modelling further elaborated that visible and tangible drought impacts on people consisted of media contents about droughts, such as water shortages (e.g., drought status), water use restrictions (e.g., hosepipe ban, water companies), and the following consequences (e.g., gardens, agricultural products, ecosystems). To summarise, media coverage about droughts tends to occur only if the meteorological conditions worsen to develop into hydrological, agricultural, and socioeconomic droughts that affect society. Thus, meteorological droughts (e.g., occasional reports by the UKCEH) do not always lead to media attention, facilitating public consensus and agenda framing for alleviation measures. In other words, the underrepresentation of media coverage at the stage of hydrological droughts can result in the lack of public awareness for ‘creeping’ droughts, failing to garner the public support for introducing preventative and adaptive measures, such as early warning systems. The media content analysis also helps navigate where the policy priorities need to be headed to handle imminent challenges the in water management system.

Code availability

The R code lines we applied for the analysis were archived in a Markdown file, which can be found in the Supplementary Material 1.

410 Supplementary Materials

All the supplementary materials mentioned in the manuscript, including the R Markdown, can be found in the online version of publication.



Author contribution

All the authors (IK, JS, RP) conceptualised and developed the research questions. IK took a lead on implementation, including data collection, formal analysis, and visualisation, upon the communication with all authors. JS and RP prepared the original draft, and all the authors (IK, JS, RP) contributed the edits, reviews, revisions for the final draft.

Competing interests

One of our co-authors is a member of the editorial board of Hydrology and Earth System Sciences (HESS).

References

- Alexander, L. V. and Jones, P. D.: Updated Precipitation Series for the U.K. and Discussion of Recent Extremes, *Atmos. Sci. Lett.*, 1, 142–150, <https://doi.org/10.1006/asle.2001.0025>, 2000.
- Antwi, S. H., Rolston, A., Linnane, S., and Getty, D.: Communicating water availability to improve awareness and implementation of water conservation: A study of the 2018 and 2020 drought events in the Republic of Ireland, *Sci. Total Environ.*, 807, 150865, <https://doi.org/10.1016/j.scitotenv.2021.150865>, 2022.
- Barker, L. J., Hannaford, J., Magee, E., Turner, S., Sefton, C., Parry, S., Evans, J., Szczykulska, M., and Haxton, T.: An appraisal of the severity of the 2022 drought and its impacts, *Weather*, 1–12, <https://doi.org/10.1002/wea.4531>, 2024.
- Bayliss, K., Mattioli, G., and Steinberger, J.: Inequality, poverty and the privatization of essential services: A ‘systems of provision’ study of water, energy and local buses in the UK, *Compet. Chang.*, 25, 478–500, <https://doi.org/10.1177/1024529420964933>, 2021.
- Benoit, K. and Matsuo, A.: spacyr: Wrapper to the 'spaCy' 'NLP' Library_. R package version 1.2.1. (<https://CRAN.R-project.org/package=spacyr>), 2020. (Last access on 30 May, 2024)
- Blei, D., Carin, L., and Dunson, D.: Probabilistic Topic Models, *IEEE Signal Process. Mag.*, 27, 55–65, <https://doi.org/10.1109/MSP.2010.938079>, 2010.
- Boykoff, M. T.: The cultural politics of climate change discourse in UK tabloids, *Polit. Geogr.*, 27, 549–569, <https://doi.org/10.1016/j.polgeo.2008.05.002>, 2008.
- Briffa, K. R., van der Schrier, G., and Jones, P. D.: Wet and dry summers in Europe since 1750: Evidence of increasing drought, *Int. J. Climatol.*, 29, 1894–1905, <https://doi.org/10.1002/joc.1836>, 2009.
- Dayrell, C., Svensson, C., Hannaford, J., McEnery, T., Barker, L. J., Baker, H., and Tanguy, M.: Representation of Drought Events in the United Kingdom: Contrasting 200 years of News Texts and Rainfall Records, *Front. Environ. Sci.*, 10, 1–18, <https://doi.org/10.3389/fenvs.2022.760147>, 2022.
- Dessai, S. and Sims, C.: Public perception of drought and climate change in southeast England, *Environ. Hazards*, 9, 340–357, <https://doi.org/10.3763/ehaz.2010.0037>, 2010.



- Dobson, B., Coxon, G., Freer, J., Gavin, H., Mortazavi-Naeini, M., and Hall, J. W.: The Spatial Dynamics of Droughts and Water Scarcity in England and Wales, *Water Resour. Res.*, 56, <https://doi.org/10.1029/2020WR027187>, 2020.
- 445 Dow, K.: News coverage of drought impacts and vulnerability in the US Carolinas, 1998-2007, *Nat. Hazards*, 54, 497–518, <https://doi.org/10.1007/s11069-009-9482-0>, 2010.
- Folland, C. K., Hannaford, J., Bloomfield, J. P., Kendon, M., Svensson, C., Marchant, B. P., Prior, J., and Wallace, E.: Multi-annual droughts in the English Lowlands: A review of their characteristics and climate drivers in the winter half-year, *Hydrol. Earth Syst. Sci.*, 19, 2353–2375, <https://doi.org/10.5194/hess-19-2353-2015>, 2015.
- 450 Guerreiro, S. B., Dawson, R. J., Kilsby, C., Lewis, E., and Ford, A.: Future heat-waves, droughts and floods in 571 European cities, *Environ. Res. Lett.*, 13, <https://doi.org/10.1088/1748-9326/aaaad3>, 2018.
- Kay, A. L., Bell, V. A., Davies, H. N., Lane, R. A., and Rudd, A. C.: The UKSCAPE-G2G river flow and soil moisture datasets: Grid-to-Grid model estimates for the UK for historical and potential future climates, *Earth Syst. Sci. Data*, 15, 2533–2546, <https://doi.org/10.5194/essd-15-2533-2023>, 2023.
- 455 Kim, S., Shao, W., and Kam, J.: Spatiotemporal patterns of US drought awareness, *Palgrave Commun.*, 5, 1–9, <https://doi.org/10.1057/s41599-019-0317-7>, 2019.
- Marsh, T.: The UK drought of 2003 - an overview, *Weather*, 59, 2004. (<https://nora.nerc.ac.uk/id/eprint/510184/1/N510184CR.pdf>)
- Marsh, T., Cole, G., and Wilby, R.: Major droughts in England and Wales, 1800-2006, *Weather*, 62, 87–93, <https://doi.org/10.1002/wea.67>, 2007.
- 460 Marsh, T., Booker, D., and Fry, M. The 2004-06 Drought 2nd edition, 2014. (<https://nora.nerc.ac.uk/id/eprint/510185/1/N510185CR.pdf>)
- Marsh, T.J., Parry, S., Kendon, M.C., and Hannaford, J.: The 2010-12 drought and subsequent extensive flooding. Centre for Ecology & Hydrology, 2013. (<https://nora.nerc.ac.uk/id/eprint/503643/1/N503643CR.pdf>)
- 465 Murphy, C., Wilby, R. L., Matthews, T., Horvath, C., Crampsie, A., Ludlow, F., Noone, S., Brannigan, J., Hannaford, J., McLeman, R., and Jobbova, E.: The forgotten drought of 1765–1768: Reconstructing and re-evaluating historical droughts in the British and Irish Isles, *Int. J. Climatol.*, 40, 5329–5351, <https://doi.org/10.1002/joc.6521>, 2020.
- Norton, C. and Hulme, M.: Telling one story, or many? An ecolinguistic analysis of climate change stories in UK national newspaper editorials, *Geoforum*, 104, 114–136, <https://doi.org/10.1016/j.geoforum.2019.01.017>, 2019.
- 470 O’Connor, P., Murphy, C., Matthews, T., and Wilby, R. L.: Relating drought indices to impacts reported in newspaper articles, *Int. J. Climatol.*, 43, 1796–1816, <https://doi.org/10.1002/joc.7946>, 2023.
- Osaka, S., Painter, J., Walton, P., and Halperin, A.: Media representation of extreme event attribution: A case study of the 2011–17 California drought, *Weather. Clim. Soc.*, 12, 847–862, <https://doi.org/10.1175/WCAS-D-19-0050.1>, 2020.
- Parker, D. E., Legg, T. P., and Folland, C. K.: A new daily central England temperature series, 1772–1991, *Int. J. Climatol.*, 12, 317–342, <https://doi.org/10.1002/joc.3370120402>, 1992.
- 475



- Quesnel, K. J. and Ajami, N. K.: Changes in water consumption linked to heavy news media coverage of extreme climatic events, *Sci. Adv.*, 3, 1–10, <https://doi.org/10.1126/sciadv.1700784>, 2017.
- Roberts, M. E., Stewart, B. M., and Tingley, D.: Stm: An R package for structural topic models, *J. Stat. Softw.*, 91, <https://doi.org/10.18637/jss.v091.i02>, 2019.
- 480 Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., and Rand, D. G.: Structural topic models for open-ended survey responses, *Am. J. Pol. Sci.*, 58, 1064–1082, <https://doi.org/10.1111/ajps.12103>, 2014.
- Robins, L., Burt, T. P., Bracken, L. J., Boardman, J., and Thompson, D. B. A.: Making water policy work in the United Kingdom: A case study of practical approaches to strengthening complex, multi-tiered systems of water governance, *Environ. Sci. Policy*, 71, 41–55, <https://doi.org/10.1016/j.envsci.2017.01.008>, 2017.
- 485 Sodge, J., Kuhlicke, C., and de Brito, M. M.: Automated spatio-temporal detection of drought impacts from newspaper articles using natural language processing and machine learning, *Weather Clim. Extrem.*, 41, 100574, <https://doi.org/10.1016/j.wace.2023.100574>, 2023.
- Spinoni, J., Naumann, G., Vogt, J. V., and Barbosa, P.: The biggest drought events in Europe from 1950 to 2012, *J. Hydrol. Reg. Stud.*, 3, 509–524, <https://doi.org/10.1016/j.ejrh.2015.01.001>, 2015.
- 490 Spinoni, J., Vogt, J. V., Naumann, G., Barbosa, P., and Dosio, A.: Will drought events become more frequent and severe in Europe?, *Int. J. Climatol.*, 38, 1718–1736, <https://doi.org/10.1002/joc.5291>, 2018.
- Stahl, K., Kohn, I., Blauhut, V., Urquijo, J., De Stefano, L., Acácio, V., Dias, S., Stagge, J. H., Tallaksen, L. M., Kampragou, E., Van Loon, A. F., Barker, L. J., Melsen, L. A., Bifulco, C., Musolino, D., De Carli, A., Massarutto, A., Assimacopoulos, D., and Van Lanen, H. A. J.: 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/nhess-16-801-2016>, 2016.
- 495 Tanguy, M., Haslinger, K., Svensson, C., Parry, S., Barker, L. J., Hannaford, J., and Prudhomme, C.: Regional Differences in Spatiotemporal Drought Characteristics in Great Britain, *Front. Environ. Sci.*, 9, <https://doi.org/10.3389/fenvs.2021.639649>, 2021.
- 500 Tennekes, M.: `_treemap: Treemap Visualization_`. R package version 2.4-4, 2023. <https://CRAN.R-project.org/package=treemap> (last access on 30 May, 2024).
- Turner, S., Barker, L. J., Hannaford, J., Muchan, K., Parry, S., and Sefton, C.: The 2018/2019 drought in the UK: a hydrological appraisal, *Weather*, 76, 248–253, <https://doi.org/10.1002/wea.4003>, 2021.
- UK CEH (UK Center for Ecology & Hydrology). (n.d.) Major Hydrological Events (<https://nrfa.ceh.ac.uk/occasional-reports>)
- 505 (last access on 30 May, 2024).
- Van Loon, A. F.: Hydrological drought explained, *Wiley Interdiscip. Rev. Water*, 2, 359–392, <https://doi.org/10.1002/WAT2.1085>, 2015.
- Wei, J., Wei, Y., Western, A., Skinner, D., and Lyle, C.: Evolution of newspaper coverage of water issues in Australia during 1843–2011, *Ambio*, 44, 319–331, <https://doi.org/10.1007/s13280-014-0571-2>, 2015.



510 Wilhite, D. A. and Glantz, M. H.: Understanding: The drought phenomenon: The role of definitions, *Water Int.*, 10, 111–120, <https://doi.org/10.1080/02508068508686328>, 1985.

Zhang, T., Shen, S., Cheng, C., Su, K., and Zhang, X.: A topic model based framework for identifying the distribution of demand for relief supplies using social media data, *Int. J. Geogr. Inf. Sci.*, 35, 2216–2237, <https://doi.org/10.1080/13658816.2020.1869746>, 2021.

515