



OPEN ACCESS

EDITED BY

Spyros Foteinis,
Heriot-Watt University, United Kingdom

REVIEWED BY

Gopal Krishan,
National Institute of Hydrology (Roorkee), India
Yolanda J. McDonald,
Vanderbilt University, United States
Amy N. Podraza,
Vanderbilt University, United States in
collaboration with reviewer YM

*CORRESPONDENCE

Catherine Christenson,
✉ cchristenson@usgs.gov
Jennifer Murphy,
✉ jmurphy@usgs.gov

RECEIVED 02 May 2025

REVISED 01 October 2025

ACCEPTED 30 October 2025

PUBLISHED 28 November 2025

CITATION

Christenson C, Murphy J and Ortiz J (2025)
Exploring the news media and scientific
conversations around water quality in a water-
rich basin of the United States.
Front. Environ. Sci. 13:1622251.
doi: 10.3389/fenvs.2025.1622251

COPYRIGHT

© 2025 Christenson, Murphy and Ortiz. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Exploring the news media and scientific conversations around water quality in a water-rich basin of the United States

Catherine Christenson^{1*}, Jennifer Murphy^{2*} and Jaqueline Ortiz²

¹United States Geological Survey, Upper Midwest Water Science Center, Madison, WI, United States, ²US Geological Survey, Central Midwest Water Science Center, Madison, WI, United States

Community concerns about water availability vary depending on local economic, regulatory, environmental, and ecological considerations. In water-rich basins, water quality is often the focus of community concerns. As such, understanding community priorities in the context of water quality is crucial for informing scientists working in water-rich basins. In this work, we compiled over 6,500 local news articles (public discourse) and 190 scientific abstracts (scientific discourse) related to water-quality issues in the water-rich Illinois River Basin (ILRB) published between 2018 and 2022. We applied a Structural Topic Model (STM) to identify key water-quality topics within both datasets and explore the variability of newspaper topics geographically across the basin. Prevalent topics in both the public (local news articles) and scientific (abstracts) discourses were agriculture, drinking water quality, PFAS (per- and polyfluoroalkyl substances), and river ecosystem/fish. Topics exclusive to public discourse included water infrastructure, community development, and public water supply, while the scientific discourse focused more heavily on a wider range of agricultural issues. Furthermore, the public discourse varied geographically across the basin. Some topics are correlated with land use or urban/rural divides within the basin, and the frequency of many topics clearly varied across state (political) boundaries. Understanding and quantifying public and scientific discourses related to water-quality are important for scientists and water managers working in the basin to improve communication of critical science to the public.

KEYWORDS

Illinois River Basin, structural topic model (STM), drinking water quality, public discourse, news analysis

1 Introduction

In water-rich basins, which typically have plentiful water resources (both surface water and groundwater) and receive substantial rainfall, water availability is often controlled by the quality of water, in the form of water's chemical and physical characteristics as opposed to the overall quantity of water. The effects of climatic and socioeconomic changes on water quantity are well studied, however less is known about their potential effects on water quality and quantity-quality interactions. As such, future climatic and socioeconomic effects on water quality are an often-unaddressed source of additional stress on water availability. In a rare example of incorporating water-quality issues (in this case nitrogen pollution) in water availability modeling, Wang et al. (2024) estimated that globally in 2010,

2.5 times the number of sub-basins were affected by quality- and quantity-induced water scarcity compared to quantity-induced scarcity alone. Furthermore, the authors found projected changes to the climate and socioeconomic factors may lead to quality-induced scarcity effecting an additional three billion people.

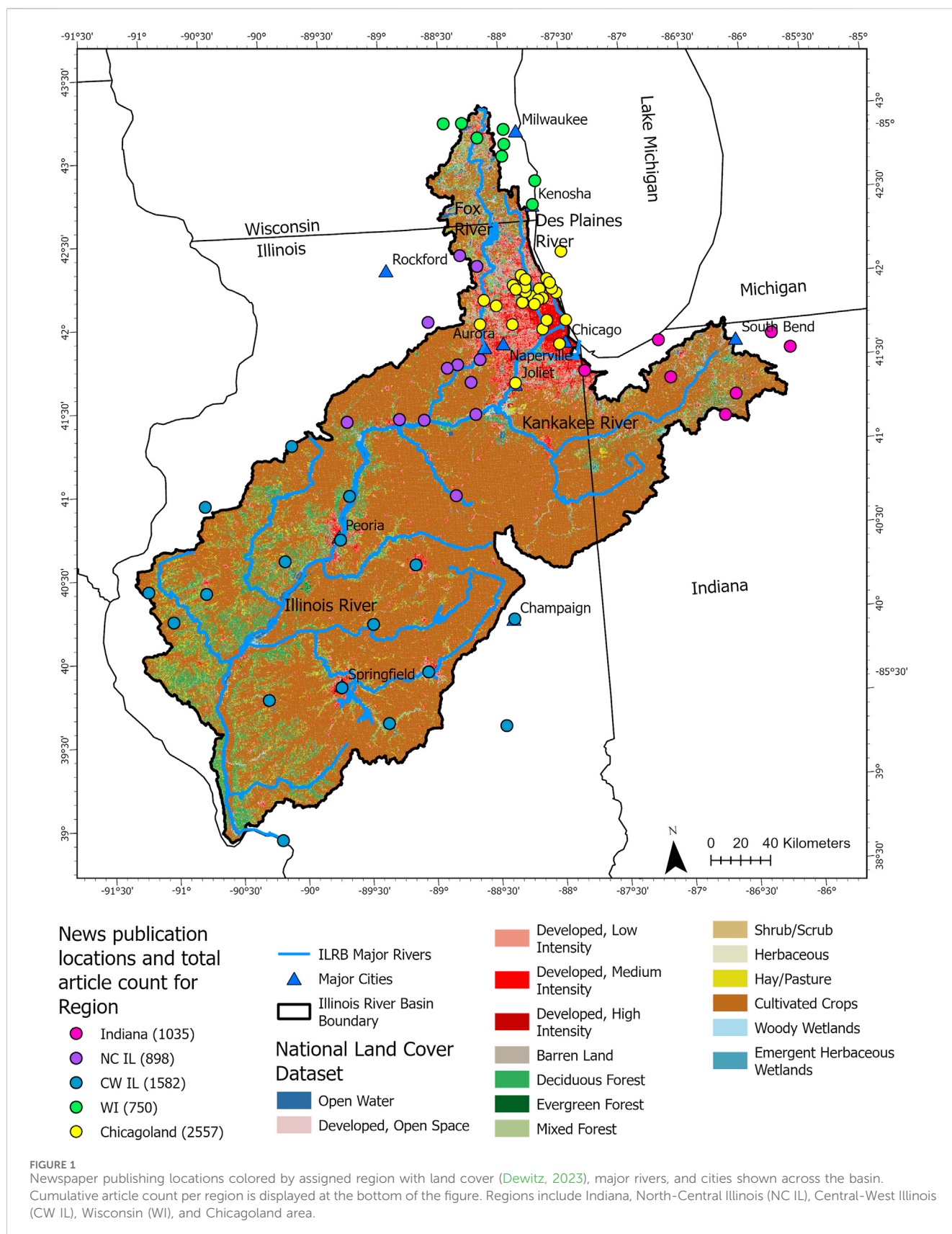
Water-quality stressors are numerous and can be characterized by the presence, concentration, or duration exceeding a threshold of common constituents (e.g., nitrogen species, sodium, chloride, and manganese), toxic contaminants (e.g., arsenic, PFAS, and microcystin), microbes (e.g., *Escherichia coli*, enteroviruses, and *Giardia lamblia*), and physical conditions (e.g., water temperature, specific conductance, and turbidity). The importance of these stressors depends on the intended water use (Van Vliet et al., 2017) and can produce undesirable effects when stressors and use conflict. Some examples include elevated salinity corroding domestic and industrial water supply infrastructure (e.g., Stets et al., 2018), higher water temperatures limiting the cooling ability of thermoelectric powerplant (e.g., Miara et al., 2013), increased frequency and (or) duration of dissolved oxygen concentrations below survival thresholds for fish and other aquatic organisms (e.g., Ficklin et al., 2013), and elevated presence of bacteria, viruses, or protozoa in groundwaters leading to gastric distress and water-borne diseases for users of domestic water systems (e.g., DeFelice et al., 2016). The geographic location of water-quality issues and a community's ability to respond and mitigate undesired consequences of acute or chronic water-quality problems are influenced by demographic and economic factors (Sohns, 2023). Public perception may also play a role in a community's response. The importance of different water-quality issues is expected to vary geographically depending on demographic and environmental characteristics (Uslu et al., 2024). Identifying water-quality issues in public conversations and where they occur geographically can provide scientists and public policymakers with knowledge of the most concerning issues for a community (Caballero et al., 2022). Furthermore, the absence of specific water-quality issues in public conversations can alert scientists and policymakers to important issues that have not been introduced to the public (The Future Water Agenda, 2025).

Newspaper media (both published online and in print) can be critical to influencing public knowledge and opinion on scientific topics, so understanding what the environmental issues are and how they are portrayed in the news can be an important indicator of public exposure to scientific issues (Boykoff and Boykoff, 2007; Corbett and Durfee, 2004). Despite the rise in digital media, local and regional newspapers still play an important role in informing citizens about local issues, risks, and management (Schulz, 2021), tracking societal values (Wei et al., 2017), and influencing citizens' opinions on topics by evoking positive or negative reactions (Rameshbhai and Paulose, 2019). Topics published by the newspaper media can influence which issues are considered most salient among their readership (McCombs et al., 2014), and news coverage has been found to be associated with public behavior related to environmental issues (Quesnel and Ajami, 2017). Furthermore, newspaper media can magnify, simplify, or downplay issues related to risk or management of resources, thereby influencing public perception and decision-making. Therefore, newspaper articles are an important source of content and can be used to identify public conversations around specific

topics. While many researchers have analyzed the content of newspaper media related to climate change and environmental risks, only a few have conducted content analysis of newspaper coverage of water to gain insights into public exposure and perception of water-related issues (using either manual or automated methods), specifically in the United States (U.S.; Mayeda et al., 2019; Sohns, 2023; Sweitzer et al., 2023; Treuer et al., 2017; Caballero et al., 2022). None that we are aware of have focused on a regional watershed, as other studies have followed national or state-level jurisdictional boundaries.

In contrast to manual reading and coding individual documents (e.g., newspaper articles) by a team of researchers, topic modeling is an automated statistical approach within the field of Natural Language Processing and is used to synthesize large volumes of text to extract latent topics or themes (Valdez et al., 2018). The Latent Dirichlet Allocation (LDA) topic model is an unsupervised generative probabilistic model that analyses word occurrence within documents (Blei et al., 2003). LDA has been widely used to study research trends in a variety of academic fields (Rahman et al., 2022; Christenson and Cardiff, 2024; Sakshi and Kukreja, 2023; Clare and Hickey, 2019; Tounsi and Temimi, 2023; Yu and Xiang, 2023). A Structural Topic Model (STM) is an extension of the LDA model that allows covariates to be assigned to documents, thereby allowing the user to explore variations in topics across factors such as publication outlet, author, date, etc. (Roberts et al., 2019). STM has been used to conduct content analyses on newspaper articles in various fields (Chandelier et al., 2018). In the field of socio-hydrology, STM was used to identify drinking water pollution topics in North Carolina from scientific abstracts (Sohns, 2023) and to explore conflict and cooperation dynamics related to water issues in the Lancang-Mekong River Basin from newspaper articles (Wei et al., 2021). Additionally, Sweitzer et al. (2023) used STM to analyze over 1.8 million newspaper articles across the United States that included the keyword "water" and identified regional and temporal variations in discussions of water. A key finding was that discussions related to water quality contaminants are more prevalent in the water-rich regions of the upper Midwest and Northeastern U.S.

The Illinois River Basin (ILRB), which receives on average 30–35 inches precipitation a year, also receives water diverted from Lake Michigan via the reversal of the Chicago and Calumet Rivers as well as the pumpage out of the Great Lakes Basin as effluent, is considered a water-rich basin where water-quality concerns play an outsized role on water availability compared to water supply (National Weather Service, n.d.; Lake Michigan Diversion Accounting Program, 2024). The ILRB encompasses 18 million acres of land, drains 44% of the land cover within Illinois (in addition to portions of Wisconsin, Indiana, and Michigan) and ultimately discharges to the Mississippi River (Figure 1; Talkington, 1991). The main "headwater rivers" of the basin include the Fox River, which drains a fast-growing, converting-to-urban and suburban area in the upper water shed, the DuPage River and Des Plaines River, which flow through the Chicago metropolitan area (population of almost 10 million people), the Kankakee River which is largely agricultural and drains northwestern Indiana, and the Chicago Sanitary and Ship Canal (CSSC) which opened in 1900 and diverts several thousand cubic feet per second of Lake Michigan water to the Illinois River (Talkington, 1991). Geographically the northern portion of the



basin is highly developed with urban and suburban structures and related infrastructure while the middle and southern portions of the basin, and the Kankakee River, are highly agricultural.

Given the diverse land uses within the ILRB, water use and community concerns related to water quality are anticipated to vary across the landscape. For example, the City of Chicago relies on

municipal water drawn from surface water in Lake Michigan, while nearby municipalities draw water from the Fox River and the Illinois River. Some cities near Chicago and further south rely on municipal groundwater supplied. The City of Joliet, a southwestern suburb of Chicago, illustrates how water quality confounds a water supply issue: groundwater levels in the deep sandstone aquifer used as the drinking water supply for Joliet are declining and chloride contamination in the shallow dolomite aquifer necessitates that the City secure another source of water by 2030 (Abrams and Cullen, 2020). Private, domestic wells provide another example of how water quality influences water availability. In the middle and southern portion of the basin, residents in rural areas generally rely on relatively shallow domestic wells, which are not regulated and regularly tested with the same frequency as municipal water systems, making the users particularly susceptible to pollution. As a final example, concerns about point source versus nonpoint source pollution relates to urban and agricultural lands, respectively. The Clean Water Act (CWA) regulates point sources discharges into navigable waters, for example, treated wastewater effluent from greater Chicago area treatment plants. Agricultural runoff, in contrast, is considered nonpoint source pollution and is not direction regulated under the CWA. The Safe Drinking Water Act (SDWA) applies to water suppliers who must treat and monitor nitrate levels (despite the source of nitrate) before delivering water to customers (U.S. Environmental Protection Agency, 1974). The distribution and source of nitrate concentrations within the Illinois River and its headwater tributaries provide an illustration: nitrate concentrations are the highest in the northern urban portion of the basin due to large amounts of treated wastewater from the Chicago area and these concentrations tend to decrease downstream but nitrate from nonpoint sources, such as synthetic fertilizer, become more common (Panno et al., 2008). As water resource managers and scientists balance complex goals within the ILRB, it is important to understand the geographic variations in priorities related to water quality.

The ILRB provides a case study to identify key water quality topics occurring in public conversations, explore these topics geographically across the basin, and compare these topics to scientific conversations. Our objectives are to:

1. Determine key water quality topics occurring in newspaper articles published between 2018 and 2022 from publishers that are located within or near the Illinois River Basin.
2. Assess the variability of these topics spatially across the basin using five geographic regions (Chicagoland, Wisconsin, Indiana, North-Central Illinois, and Central-West Illinois) based on the location of the publisher.
3. Identify key water quality topics occurring in scientific abstracts from research projects occurring within or near the Illinois River Basin published between 2018 and 2023, and compare these to the water quality topics identified via newspaper articles.

The topics derived from the newspaper articles and scientific abstracts serve as proxies for public and scientific conversations, respectively, regarding water quality in the ILRB. We use a STM to examine these public and scientific topics. By illuminating public

conversations around water quality in this water-rich basin, we identify topics of concern and interest among the public and describe how uniform or dispersed these topics are spatially. Furthermore, analyzing the overlap (or lack thereof) between public and scientific conversations is intended to reveal pressing water quality issues and those that may not have been effectively communicated between scientists and the public (or *vice versa*). Addressing water availability is a multifaceted challenge, and understanding the social landscape surrounding these topics can better inform research, outreach, and strategies for tackling environmental problems.

2 Materials and methods

2.1 Newspaper corpus

To identify newspaper articles related to water quality in the ILRB, we explored multiple online newspaper archives and iterated on the query request during these explorations. We ultimately selected Access World News from NewsBank as the newspaper archive because it had the largest local newspaper collection for Illinois, producing thousands more results than other archives (Access World News, 2023). The search query used within the database was: (“surface water” OR “water” OR “lake*” OR “river*” OR “stream” OR “streams” OR “groundwater” OR “ground water” OR “aquifer*” OR “well*”) near5 (“quality” OR “study” OR “evaluation” OR “monitoring” OR contam* OR “pollution” OR “exceedance” OR “degrad*”). The intention of this search was to identify articles that reference water quality of either surface- or groundwater within the basin, and to achieve this our search query format used two key words (one related to a water feature, and one related to a quality term). The near5 operator requested the terms to be within five words of one another; the * is used as a wildcard. We ran the query to yield related newspaper articles published within the ILRB during the five-year period, 2018–2022.

Before compiling the final set of newspaper articles, we ran the query to identify the newspaper publishers that were not located within the ILRB and thus needed to be removed from the query. This exploratory query was run for the entire states of Illinois, Indiana, and Wisconsin from 2000 through September 2023. The city of each publisher was used to determine proximity to the ILRB; publication locations that fell within 40 miles of the basin were included in the search to allow a consistent geographic buffer of areas that may be reporting on overlapping issues close to the basin. Publishers that fell further than 40 miles from the ILRB were excluded in the final query.

Once the final query was run, a subset of the data was reviewed manually to assess what further preprocessing steps to filter the dataset would be necessary in the analysis. This full data set compiled from the final query resulted in 8,549 articles. For each of the three states in the ILRB (WI, IL and IN), we reviewed 5% of the articles (430 articles in total) to gauge their relevance to water-quality topics. This percentage was distributed equally across articles occurring in the beginning, middle, and end of the compiled period. When reviewing the articles, we found that the non-related articles primarily related to agriculture, farming, or government projects, with limited, brief references to water quality. There were also

articles that contained water-related terms such as a proper noun, either as a name or location name (e.g., Lake County, Water Street). This gave us an introduction to the topics found in our database of newspapers. Relevant articles were less common than non-relevant articles in these sets within the broad dataset, with relevant articles comprising 23%, 18%, and 36% of the quality control articles (the 5% that were reviewed) articles from Indiana, Illinois, and Wisconsin, respectively. Due to the high degree of non-water-quality-related articles within the dataset, we followed a similar approach utilized by Sohns (2023) to employ a two-step modeling process to filter the dataset toward topics of interest (discussed in detail in the *Structural Topic Modeling (STM)* workflow subsection of the methods).

2.2 Basin region subdivision

Geographic categories were assigned to break the dataset into regions to later use as a covariate in the model. The ILRB was divided into five regions: Wisconsin, Indiana, North-Central Illinois (NC Illinois), Central-West Illinois (CW Illinois), and Chicagoland. Since most of the articles are from Illinois, we decided to further divide the Illinois region. We labeled Chicagoland as the cities that fell between the city and a 40-mile radius to the Chicago Loop, roughly following the overall transition in land from developed (any intensity) to more predominately cropland (Figure 1). Additionally, we used the area near Starved Rock State Park to act as a boundary between North-Central and Central-West Illinois regions. The locations of newspaper publishers within our study area, as well as the cumulative number of articles in the corpus for each region, are shown in Figure 1.

2.3 Academic corpus

We compiled a dataset of scientific abstracts from published journal articles related to water quality in the ILRB to provide a comparison of the newspaper portrayal of water quality issues to scientific portrayal. We utilized the SCOPUS database, which broadly includes peer-reviewed journal articles, conference proceedings, books and book chapters, searching all sources within the database over a five-year period between 2018 and 2023 (Scopus, 2024). We used a very similar search criteria across SCOPUS as we did for the newspaper articles, with the addition of including state names to geographically locate research taking place in the basin. The full search criterion was applied within the article title and abstract. The keywords we used were: (Illinois OR Wisconsin OR Indiana) AND ((“surface water” OR “water” OR lake* OR river* OR “stream” OR “streams” OR “groundwater” OR ground water” OR “aquifer” OR “well*”) W/5 (“quality” OR “monitoring” OR “contam*” OR “pollution” OR “exceedance” OR “degred*”). This was intended to be spatially inclusive of the states across which the ILRB spans and resulted in 557 results. As previously noted, there is a very small portion of Michigan included in the ILRB, but we did not include Michigan in this search due to the small surface area and sparse population in this area. Each title and abstract were manually reviewed to determine whether the study area took place within the ILRB or not, as well as its relevance to water quality. Almost all the articles removed related

to water quality studies in Wisconsin and Indiana, but not in the ILRB, which was expected using this state inclusive search. After manual removal of non-geographically relevant abstracts, 190 academic abstracts published in 121 discrete academic journals and conference proceedings remained.

2.4 Structural topic model (STM) workflow

Raw article portable document files (PDFs) were preprocessed using the *Python (version 3.13)* programming language to extract the text of the article body and associated metadata (publication date and publisher). Given that our initial quality control review revealed a substantial number of articles not entirely related to water quality, we did a first pass of our modeling to filter the dataset following Sohns (2023) to articles more related to water quality. We ran the full dataset of 8,549 articles through the LDA model using 30 topics to assess baseline topics within the dataset (Supplementary Figure 1). 12 of the 30 topics were determined by the authors to not be related to water quality based on the top associated keywords. For each document within the corpus, a top-aligned topic can be identified. We filtered out articles for which the most highly related topic was one of the 12 determined to be unrelated to water quality. The filtered dataset contains 6,822 articles. The STM model was conducted using the STM package in the *R (version 4.4.1)* programming language (Roberts et al., 2019). The code and metadata used to produce these model results can be found in Christenson et al. (2025).

2.4.1 Topic number selection

The framework provided in Weston et al. (2023) was followed to determine the number of topics used in the STM model. Model metrics, including exclusivity, variational lower bound, residuals, and semantic coherence, were calculated for possible model results between a range of 10 and 40 topics (Figure 2) for the newspaper dataset. This range was chosen at the authors discretion as representing a reasonable range for interpretability. Among these metrics, better model solutions are reflected by lower residuals and higher exclusivity, semantic coherence, and variational lower bound. Previous research found that increasing the number of topics improves fit metrics except for semantic coherence, so it is therefore important to balance these metrics in the selection process and look for local maxima metrics or select models that provide substantive improvement in fit relative to preceding models (Fu et al., 2021). Three model candidates (14, 24, and 28 topics) were selected for further evaluation based on the fit statistic results show in Figure 2. It is important to note that when selecting the number of topics used in topic model, there is no “ideal” number of topics for the corpus, but are likely several good options, each of which may be useful for different research objectives. Choosing the ultimate topic model number is a somewhat subjective process (Weston et al., 2023). Model results were analyzed for each 14, 24, and 28 topic number solutions. Upon reviewing 14 topics, it was clear that more specificity in the topics was needed for improved resolution of this research objective. The final model of 28 topics was selected, as it provided more information than 14 and 24 topic number solution, and better topic interpretability. Topic names (which are assigned to select topics within the presented results) were assigned by the

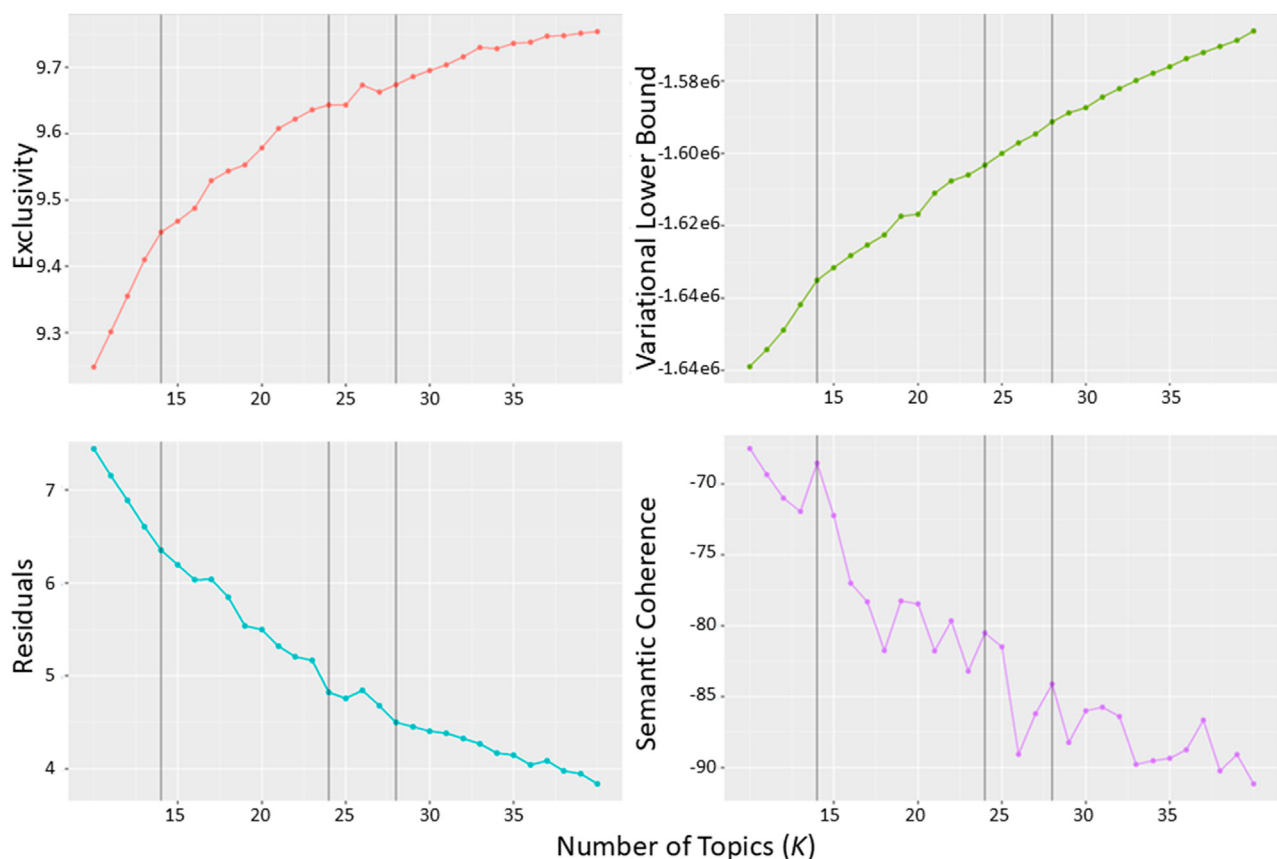


FIGURE 2
Model metrics including exclusivity, variational lower boundary, residuals, and semantic coherence (which are dimensionless values) were calculated for every solution between 10 and 40 topics. Bars identify candidate models that were considered for final model selection.

authors based on keywords within the topics and top associated documents from the corpus, which were derived using the *findThoughts* function in the STM package.

For the scientific abstract dataset, a similar approach was taken to optimize model metrics to select the topic number for this dataset. Due to the much smaller dataset size, model metrics were calculated for every two model numbers between 4 and 20 topics (Supplementary Figure 2). Among these results, 14 topics clearly optimized semantic coherence while minimizing residuals, and resulted in interpretable results. Solutions for 12 and 16 topics were also qualitatively evaluated by the authors, and provided minimal changes to the topics. The dataset, code and metadata used to produce these model results can be found in Christenson et al. (2025).

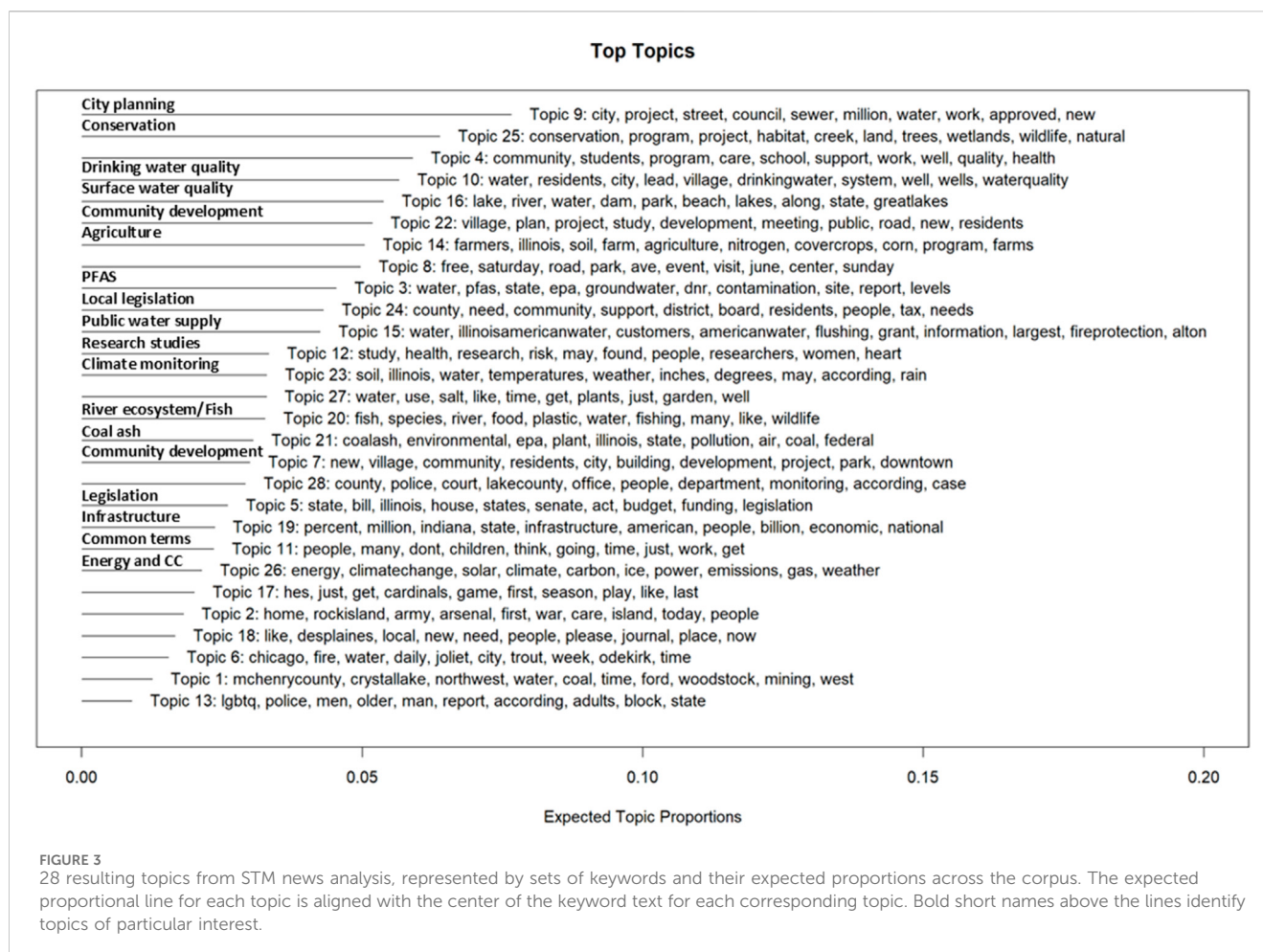
3 Results

3.1 Prevalent topics within the news corpus

Model results, presented as sets of 10 top keywords (words that occur most frequently in articles most representative of a particular topic), are presented in Figure 3. The x-axis represents the expected topic proportion, i.e., what percentage of the news article dataset

relates to a given topic. The 28 resulting topics are ranked on the plot in order of prevalence within the corpus. The names of the topics shown to the left of the primary keywords were assigned by the authors of the paper. A few of the topics are unrelated to water quality issues such as those related to local events (T8-N (-N refers to “news”, to differentiate the topics assigned to the scientific abstracts)), sports (T17-N) and two containing common terms in English (T11-N) and localized news (T18-N). Other topics that are unnamed in Figure 3 were determined to be unrelated to water quality. These topics exist because some articles in the dataset, while in part relating to water quality, include discussions of other newsworthy topics. A few of the topics, including the single most prevalent topic found in the corpus, broadly relate to water infrastructure and management issues on different scales (T9-N, T22-N, T15-N, T7-N, & T19-N), ranging from a specific utility company providing water and wastewater services to Illinois residents (T15-N), to a more nationally focused infrastructure (T19-N). Collectively, infrastructure related topics reflect over 20% of the corpus. Two topics related to legislation, both locally and on a broader scale (T24-N & T5-N), accounting for about 7% of the corpus.

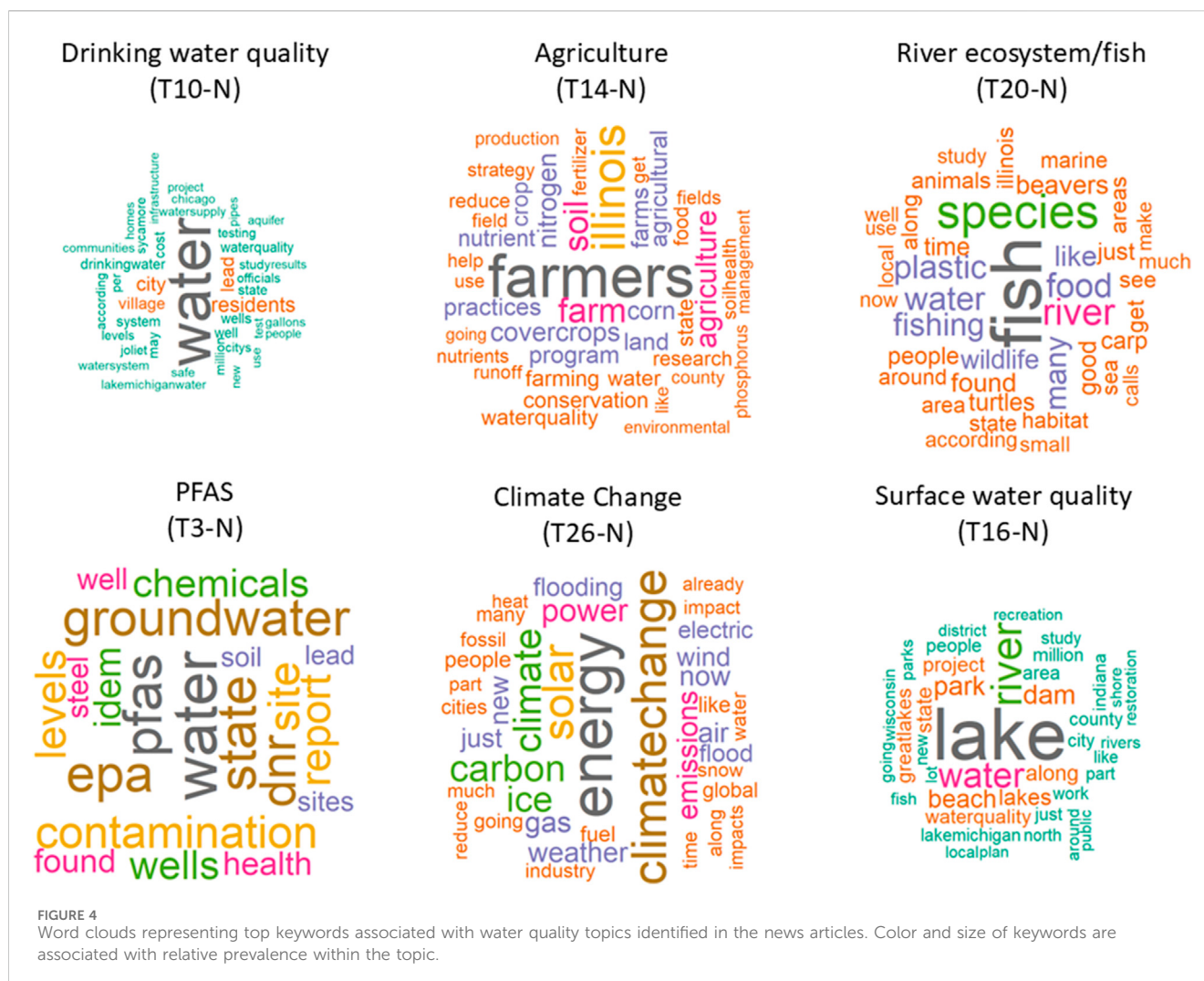
Six topics related to water quality were selected to more closely analyze, including drinking water quality (T10-N), agriculture (T14-N), PFAS (T3-N), energy and climate change (T26-N), river ecosystems/



fish (T20-N), and surface water quality (T16-N). The top keywords associated with each of these topics are conveyed in word clouds in [Figure 4](#). Word clouds for all 28 topics are provided in the [Supplementary Material](#). Drinking water quality (T10-N) ranks 4th out of the 28 identified topics, and accounts for 6% of the corpus overall. Key words such as “study results”, “safe”, “levels”, “test”, in addition to place-based words such as “Lake Michigan water”, “aquifer”, and “wells”, all common drinking water sources for those in the ILRB, suggest that this topic relates to articles conveying the monitoring and safety of drinking water sources. “Lead” is the only specific water quality contaminant observed in the top 10 keywords for the drinking water quality topic (T10-N). This topic also contains additional, less frequent (i.e., below top 10), place-based words such as “Chicago”, “Joliet”, “Sycamore”, and “communities”. Agriculture (T14-N) is the seventh most prevalent topic, accounting for 5% of the corpus overall. Interestingly, the top keywords affiliated with this topic relate to harm reduction strategies, such as “conservation,” “practices,” “reduce,” “research,” “strategy,” and phrases including “soil health” and “cover crops.” While “nitrogen,” “nutrients,” “fertilizer,” “runoff,” and “phosphorus,” which relate to contaminants and contaminant sources/mechanisms do arise - this topic is overwhelmingly focused on best management practices and nutrient reduction, rather than the harmful effects of nitrate contamination. In fact, we do not see the word contamination or pollution appear in the top words associated with

agriculture, nor do we see words such as cancer, or other health related negative societal impacts of agricultural contamination. PFAS ranks as the 9th most prevalent topic and covers about 4.5% of the corpus overall (T3-N), and includes words such as “pfas,” “contamination,” “groundwater,” “levels” and “wells,” indicating PFAS contamination in groundwater is of greater concern than surface water. Keywords referring to regulatory agencies such as “epa” (Environmental Protection Agency) and “dnr” (Department of Natural Resources) are also present in the keywords for PFAS (T3-N). Energy and climate change (T26-N), accounts for less than 3% of the corpus, and incorporates terms related to both climate change causes (“fossil,” “fuel,” “emissions”), impacts (“heat,” “flooding”) and renewable energy sources (“wind,” “electric,” “solar”). Notable other water-quality topics relate to river ecosystems/fish (T20-N) and coal ash (T21-N), each accounting for about 3% of the corpus. River ecosystems/fish (T20-N) includes references to various riverine “species” including “fish,” “turtles,” “carp,” “beavers” and “wildlife” more generally, in addition to “fishing,” and notably the keyword “plastic” appears.

[Figure 5](#) illustrates the correlations observed between topics in the news corpus, developed by adapting the *topicCorr* function in the STM package. Each node in the plot represents a topic, with the node’s size being proportional to the relative prevalence of the topic within the corpus. The weight of the edges linking nodes is relative to the correlation found between the topics. Topics without

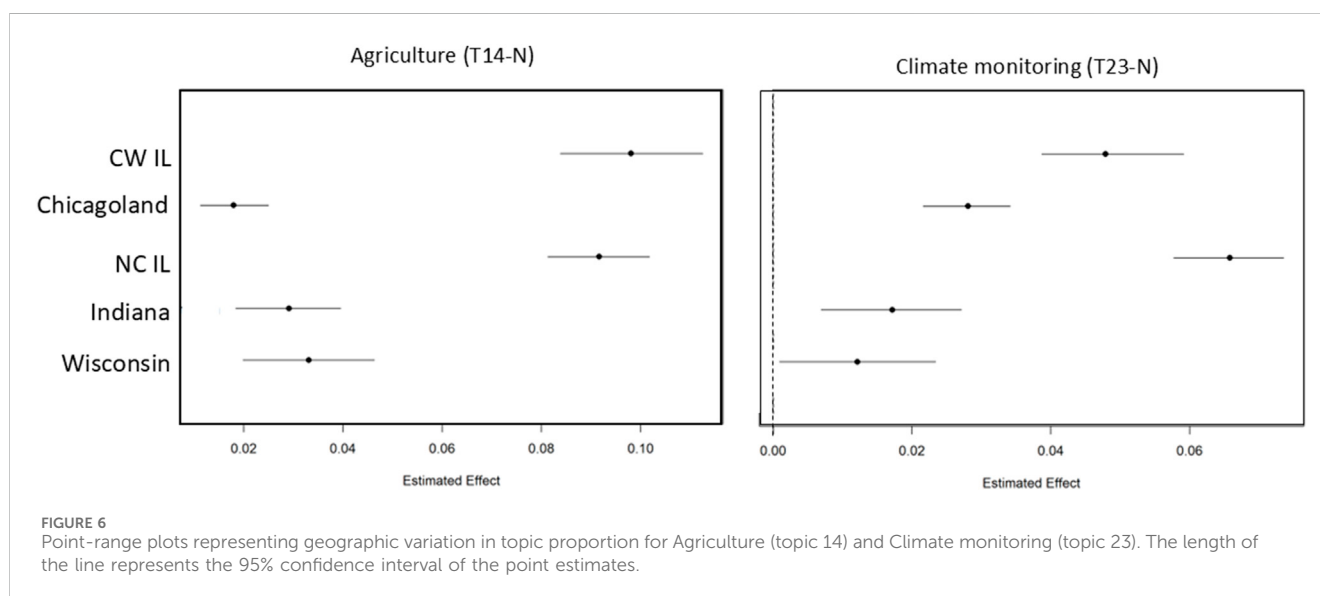
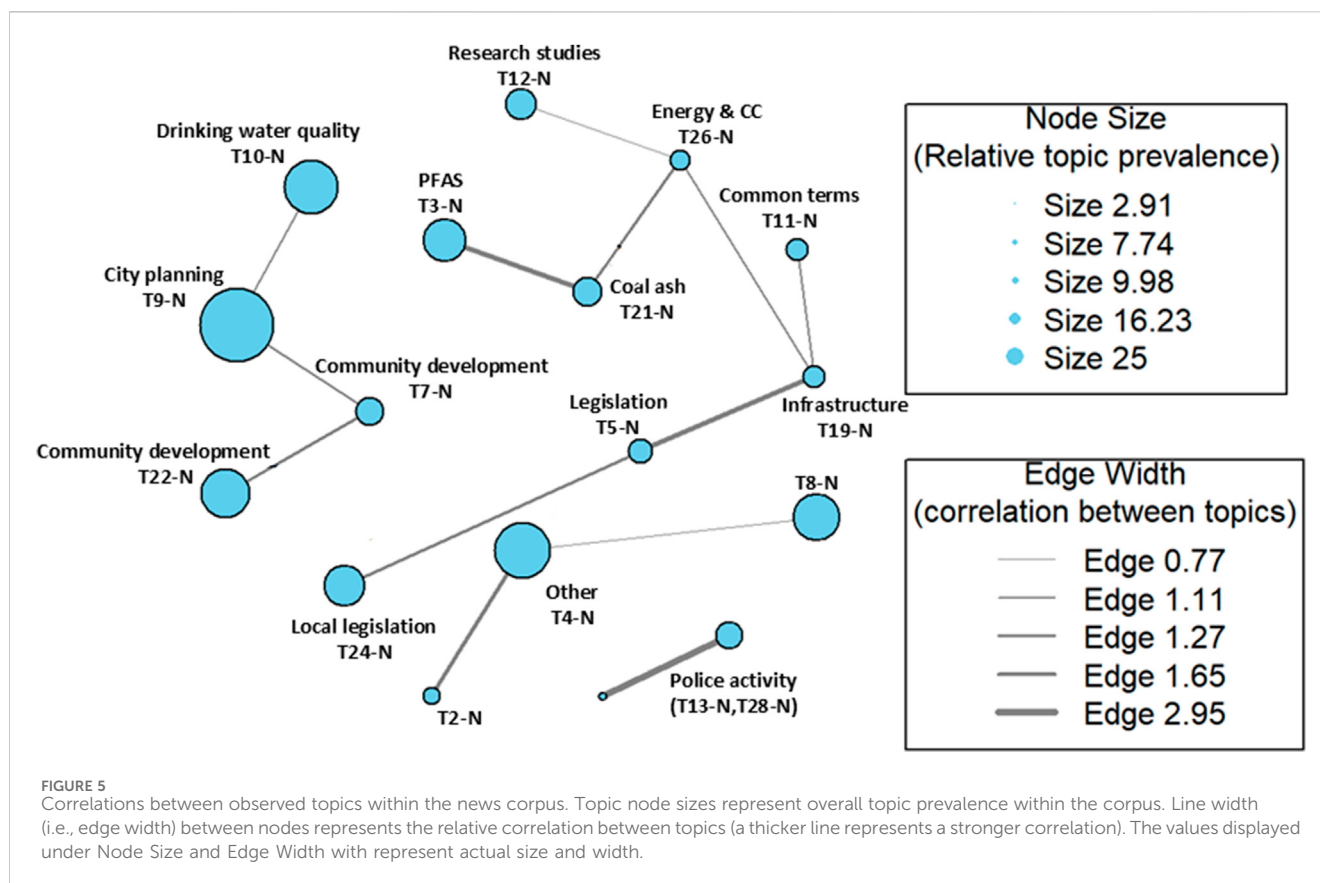


connectivity (or, without connectivity exceeding the chosen threshold) were removed from the figure. Energy and climate change (T26-N) is the topic most correlated to others, with correlations to research studies (T12-N), infrastructure (T19-N) and coal ash (T21-N), implying the interdisciplinary effects of climate change that are represented in the news. PFAS (T3-N) is correlated to coal ash (T21-N), and notably, not drinking water quality (T10-N). Drinking water quality (T10-N) is correlated with city planning (T9-N). Topics in the realm of infrastructure (T19-N) and legislation (T5-N, T24-N) are correlated, as well as community development (T7-N, T22-N) to city planning (T9-N).

3.2 Geographic variations in news results

The *estimateEffects* function in the STM R package was used to assess how newspaper publishing trends differ across five regions in the watershed (Chicagoland, Central-West Illinois, North-Central Illinois, Wisconsin and Indiana, displayed in Figure 1). This analysis is used to visualize how topic proportions vary in each region, or the percent of total news published in each region that is focused on a

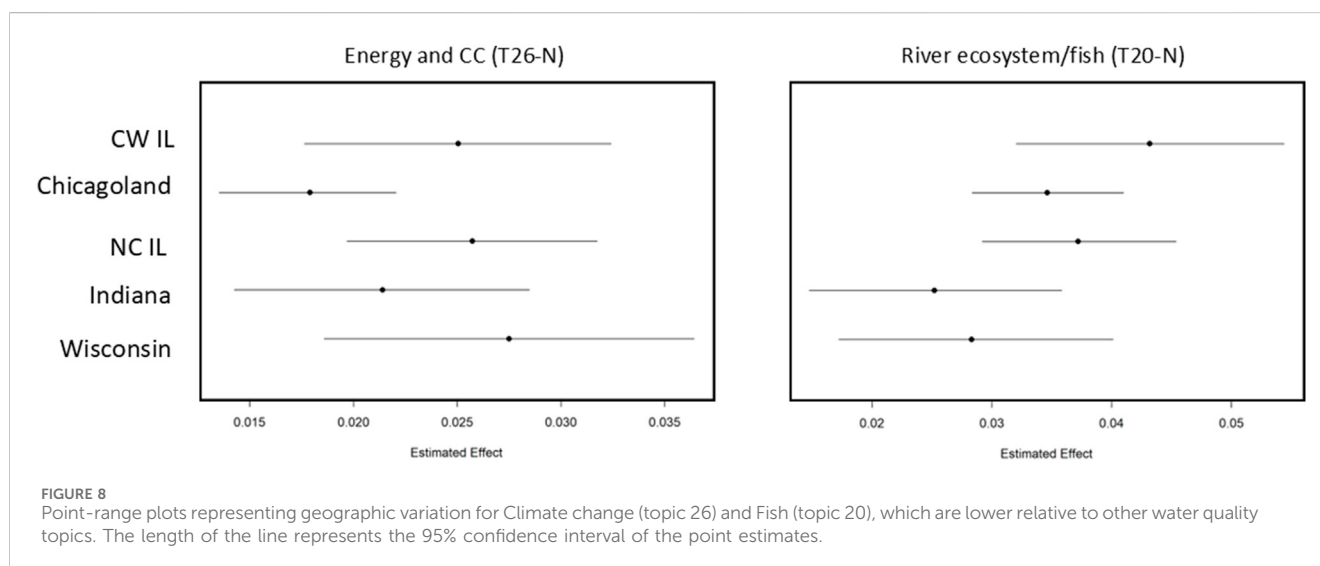
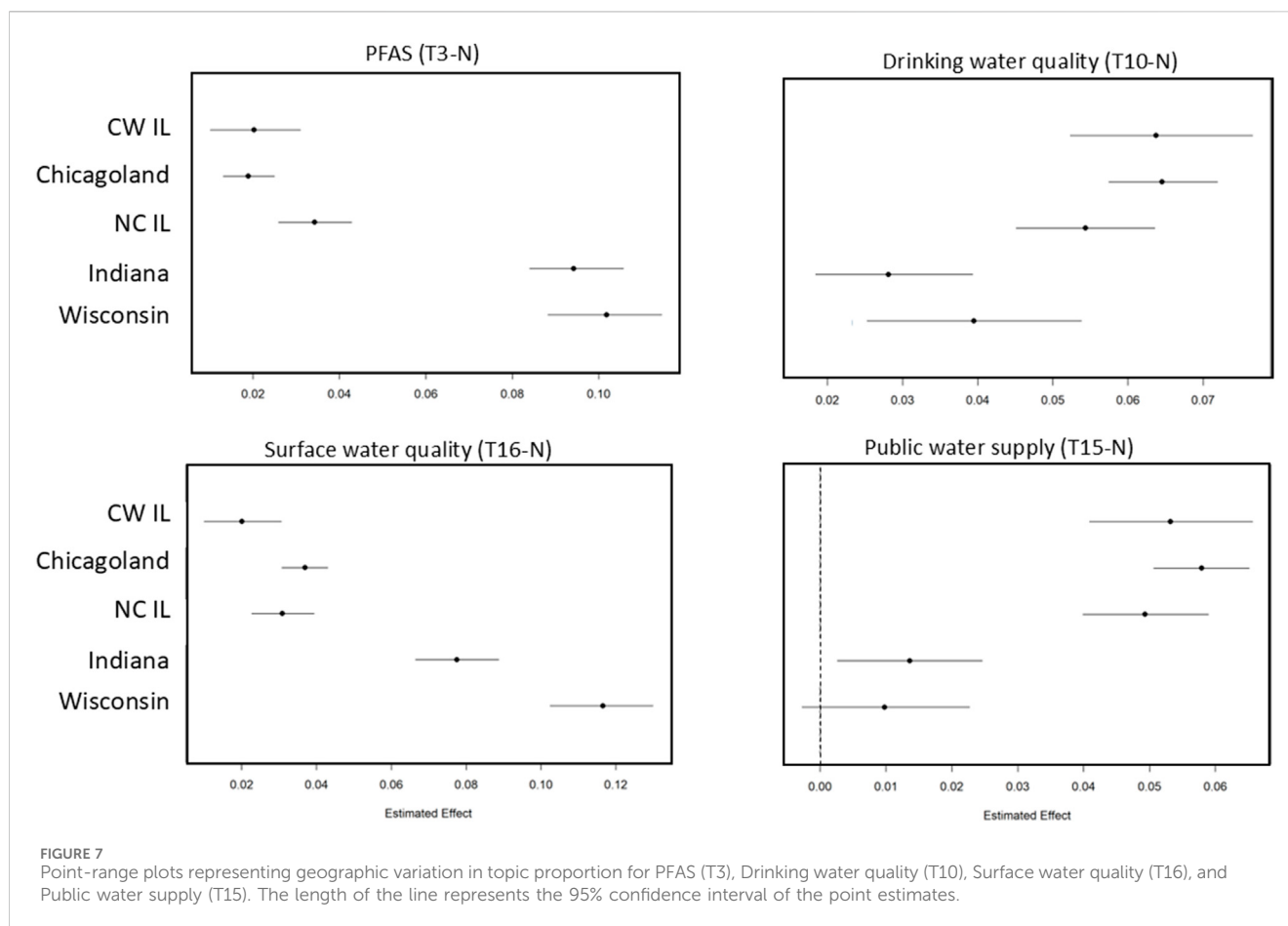
specific topic. This provides insight on how newspapers in different watershed subregions prioritize water-quality topics. Figures 6–8 illustrate how several topics vary as a portion of the newspaper corpus in the five subregions of the ILRB. Agriculture (T14-N) accounts for a much greater portion of the news in Central-West and North-Central Illinois regions (9% and 10% respectively) than it is in the other three regions of the basin (between 2% and 3%) (Figure 6). Climate monitoring (T23-N), not to be confused with the climate change topic, which includes keywords related to climate data collection such as “temperature”, “inches”, “degrees” and “rain,” follows a similar geographic trend to Agriculture (T14-N) and is most prevalent in North-Central Illinois and Central-West Illinois, followed by Chicagoland, and much less prevalence in Indiana and Wisconsin (Figure 6). Drinking water quality (T10-N) is a more prevalent topic in the three subregions that exist within Illinois (Central-West IL, Chicagoland, and North-Central Illinois) than in Indiana or Wisconsin (Figure 7). A similar state-based distinction exists for public water supply (T15-N), which is more prevalent in the Illinois subregions than in Indiana or Wisconsin. In direct contrast, both PFAS (T3-N) and surface water quality (T16-N) are more prevalent topics in Wisconsin and Indiana than in any



of the Illinois regions. Much less geographic variation is observed related to topics such as climate change (T26-N) and river ecosystems/fish (T20-N), with less absolute range of topic proportion between the geographic regions, as shown in Figure 8. The results of the *estimateEffects* function in the STM R package for regional analysis for all 28 topics is provided in the Supplementary Material.

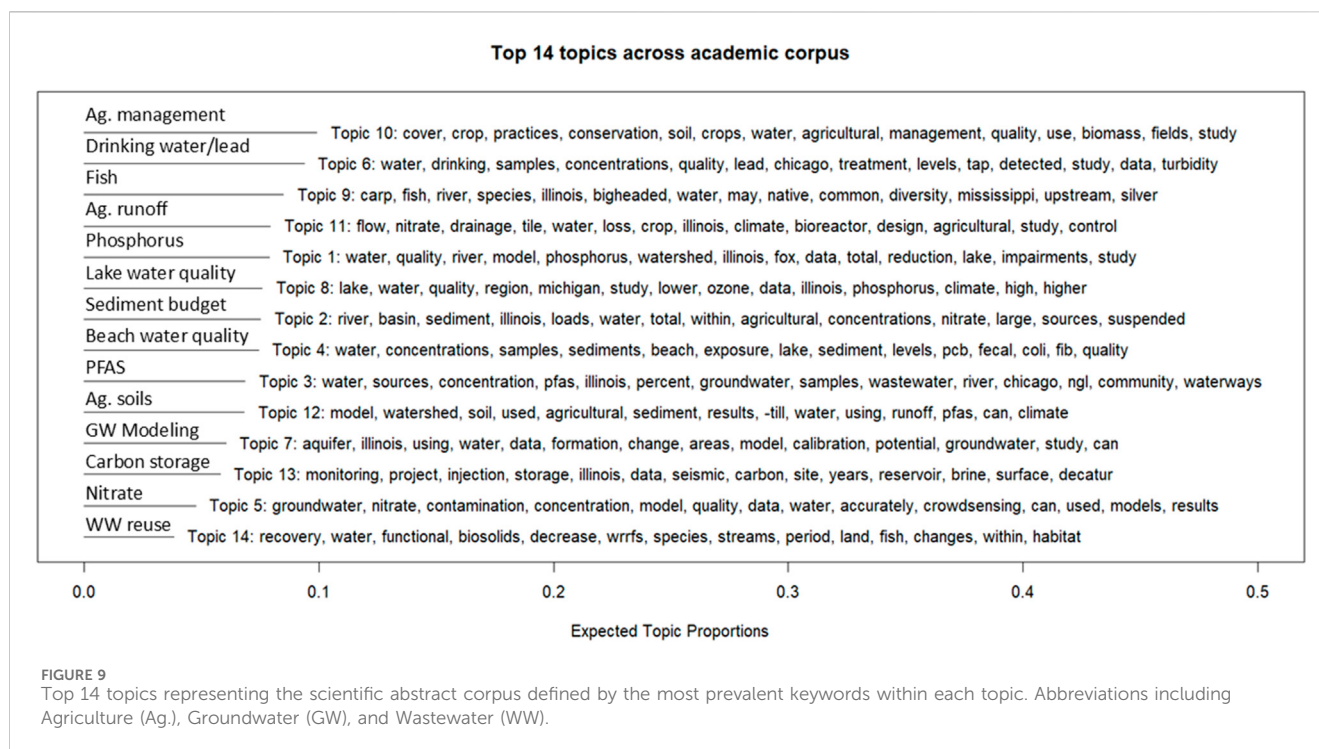
3.3 Scientific corpus results

A STM was also conducted on the corpus of 190 scientific abstracts published related to water-quality in the basin over a similar period as the collected news articles, results of which are presented in Figure 9. Fourteen topics were selected for the model as described in the methods section. The highest proportion topic



across the dataset, T10-A (-A refers to “abstracts”, to differentiate the topics assigned to the newspaper corpus), relates to ag. (agricultural) management and accounts for 10% of the scientific abstract corpus. The top words within this topic relate to best management practices such as “cover crop” and “conservation”. Two other discrete agricultural articles emerge within the dataset, ag.

runoff (T11-A) and ag. soils (T12-A), accounting for 8% and 7% of the scientific corpus respectively. While “nitrate” is a keyword within the ag. runoff (T11-A) topic, a separate nitrate focused topic also exists in the corpus. The nitrate topic (T5-A) relates more exclusively to nitrate contamination, concentrations, and water quality models. Agricultural topics (including the discrete nitrate



topic) account for 29% of the total corpus. Drinking water quality (T6-A) was the second more prevalent topic within the scientific corpus, with the key word of “lead” appearing as the top contaminant, in addition to “Chicago”. Fish (T9-A) was the third most prevalent topic, with the top keyword being “carp”, and other references to “diversity” within the top keywords.

4 Discussion

Some prominent topics identified within the water-quality news analysis were thematically aligned with results from previous media analyses related to water. For example, over 7% of the total news discussion was related to either legislation or local legislation (T24-N, T5-N). This finding is consistent with a media analysis of water-issues coverage in Australia that identified policy as a top issue discussed in news related to water (Hurlimann and Dolnicar, 2012). Legislation is inextricably tied to water issues: it is crucial in managing the use of water resources by municipalities and industries, establishing and enforcing water-quality standards, and promoting conservation practices (Milazzo, 2006). Additionally, our finding that the emphasis within a general drinking water quality topic (T10-N) relates to lead has some precedence in prior studies. A study evaluated the prevalence of water-quality issues relating to five rules in the SDWA, and found that of these rules, the lead and copper rule was by far the most discussed in news across the country (including in the region encompassing the ILRB) (Caballero et al., 2022). The lead and copper rule was reported on more than twice as much as other water-quality rules included in the SDWA.

In contrast, the absence of some important water-quality topics was unexpected. Arsenic and nitrate and have previously been

identified as the main water-quality concerns to groundwater in the ILRB (Kelly et al., 2018). While “nitrogen” is referenced as a keyword in the agriculture (T14-N) topic, arsenic is not seen in any of the top keywords across the 28 identified topics. Arsenic can be both a geogenically and anthropologically sourced contaminant, found at varying levels in drinking water across the state of Illinois (Kelly et al., 2018). Additionally, water-borne disease, bacteria, toxic algae blooms, have all been identified in previous media-analyses related to water in the U.S., and are absent from our 28 topics (Mayeda et al., 2019). Our results do not imply that these topics are not discussed at all within the dataset, but it does imply these missing topics are less prominent within local news than the 28 most salient topics displayed in Figure 3.

Two noteworthy findings from the correlation analysis are, first, among all the pairwise correlations of news topics shown in Figure 5, PFAS (T3-N) and coal ash (T21-N) have the strongest correlation. Hypotheses for this correlation are these two are groundwater contaminants often found near industrial sites; these are identified as priorities by the EPA for enforcement and mitigation (U.S. Environmental Protection Agency, 2023). Next, energy and climate change (T26-N) is correlated with three other topics (research studies (T12-N), coal ash (T21-N) and infrastructure (T19-N)), making it the most intercorrelated topic. Given that energy and climate change is a relatively less prominent topic with our results (ranked 21 of 28), we hypothesize that it is frequently discussed in the context of other topics, such as those with which it is highly correlated.

Using the analysis of topic prevalence within the news across the five geographic regions reveal two general findings: 1) some topics are correlated with land use or urban/rural divides within the basin, and 2) the frequency of many topics clearly varied across state (political) boundaries. With regards to the first finding, a much higher prevalence in agriculture (T14-N) is observed in the regions

that are more rural and have higher portions of agricultural land use (North-Central Illinois and Central-West Illinois regions, [Figures 1, 6](#)). This suggests that the scale of local news coverage within the basin may be quite localized, despite the decades-long consolidation of news media outlets ([Garz and Ots, 2025](#)). Climate monitoring (T23-N) is another topic that occurred more frequently in the two rural regions of the basin ([Figure 6](#)). In the context of this work, we are classifying Central-West and North-Central Illinois as “rural” due to the primary land use being Cropland rather than developed (either open space, low, medium, or high intensity), as shown in [Figure 1](#). One hypothesis for this is that attention to climate monitoring (such as rainfall, temperature, storms, etc.) may be of particular importance to agricultural communities who rely on climate and weather information to manage their farms (e.g., monitoring rainfall to inform irrigation management) ([Ziolkowska and Zubillaga, 2018](#)).

The second finding revealed from the geographic analysis is that the frequency of some topics varied across state, or political boundaries. This pattern is evident for the four topics displayed in [Figure 8](#). Drinking water quality (T10) and public water supply (T15) were reported more frequently in the Illinois regions of the basin than Wisconsin and Indiana. Keyword context may explain some of this trend. For example, in [Figure 3](#) there is a direct reference within the keywords for public water supply (T15) to Illinois American Water, which is the utility that provides water and wastewater services to Illinois residents ([Illinois American Water, n.d.](#)). This topic would not logically be discussed as frequently in regions (Indiana, Wisconsin) that Illinois American Water does not serve. In contrast, PFAS (T3) and surface water quality (T16) are more frequently reported on in Indiana and Wisconsin compared to Illinois ([Figure 8](#)). PFAS are more colloquially known as “forever chemicals” and are a contaminant of emerging concern that have been the subject of increasing academic study for the past 15 years ([Domingo and Nadal, 2019](#)). PFAS (T3) dominates 10% of the water-quality news related discussion in Wisconsin in contrast to 2%–3% of the discussion in the three Illinois regions of the basin. This is an unexpected finding given PFAS-related research and interest in Illinois in recent years. Recent PFAS-focused activities in Illinois range from a statewide community-supply network sampling campaign conducted by the Illinois EPA ([Illinois Environmental Protection Agency, 2025](#)) and the state legislature unanimously passing the PFAS Reduction Act in 2021, which bans and regulates the manufacturing and release of PFAS products ([Illinois General Assembly, 2021](#)). Further research is needed to assess whether the scale of PFAS-focused activities was significantly higher statewide in Wisconsin or Indiana than Illinois. Another hypothesis as to why certain topics differ by state is that local news reporting publishes not only within their local watershed, but are publishing on statewide topics; Indiana and Wisconsin publishers are likely publishing news from across the state related to water issues or studies generally. Watershed boundaries do not coincide with jurisdictional boundaries, which has been the cause for much conflict over water allocation in the U.S. historically ([Caccese and Fowler, 2020](#)), and our results suggest state identity may be as important to the water-quality news agenda as basin identity.

Examination of the differences between how news related to water quality in a relatively small area (compared to a national lens) differs from the scientific publishing taking place in the basin can

provide insight for researchers and scientists into what issues are salient and being publicly communicated. However, it is also important to consider that the scope and purpose of news media and scientific publishing differ ([Erduran, 2025](#)), and the prevalence of topics within the two datasets were not expected to completely align. There are topics exclusive to newspapers or scientific abstracts. For example, the newspaper articles exclusively contain topics related to infrastructure (T19-N), community development (T7-N, T22-N), and legislation (T5-N). The scientific abstracts exclusively contain topics such as sediment budgets (T2-A), GW modeling (T7-A), carbon storage (T13-A) and wastewater reuse (T14-A), all of which focus on scientific or environmental management processes. A previous study illustrated that a low percent (14%) of statements in water-related news coverage were supported by scientific evidence ([Hurlimann and Dolnicar, 2012](#)), and while we did not analyze statements for scientific support, the news topic model results do not include references to primary scientific methods and environmental management processes that are included in the academic abstracts. At a high level, certain areas of focus within the newspaper articles and scientific abstracts aligned: agriculture, drinking water quality, wastewater (in the news contained in public water supply), beach/surface water quality, and PFAS all emerge as highly prominent topics within both the news and scientific datasets. Drinking water quality (T10-N, T6-A) is among the most prevalent topics in both the news articles and scientific abstracts, and the single water-quality contaminant to appear among the keywords for both datasets is lead. No level of lead in drinking water is considered safe by the American Academy of Pediatrics ([Council on Environmental Health, 2016](#)), and recent studies in Illinois have detected lead in as many as 48.3% of homes with private wells ([Geiger et al., 2021](#)). While private wells are more pervasive in rural areas, lead in drinking water is a concern in urban settings of Illinois as well. Disturbances to water distribution systems, such as water mains replacements, have been found to elevate lead levels in domestic water systems, and this has been the subject of concern and study in the Chicago water system ([Batterman et al., 2019](#)).

Topics that were generally included in both the news articles and scientific abstracts, the extent to which these topics are the subject of focus within their respective dataset differ. Agriculture is a primary example of this idea. The scientific abstracts are more heavily focused on agricultural topics than the news overall. In the newspaper corpus, agriculture (T14-N) was found to be one of the most prevalent water-quality topics, but did not supersede a number of more urban-focused topics, and was contained to a single topic among 28 accounting for less than 5% of the news corpus. In contrast, the scientific abstract corpus contained three discrete agriculture related topics relating to agricultural management (T10-A), runoff (T11-A), and soils (T12-A), in addition to two agricultural adjacent topics: phosphorus (T1-A) and nitrate (T5-A). The scientific abstracts provided a wider breath of nuance related to agriculture, describing mechanisms of contamination with more detail, as well as scientific and field-based approaches for remediation, whereas the news agricultural topic related mostly to management strategies. We do see the keyword “research” appear with the news article agricultural topic (T14-N), as well as references to “phosphorus” and “nitrogen”, which implies that the news is picking up the breadth of academic agricultural research to some degree, but we do not see any other

keywords related to academic study (such as “study”, “model”, etc.). Studies related to impacts of agriculture on water quality in the ILRB have been ongoing for decades (David and Gentry, 2000), and the news attention cycle is quite short and favors newer information, so it may not be a surprise that agriculture would not remain at the forefront of the water-quality news agenda, despite its continued importance and impact on the basin.

This work includes some limitations within its scope, and additionally introduces several opportunities for extensions of future work. While our compilation of newspaper articles provides many useful insights, it does not include all forms of media that are being consumed by the public such as digital, television, or social media, all of which are also providing water-quality information to the public. The scientific abstract dataset compiled in this text includes peer-reviewed scientific abstracts, but does not include grey literature (such as those produced by state-based organizations), abstracts submitted to scientific conferences, or scientific proposals, all of which can be considered relevant to the “scientific conversation.” While SCOPUS is a widely expansive multidisciplinary database, it is not inclusive of every single academic journal in the literature. It is possible more water-quality related articles could be found in alternative scientific databases such as PubMed or Medline. Additionally, the search query used to for the SCOPUS database may missed some relevant sources by not including contaminant specific terminology (e.g., “harmful algal blooms”), if the abstract did not also reference more general water quality terms. These contaminant specific terms were intentionally avoided to not bias the dataset towards any particular issue, but some scientific articles may be so niche that they only include very specific terminology.

An interesting extension of analysis could apply sentiment analysis methods on prominent issues related to identified topics. For example, specific proper nouns (such as water related government institutions, or treatment facilities) can be identified and then sentiment of words used to reference those nouns can be quantified, which could give insight into public sentiment around various entities (Nazir et al., 2022). It is important to note that the scope of this work spans only 6 years (2018–2023), and while the news media cycle can be short, research often takes several years from project development to publication and finally outreach of results to the public. We accounted for this somewhat by slightly broadening the period of abstract selection (from 5 years for the news articles, to 6 years for the scientific abstracts), however a future study utilizing a longer period of record for both sources could reveal further insights between the two, in addition to changes in topics over time. Additionally, the timing between scientific publication timing and news focus on specific topics of interest could be further evaluated. A factor that could influence issues of prevalence of topics within the news is climate factors, for example, whether the region is experiencing drought or flood conditions within the period of analysis. While the focus of this particular analysis relates to water-quality, and as the ILRB is generally considered a water-rich basin, it would also be interesting to expand the analysis toward water-quantity issues and assess how the balance and interaction of these issues changes based on climate conditions on a year-to-year basis. It is likely we could observe a shift in the discussion towards climate factors (such as T23-N). It can be hypothesized that scientific literature trends are less affected by year-to-year changes in climatic factors, given the long-term nature of

scientific research, whereas the news cycle is much shorter and more cyclical. Expanding the length of both the news and abstract datasets to a decadal scale and applying a similar STM analysis could provide insights into the nature and timing of these relationships.

In this work, we assess water-quality topics within local newspaper articles in the Illinois River Basin, evaluate how the prevalence of these topics varies throughout the basin, and compare the publishing trends to water-quality related scientific abstracts published in the basin between 2018 and 2022. To the best of our knowledge, this is one of the only efforts that directly compares news media and scientific publishing trends in the field of hydrology. Because news media continue to shape much of the public’s scientific understanding—and because public interest and investment influence how resources are allocated and policies are created—this work holds significant value for scientists and other stakeholders in the region who aim to address water availability challenges and effectively communicate emerging hydrologic insights. Our results illustrate a high level of crossover between scientific and newspaper media, with agriculture, drinking water quality, public water supply/wastewater, beach/surface water quality, and PFAS all appearing as prevalent water quality topics. Topics exclusive to news articles, such as infrastructure, city development, and legislation, or exclusive to scientific abstracts, such as sediment loads, hydrologic models, and carbon storage, align with differences in the publishing agendas of these two outlet types, but also imply a gap in communication between scientists and the public. Additionally, the lack of specific, yet important water quality topics (such as arsenic) in public and scientific conversations may be due to a variety of causes, such as consolidation of news organizations, influence of outside interest groups, and funding structures, only some of which scientists and the public are able to remedy. We identified subregional differences in water-quality news reporting, such an urban and rural divide in agricultural reporting, and differences in proportion of topic reporting based on state boundaries. This points to the importance of local reporting that reflects the environmental issues and challenges facing communities. However, some topics were equally prevalent over the entire study area, such as energy and climate change and river ecosystems/fish, which suggests an opportunity for improved inter-basin communication. Application of STM to a broader temporal set of news articles and scientific abstracts, as well as to water-quantity issues in the Illinois River Basin and beyond, would provide opportunities to reflect on the relevance, community interest, and responsiveness of scientific research to the public. As water availability becomes a more pressing issue in the coming years, identifying and quantifying water-quality topics in public and scientific conversations can help scientists, policymakers, and resource managers understand the salient issues within a region.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: A USGS data release accompanies this publication (Christenson et al., 2025). Requests to access these datasets should be directed to Catherine Christenson, cchristenson@usgs.gov.

Author contributions

CC: Conceptualization, Writing – review and editing, Investigation, Methodology, Writing – original draft, Visualization. JM: Project administration, Supervision, Writing – review and editing, Resources, Writing – original draft. JO: Writing – original draft, Data curation.

Funding

The authors declare that financial support was received for the research and/or publication of this article. Funding was provided by the USGS Water Availability and Use Science Program and the National Water Quality Program.

Acknowledgements

This work was conducted as part of the USGS Integrated Water Availability Assessments (IWAAAs) Program, which examines the spatial and temporal distribution of water quantity and quality in both surface and groundwater, as related to human and ecosystem needs and as affected by human and natural influences. Reviewers, including James Duncker and David Dupre (both USGS), are thanked for their thoughtful review and comments on the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

References

- Abrams, D. B., and Cullen, C. (2020). *Analysis of risk to sandstone water supply in the southwest suburbs of Chicago (ISWS contract report no. CR-2020-04; p. 59)*. Champaign, IL: Illinois State Water Survey. Available online at: <http://hdl.handle.net/2142/109174>.
- Access World News (2023). Access world news: a NewsBank database. Available online at: <https://www.newsbank.com> (Accessed June 1, 2023).
- Batterman, S. A., McGinnis, S., DeDolph, A. E., and Richter, E. C. (2019). Evaluation of changes in lead levels in drinking water due to replacement of water mains: a comprehensive study in Chicago, Illinois. *Environ. Sci. & Technol.* 53 (15), 8833–8844. doi:10.1021/acs.est.9b02590
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.* 3, 993–1022. doi:10.5555/944919.944937
- Boykoff, M. T., and Boykoff, J. M. (2007). Climate change and journalistic norms: a case-study of US mass-media coverage. *Geoforum* 38 (6), 1190–1204. doi:10.1016/j.geoforum.2007.01.008
- Caballero, M. D., Gunda, T., and McDonald, Y. J. (2022). Pollution in the press: employing text analytics to understand regional water quality narratives. *Front. Environ. Sci.* 10, 770812. doi:10.3389/fenvs.2022.770812
- Caccese, R. T., and Fowler, L. B. (2020). Reasonable use? The challenges of transboundary groundwater regulation in the eastern United States. *JAWRA J. Am. Water Resour. Assoc.* 56 (3), 379–386. doi:10.1111/1752-1688.12840
- Chandelier, M., Steuckardt, A., Mathevet, R., Diwersy, S., and Gimenez, O. (2018). Content analysis of newspaper coverage of wolf recolonization in France using structural topic modeling. *Biol. Conserv.* 220, 254–261. doi:10.1016/j.biocon.2018.01.029
- Christenson, C. A., and Cardiff, M. (2024). Where has hydrogeologic science been, and where is it going? Research trends in hydrogeology publishing over the past 60 years. *Hydrogeology J.* 32 (7), 1787–1800. doi:10.1007/s11842-024-02829-4
- Christenson, C. A., Murphy, J. C., and Ortiz, J. (2025). Structural topic models of water-quality related news articles and scientific abstracts in the Illinois River Basin. *U.S. Geol. Surv.* doi:10.5066/P1JZFCVA
- Clare, S. M., and Hickey, G. M. (2019). Modelling research topic trends in community forestry. *Small-Scale For.* 18 (2), 149–163. doi:10.1007/s11842-018-9411-8
- Corbett, J. B., and Durfee, J. L. (2004). Testing public (Un)Certainty of science: media representations of global warming. *Sci. Commun.* 26 (2), 129–151. doi:10.1177/1075547004270234

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fenvs.2025.1622251/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

LDA topic modeling results for 30 topics from initial filtering of newspaper dataset. Keywords for each topic are ranked in order of relevance to the topic.

SUPPLEMENTARY FIGURE 2

Model metrics including exclusivity, variational lower boundary, residuals, and semantic coherence (which are dimensionless values) were calculated for every solution between 4 and 20 topics for scientific abstract dataset. Bars identify candidate models that were considered for final model selection.

- Council On Environmental Health, Lanphear, B. P., Lowry, J. A., Ahdoor, S., Baum, C. R., Bernstein, A. S., et al. (2016). Prevention of childhood lead toxicity. *Pediatrics* 138 (1), e20161493. doi:10.1542/peds.2016-1493
- David, M. B., and Gentry, L. E. (2000). Anthropogenic inputs of nitrogen and phosphorus and riverine export for Illinois, USA. *J. Environ. Qual.* 29 (2), 494–508. doi:10.2134/jeq2000.00472425002900020018x
- DeFelice, N. B., Johnston, J. E., and Gibson, J. M. (2016). Reducing emergency department visits for acute gastrointestinal illnesses in North Carolina (USA) by extending community water service. *Environ. Health Perspect.* 124 (10), 1583–1591. doi:10.1289/EHP160
- Dewitz, J. (2023). National land cover database (NLCD) 2021 products. *U.S. Geol. Surv. data release*. doi:10.5066/P9JZ7AO3
- Domingo, J. L., and Nadal, M. (2019). Human exposure to per- and polyfluoroalkyl substances (PFAS) through drinking water: a review of the recent scientific literature. *Environ. Res.* 177, 108648. doi:10.1016/j.envres.2019.108648
- Erduran, S. (2025). Beyond misalignment of science in the news and in schools. *Science*. 387:eadu7468. doi:10.1126/science.adu7468
- Ficklin, D. L., Stewart, I. T., and Maurer, E. P. (2013). Effects of climate change on stream temperature, dissolved oxygen, and sediment concentration in the sierra Nevada in California. *Water Resour. Res.* 49 (5), 2765–2782. doi:10.1002/wrcr.20248
- Fu, Q., Zhuang, Y., Gu, J., Zhu, Y., and Guo, X. (2021). Agreeing to disagree: choosing among eight topic-modeling methods. *Big Data Res.* 23, 100173. doi:10.1016/j.bdr.2020.100173
- Garz, M., and Ots, M. (2025). Media consolidation and news content quality. *J. Commun.* 75 (3), 195–206. doi:10.1093/joc/jqae053
- Geiger, S. D., Bressler, J., Kelly, W., Jacobs, D. E., Awadalla, S. S., Hagston, B., et al. (2021). Predictors of water lead levels in drinking water of homes with domestic Wells in 3 Illinois counties. *J. Public Health Manag. Pract.* 27, 567–576. doi:10.1097/phh.0000000000001255
- Hurlimann, A., and Dolnicar, S. (2012). Newspaper coverage of water issues in Australia. *Water Res.* 46 (19), 6497–6507. doi:10.1016/j.watres.2012.09.028
- Illinois American Water (n.d.). Who we are [fact sheet]. *Ill. Am. Water*. Available online at: <https://www.amwater.com/ilaw/resources/PDF/about-us/illinois-amwater-WhoWeAre-factSheet.pdf>.
- Illinois Environmental Protection Agency (2025). *Illinois EPA PFAS sampling network (2020-2021) dashboard*. Springfield, IL: Illinois Environmental Protection Agency. Available online at: <https://illinois-epa.maps.arcgis.com/apps/dashboards/bd61162a7f74cfe88b6928c926416c3> (Accessed February 15, 2025).
- Illinois General Assembly (2021). SB0561: PFAS reduction bill (SB 0561). *Ill. General Assem.* Available online at: <https://legiscan.com/IL/bill/SB0561/2021> (Accessed February 15, 2025).
- Kelly, W. R., Abrams, D. B., Knapp, H. V., Zhang, Z., Dziegielewski, B., Hadley, D. R., et al. (2018). *Water supply planning: middle Illinois assessment of water resources for water supply final report (ISWS contract report nos. 2018–02; p. 124)*. Available online at: <http://hdl.handle.net/2142/101848>.
- Lake Michigan Diversion Accounting Program (2024). *Lake Michigan diversion accounting program*. Chicago, IL: US Army Corps of Engineers. Available online at: <https://www.lrd.usace.army.mil/Missions/Projects/Article/3639032/lake-michigan-diversion-accounting-program> (Accessed April 15, 2025).
- Mayed, A. M., Boyd, A. D., Paveglio, T. B., and Flint, C. G. (2019). Media representations of water issues as health risks. *Environ. Commun.* 13 (7), 926–942. doi:10.1080/17524032.2018.1513054
- McCombs, M., Shaw, D., and Weaver, D. (2014). New directions in agenda-setting theory and research. *Mass Commun. Soc.* 17, 781–802. doi:10.1080/15205436.2014.964871
- Miara, A., Vörösmarty, C. J., Stewart, R. J., Wollheim, W. M., and Rosenzweig, B. (2013). Riverine ecosystem services and the thermoelectric sector: strategic issues facing the northeastern United States. *Environ. Res. Lett.* 8 (2), 025017. doi:10.1088/1748-9326/8/2/025017
- Milazzo, P. C. (2006). *Unlikely environmentalists: Congress and clean water, 1945–1972*. Lawrence, KS: University of Kansas Press.
- National Weather Service (n.d.). *Hydrology education: climate and climate changes*. Chicago, IL: U.S. Department of Commerce. Available online at: https://www.weather.gov/lot/hydrology_education_climate (Accessed November 1, 2024).
- Nazir, A. Y., Rao, L. W., and Sun, L. (2022). Issues and challenges of aspect-based sentiment analysis: a comprehensive survey. *IEEE Trans. Affect. Comput.* 13, 845–863. doi:10.1109/TAFFC.2020.2970399
- Panno, S. V., Kelly, W. R., Hackley, K. C., Hwang, H.-H., and Martinsek, A. T. (2008). Sources and fate of nitrate in the Illinois River Basin, Illinois. *J. Hydrology* 359 (1), 174–188. doi:10.1016/j.jhydrol.2008.06.027
- Quesnel, K. J., and Ajami, N. K. (2017). Changes in water consumption linked to heavy news media 584 coverage of extreme climatic events. *Sci. Adv.* 3, 10. doi:10.1126/sciadv.1700784
- Rahman, M., Frame, J. M., Lin, J., and Nearing, G. S. (2022). Hydrology research articles are becoming more topically diverse. *J. Hydrology* 614, 128551. doi:10.1016/j.jhydrol.2022.128551
- Rameshbhai, C. J., and Paulose, J. (2019). Opinion mining on newspaper headlines using SVM and NLP. *Int. J. Electr. Comput. Eng. (IJECE)* 9 (3), 2152. doi:10.11591/ijece.v9i3.pp2152-2163
- Roberts, M. E., Stewart, B. M., and Tingley, D. (2019). Stm: an R package for structural topic models. *J. Stat. Softw.* 91, 2. doi:10.18637/jss.v091.i02
- Sakshi, and Kukreja, V. (2023). Recent trends in mathematical expressions recognition: an LDA-based analysis. *Expert Syst. Appl.* 213, 119028. doi:10.1016/j.eswa.2022.119028
- Schulz, A. (2021). *Local news unbundled: where audience values still lies*. Oxford, UK: Reuters Institute. Available online at: <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2021/local-news-unbundled-where-audience-value-still-lies>.
- Scopus (2024). Scopus: abstract and citation database. Available online at: <https://www.scopus.com> (Accessed October 15, 2024).
- Sohns, A. (2023). Differential exposure to drinking water contaminants in North Carolina: evidence from structural topic modeling and water quality data. *J. Environ. Manag.* 336, 117600. doi:10.1016/j.jenvman.2023.117600
- Stets, E. G., Lee, C. J., Lytle, D. A., and Schock, M. R. (2018). Increasing chloride in rivers of the conterminous U.S. and linkages to potential corrosivity and lead action level exceedances in drinking water. *Sci. Total Environ.* 613–614, 1498–1509. doi:10.1016/j.scitotenv.2017.07.119
- Sweitzer, M. D., Gunda, T., and Gilligan, J. M. (2023). Water narratives in local newspapers within the United States. *Front. Environ. Sci.* 11, 1038904. doi:10.3389/fenvs.2023.1038904
- Talkington, L. M. (1991). *The Illinois River: working for our State (ISWS miscellaneous publication no. MP-128; p. 57)*. Champaign, IL: Illinois State Water Survey. Available online at: <https://www.ideals.illinois.edu/items/48971>
- The Future Water Agenda (2025). “How water can lead the way for sustainability and collective action,” in *WWF and GlobeScan*. Available online at: https://globescan.wpenginepowered.com/wp-content/uploads/2025/03/The-Future-Water-Agenda-Report-GlobeScan-WWF-March-2025_Final-Version.pdf.
- Tounsi, A., and Temimi, M. (2023). A systematic review of natural language processing applications for hydrometeorological hazards assessment. *Nat. Hazards* 116 (3), 2819–2870. doi:10.1007/s11069-023-05842-0
- Treuer, G., Koebele, E., Deslatte, A., Ernst, K., Garcia, M., and Manago, K. (2017). A narrative method for analyzing transitions in urban water management: the case of the Miami-Dade water and sewer department. *Water Resour. Res.* 53 (1), 891–908. doi:10.1002/2016WR019658
- U.S. Environmental Protection Agency (1974). Safe drinking water act. *Public Law*, 93–523. Available online at: <https://www.epa.gov/sdwa>.
- U.S. Environmental Protection Agency (2023). EPA announces federal enforcement priorities to protect communities from pollution. Available online at: <https://www.epa.gov/newsreleases/epa-announces-federal-enforcement-priorities-protect-communities-pollution> (Accessed June 1, 2024).
- Uslu, A., Dugan, S. T., El Hmadi, A., and Muhammetoglu, A. (2024). Comparative evaluation of spatiotemporal variations of surface water quality using water quality indices and GIS. *Earth Sci. Inf.* 17 (5), 4197–4212. doi:10.1007/s12145-024-01389-1
- Valdez, D., Pickett, A. C., and Goodson, P. (2018). Topic modeling: latent semantic analysis for the social sciences. *Soc. Sci. Q.* 99 (5), 1665–1679. doi:10.1111/ssq.12528
- Van Vliet, M. T. H., Flörke, M., and Wada, Y. (2017). Quality matters for water scarcity. *Nat. Geosci.* 10 (11), 800–802. doi:10.1038/ngeo3047
- Wang, M., Bodirsky, B. L., Rijnveld, R., Beier, F., Bak, M. P., Batool, M., et al. (2024). A triple increase in global river basins with water scarcity due to future pollution. *Nat. Commun.* 15 (1), 880. doi:10.1038/s41467-024-44947-3
- Wei, J., Wei, Y., and Western, A. (2017). Evolution of the societal value of water resources for economic development versus environmental sustainability in Australia from 1843 to 2011. *Glob. Environ. Change* 42, 82–92. doi:10.1016/j.gloenvcha.2016.12.005
- Wei, J., Wei, Y., Tian, F., Nott, N., de Wit, C., Guo, L., et al. (2021). News media coverage of conflict and cooperation dynamics of water events in the Lancang—Mekong River basin. *Hydrology Earth Syst. Sci.* 25 (3), 1603–1615. doi:10.5194/hess-25-1603-2021
- Weston, S. J., Shryock, I., Light, R., and Fisher, P. A. (2023). Selecting the number and labels of topics in topic modeling: a tutorial. *Adv. Methods Pract. Psychol. Sci.* 6, 251524592311601. doi:10.1177/25152459231160105
- Yu, D., and Xiang, B. (2023). Discovering topics and trends in the field of artificial intelligence: using LDA topic modeling. *Expert Syst. Appl.* 225, 120114. doi:10.1016/j.eswa.2023.120114
- Ziolkowska, J. R., and Zubillaga, J. (2018). Importance of weather monitoring for agricultural decision-making—An exploratory behavioral study for Oklahoma mesonet. *J. Sci. Food Agric.* 98 (13), 4945–4954. doi:10.1002/jsfa.9027