



# Human Social Sensing for Characterizing Long-Term Water-Related Discourse Dynamics: Implications for Sociohydrological Analysis

5 Juseong Lee<sup>1</sup>, Wuseong Jeong<sup>1</sup>, Young Gu Her<sup>2</sup>, JungJin Kim<sup>3</sup>, Rabin Bhattarai<sup>4</sup>, David J. Yu<sup>5,6</sup>,  
Hanseok Jeong<sup>1,3</sup>

<sup>1</sup>Department of Environmental Engineering, Seoul National University of Science and Technology, Seoul, 01811, Republic of Korea

10 <sup>2</sup>Department of Agricultural and Biological Engineering, UF/IFAS Tropical Research and Education Center, Homestead, Florida, 33031, United States

<sup>3</sup>Institute of Environmental Technology, Seoul National University of Science and Technology, Seoul, 01811, Republic of Korea

<sup>4</sup>Department of Agricultural and Biological Engineering, University of Illinois at Urbana-Champaign, Illinois, 61801, United States

15 <sup>5</sup>Department of Civil Engineering, Purdue University, West Lafayette, Indiana, 47906, United States

<sup>6</sup>Department of Political Science, Purdue University, West Lafayette, Indiana, 47907, United States

*Correspondence to:* Hanseok Jeong (hanjeong@seoultech.ac.kr)

**Abstract.** Sociohydrological models seek to represent co-evolving human and water systems. Yet, static assumptions or poorly constrained parameters may limit their capacity to reproduce long-term transitions in societal discourse and policy priorities that shape hydrologic outcomes. To address this gap, we propose and evaluate a human social sensing (HSS) framework that converts long-term public discourse into quantitative, time-resolved proxy indicators associated with water-related discourse dynamics in sociohydrological systems. HSS integrates large-scale text mining with structured manual coding to capture scalable patterns and interpretable context in water-related narratives. We applied this framework to 61 years (1960–2020) of South Korean newspaper coverage, revealing a clear transition from disaster- and development-centered framing to sustained attention on water quality, environmental management, and ecosystem protection, alongside a tonal reorientation from economic development priorities toward environmental sustainability. Text mining identified structural changes in themes and salient terms, including the rising prominence of water-quality and water-resource themes. Meanwhile, manual coding provided an interpretable contextual benchmark for evaluating computational outputs and for identifying periods in which discourse shifts preceded institutional change. Together, the proposed framework generates high-resolution indicators of evolving water-related societal discourse and demonstrates the complementary strengths of computational analysis and human interpretation. By generating empirically derived discourse-level proxy indicators, HSS offers a structured framework for incorporating long-term societal framing and public discourse dynamics into sociohydrological analysis, while preserving interpretive caution about the indirect relationship between media discourse and public perception.



## 35 **1 Introduction**

Water management decisions heavily depend on changing social values and public perceptions (Schulz et al., 2024; Yang and Sato, 2025). These sociocultural factors evolve continuously in response to environmental change (Azadi et al., 2025), and during this evolution, societies do not remain passive recipients of hydrological extremes. Instead, they actively respond through actions such as infrastructure development, institutional reform, and cultural adaptation, thereby reshaping future hydrological risks (Coletta et al., 2024; Malede et al., 2025). Sociohydrological models are uniquely positioned to capture such feedback between human behavior and hydrological processes (Han et al., 2025; Wang et al., 2024). By simulating how human responses alter water systems, these models help us better understand and predict the co-evolution of people and water (Li and Sivapalan, 2020; Razavi et al., 2025).

Despite its promise, sociohydrological modeling still struggles to represent the adaptive and context-dependent nature of human responses (Mijic et al., 2024; Xia et al., 2022). Many models rely on static or overly simplified representations of social processes, overlooking how behavior evolves with environmental change and policy interventions (Constantino et al., 2021; Herrera et al., 2022; Javanbakht-Sheikhahmad et al., 2024; Tallman et al., 2024). Among these social processes, public perception, which refers to the collectively expressed view and evaluations of environmental conditions and management (Bennett, 2016; Lee et al., 2015), plays an important role because it influences behavioral change and policy through risk awareness, institutional trust, and interpretations of past experiences (Gu et al., 2021; Kherazi et al., 2024). Despite its importance, however, public perception remains difficult to quantify because it is not directly observable and is shaped by qualitative, dynamic, and context-dependent processes (de Brito et al., 2024; T. Xu and Liang, 2021). It is therefore typically inferred from indirect indicators including surveys or policy changes, which limit their usefulness in dynamic modeling frameworks (Muñoz et al., 2024).

Textual records such as newspaper archives provide a valuable basis for examining long-term changes in water-related public discourse, but using them to infer public perception raises three methodological challenges (Hovy and Spruit, 2016; Krippendorff, 2018). First, historical archives create a scalability challenge. They often contain large volumes of articles spanning multiple decades, making it difficult for manual reading alone to identify macroscopic patterns, recurring themes, and long-term shifts in discourse. Computational text analysis helps address this challenge by enabling faster and more systematic analysis of large corpora (Galesic et al., 2021; Miao et al., 2024). Second, however, scalability can come at the cost of interpretive validity. Public perception and environmental values are latent and context-dependent constructs, whereas textual data provide only indirect traces of how such constructs are expressed (Grubert et al., 2017; Ulibarri et al., 2019). Automated analysis, if applied without sufficient contextual interpretation, may miss the framing, tone, evaluative meaning, and historical circumstances through which water-related concerns are articulated (Hovy and Spruit, 2016; Krippendorff, 2018; Yi et al., 2025). Third, temporal patterns in text do not necessarily indicate temporal changes in public priorities. Increases or decreases in the frequency of particular terms, themes, or frames may reflect substantive shifts in societal perception, but they may also result from changes in vocabulary, journalistic conventions, archival composition, or



analytical specifications (Rule et al., 2015; Ulibarri et al., 2026). The central methodological challenge, therefore, is to analyze textual records at scale while preserving the contextual and temporal validity needed to interpret them as indicators of evolving public perception (Kenter et al., 2025; Scoones, 2016).

To address this scale–validity tradeoff, we employ a human social sensing (HSS) framework for analyzing long-term water-related discourse. HSS uses human-generated textual data, such as newspaper articles, to detect discourse-level signals that may reflect broader societal perceptions, issue framings, and behavioral orientations (Galesic et al., 2021). Prior work has shown the value of historical newspapers and manual textual analysis for examining social dynamics, evaluative framing, and water-related discourse that are difficult to observe directly (Chen et al., 2016; Wei et al., 2017). Other studies have demonstrated the usefulness of computational text analysis for identifying large-scale themes and temporal patterns across extensive corpora (Galesic et al., 2021; Miao et al., 2024). However, when applied separately, each approach remains incomplete. Manual coding preserves contextual meaning but is difficult to scale, whereas computational analysis scales efficiently but can weaken interpretive validity. In this study, we therefore use HSS as an integrative framework that combines manual coding and computational text analysis. Manual coding grounds the interpretation of water-related discourse in contextual and evaluative meaning, while computational text analysis extends those interpretations across a larger historical corpus. By linking interpretive coding with scalable text mining, the framework is designed to preserve contextual validity while producing quantitative indicators of long-term discourse dynamics.

Building on this framework, this study analyzes six decades of newspaper coverage on water-related issues to examine how public discourse has changed in structure, tone, and thematic emphasis over time. We use manual coding to establish contextually grounded categories of water-related discourse and framing, and then apply computational text analysis to extend these categories across the historical corpus. This approach allows us to derive quantitative discourse indicators that capture temporal variation in how water issues are publicly represented, evaluated, and prioritized. Rather than treating these indicators as direct measures of public perception, we use them as empirically grounded proxies for perception-related discourse dynamics. In doing so, the study demonstrates how historical textual records can be converted into interpretable, model-relevant social indicators for sociohydrological analysis.

## 2 Methods

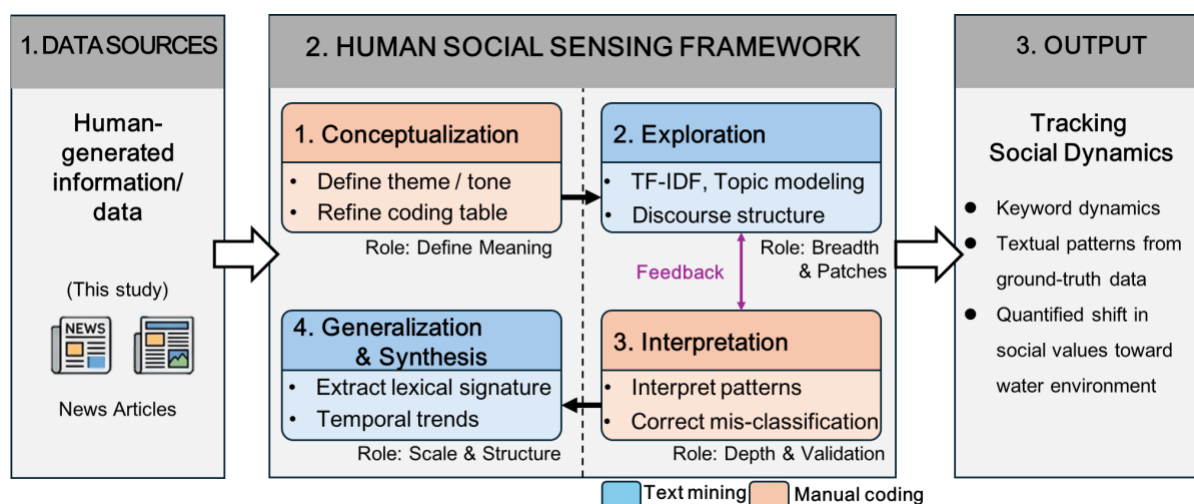
### 2.1 Overview

We operate the HSS framework to identify discourse-level indicators associated with societal priorities by integrating manual coding with computational text mining (Fig. 1). The workflow begins with a conceptualization phase, in which a structured coding table (Wei et al., 2017) is developed to define article metadata, thematic categories, and evaluative tone as the primary dimensions of water-related social values. In the exploration phase, text mining techniques are applied to explore linguistic patterns and thematic structures across the corpus. Term Frequency–Inverse Document Frequency (TF-IDF) is used to extract representative keywords across the full corpus and within subsets defined by topic and tone, while latent



100 Dirichlet allocation (LDA) is employed to uncover latent themes. The interpretation phase leverages the manually coded subset as an interpretive benchmark. Manual coding provides contextual understanding, resolves ambiguous cases, and refines category boundaries, while also guiding the validation and targeted reanalysis of computational outputs. Within each human-defined category, the integration of TF-IDF and topic modeling translates qualitative interpretations into reproducible linguistic signatures. Finally, in the synthesis phase, manually coded data and computational outputs are integrated to

105 validate the framework and examine longitudinal shifts in water-related social values. This workflow was designed not only to identify temporal patterns in water-related discourse, but also to evaluate whether computationally derived patterns were consistent with contextually interpreted categories from manual coding. In this way, the method links scalability with interpretability, which is central to the HSS framework.



110 **Figure 1. Overview of the proposed human social sensing framework applied to human-generated data. The method integrates text mining and manual coding to detect changes in social value. By comparing computational patterns against human-interpreted reference data, the framework validated indicators of social dynamics.**

## 2.2 Data

Newspaper articles were selected from two major South Korean daily newspapers, Chosun Ilbo and Donga Ilbo, covering a

115 61-year period from 1960 to 2020 due to their wide readership and long-standing publication history. The time periods for the collected news articles were allowed to track long-term trends in social values related to the water environment by covering economic development and environmental change in South Korea. The articles were sampled using a hybrid approach: two randomly selected consecutive weeks per year to ensure a non-biased baseline, and two fixed periods, the third week of July and the third week of February, representing peak and minimal precipitation weeks in a year, respectively.

120 This strategy was adopted to ensure a representative dataset that captures both hydrological extremes and year-round baseline trends, while maintaining the feasibility of in-depth manual coding over the 61-year study period. The combination of fixed hydrologic periods and randomly selected baseline periods allowed the dataset to capture both event-sensitive



discourse and broader long-term media attention. This method yielded a total of 644 articles for analysis: 319 (49.5%) collected via the fixed-period approach and 325 (50.5%) via the random-period approach.

125 **2.3 Manual Coding**

Manual coding was required because key variables, especially theme and tone, depend on context, implied framing, and culturally specific language that automated classifiers can misinterpret without strong labeled training data. It also provided a consistent interpretive benchmark for evaluating and contextualizing downstream text-mining outputs. Manual coding was conducted using a structured coding framework (Table 1). Articles were coded based on three primary dimensions: (1) metadata, such as publication date and mentioned institutions, to track temporal dynamics and the salience of key actors; (2) themes, representing the primary water-related issue addressed in each article (e.g., natural hazards, water policy, or water quality), which enables tracking of shifts in topical focus over time; and (3) tone, capturing the dominant evaluative framing of the issue.

In this study, while theme identifies what the article is about, tone reflects how the issue is framed. In this study, tone was categorized as economic development, environmental sustainability, or neutral, distinguishing articles that emphasize development-oriented narratives from those focused on environmental protection, while separating purely informational coverage without a clear value-based orientation. Treating theme and tone as complementary but distinct enables us to examine whether shifts in topical attention occur alongside shifts in framing, even when the underlying water issue remains the same. This distinction is important because public discourse may change not only in what issues receive attention, but also in how those issues are interpreted, evaluated, and connected to broader societal priorities. Articles containing only brief factual notices without sufficient contextual information for thematic or tonal interpretation were excluded from the analysis. Newspaper articles were independently coded by two graduate-level researchers with backgrounds in water and environmental studies. The coding process focused on thematic and tonal classification, and classification discrepancies were resolved through a third independent coder.

145 **Table 1. Structured coding framework for newspaper articles: items for metadata, themes, and tone.**

Variables	Items
Metadata	-Publication date: day, month, and year -Institution: organizations and individuals that are involved in the articles
Themes	Water for agriculture, water for urban use, water for the environment, natural hazards, water storage, water policy, and water quality
Tone	-Economic development: contents including the construction of water resource storage facilities, irrigation for water resource utilization, and economic development -Environmental sustainability: contents include issues such as river restoration, ecosystem pollution, and others



## 2.4 Text mining

To systematically characterize water-related discourse, we implemented a multi-level text-mining workflow that captures lexical, relational, and thematic dimensions of the corpus. The workflow consists of three steps. First, text was preprocessed  
150 to ensure consistent linguistic representation suitable for computational analysis. Second, keyword-based analyses were conducted to identify salient terms, their contextual associations, and their evolution across time. Third, topic modeling was employed to identify recurrent discourse configurations and to examine their temporal dynamics and relationships with evaluative tone. This hierarchical design enables a comprehensive analysis of how language is used, discourse patterns, and value-oriented framing co-evolve over time.

### 155 2.4.1 Preprocessing

Given the linguistic complexity of Korean (e.g., subject-object-verb structure, postpositions, and compound nouns), preprocessing was essential for enabling reliable and meaningful text mining. For example, morphological analysis was required to separate root words from attached suffixes, and specialized tokenization tools were used to accurately segment compound nouns and identify grammatical roles. The KoNLPy library and MeCab morphological analyzer (Park and Cho,  
160 2014) were utilized to tokenize text at morpheme level. Standard preprocessing steps were applied (Jurafsky and Martin, 2026), including the removal of stop words, punctuation, and irrelevant elements (e.g., media names). The stop word removal step filtered particles, common pronouns, conjunctions, filler words, and single-character tokens that carried no semantic weight for the analysis.

### 2.4.2 Keyword and semantic analysis

165 Multiple complementary keyword-based metrics were employed to identify salient lexical features and their temporal dynamics. Term frequency and TF-IDF were used to quantify both overall prominence and category-specific importance of terms across the corpus and within subsets defined by theme, tone, and period. For each subset, mean TF-IDF scores were computed across documents to extract representative keywords (Qaiser et al., 2018).

Co-occurrence patterns were constructed based on within-document term co-presence to capture relational structure beyond  
170 individual term importance (Mikolov et al., 2013). In addition, semantic shifts were quantified by comparing word representations across time periods. Specifically, cosine similarity was computed between vector representations of the same term in different time periods to measure changes in usage context (Manning et al., 2008).

Neighbor-based Jaccard similarity was computed by comparing the sets of top 10 most associated terms to further assess local contextual changes. (i.e., nearest neighbors based on co-occurrence or vector similarity) for each keyword across time  
175 periods (Hamilton et al., 2016; Kulkarni et al., 2015). This measure captures the extent to which the semantic neighborhood of a term is preserved or reconfigured over time.



### 2.4.3 Topic modeling

To characterize higher-order thematic structures and their evolution, topic modeling was applied to a subset of articles with clearly defined evaluative tone labels. Focusing on value-laden discourse, Latent Dirichlet Allocation (LDA) was used to identify dominant thematic configurations within the corpus (Jelodar et al., 2019). Rather than modeling each tone category independently, a topic model ( $K = 3$ ) was estimated to capture the primary thematic structure underlying tone-classified articles. The number of topics was selected to balance interpretability and thematic resolution. For each topic, representative keywords were identified based on their posterior probabilities, providing interpretable lexical signatures of the underlying themes. To examine temporal dynamics, topic prevalence was tracked across time periods, enabling analysis of how the relative importance of each theme evolves. In addition, the distribution of evaluative tone within each topic was analyzed to assess how thematic structures are associated with shifting value orientations.

### 2.5 Determining the pattern of societal values with transition theory

While annual snapshots of value prevalence offer descriptive insights, they do not capture the systemic trajectory through which values emerge, intensify, and potentially decline. To examine these long term dynamics, Transition Theory was adopted as a conceptual reference (Rotmans, 2005; Rotmans et al., 2001). The theory describes transformation processes across four phases: pre-development, take-off, acceleration, and stabilization. Each phase captures distinct system dynamics, from subtle shifts to rapid changes and eventual equilibrium. Additionally, transitions may also fail, resulting in system breakdown, public backlash, or lock-in effects. Rather than providing detailed theoretical interpretation, Transition Theory was used here as a structural guide for identifying shifts in value trajectories. To identify transition phases empirically, nonlinear regression was applied to manually coded data. The first order derivative was utilized to quantify the direction and magnitude of the change trend, while the second order derivative was utilized to describe the dynamics of the change rate (i.e., acceleration or deceleration) by capturing curvature shifts in the fitted trajectory. This procedure offered a consistent basis for distinguishing patterns of value emergence and stabilization without assigning causal explanations to the observed transitions.

## 200 3 Results

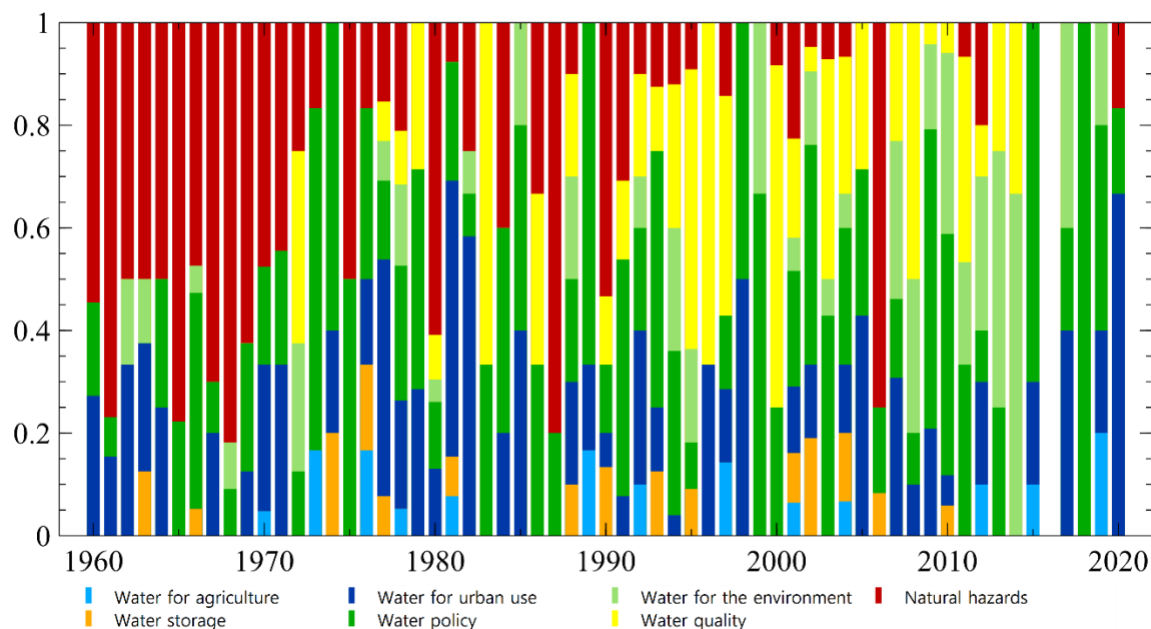
### 3.1 Manual coding

#### 3.1.1 Theme analysis

A total of 644 newspaper articles published between 1960 and 2020 were classified into seven thematic categories using the manual coding framework (Fig. 2). Of the full sample, natural hazards accounted for 180 articles (28%), water policy for 166 (26%), urban water use for 111 (17%), water quality for 96 (15%), environmental themes for 54 (8%), water storage for 20



(3%), and agricultural water use for 17 (3%). These proportions provide a baseline distribution of media attention across major water-related issues in South Korea.



210 **Figure 2. Temporal distribution of water-related themes in Korean newspaper articles (1960–2020). Articles from 2016 were not retrieved during the random sampling process, resulting in a gap in the distribution.**

215 Across the six decades of coverage, the distribution of themes shows clear temporal variation. From the 1960s through the early 1980s, articles related to natural hazards accounted for the largest share, reflecting sustained media attention to floods, droughts, and related emergency responses during this period. Beginning in the mid-1980s, the prominence of water policy and urban water use increased, coinciding with broader national shifts toward urban expansion and institutional development.

220 Attention to environmental themes and water quality rose markedly from the late 1990s onward, and these categories continued to grow through the 2000s and 2010s. Meanwhile, the proportion of articles addressing agricultural water use declined steadily after the 1980s. In more recent decades, coverage became more evenly distributed across several themes, indicating a broadening of media focus on diverse aspects of water management. These temporal shifts suggest that the dominant drivers of public attention moved from hazard response toward governance and water-quality concerns over time, motivating further interpretation in the Discussion.

### 3.1.2 Tone analysis

Manual coding classified the 644 articles into three tone groups: 151 articles (23%) associated with economic development, 123 articles (19%) with environmental sustainability, and 370 articles (58%) with a neutral tone. The temporal distribution of these categories shows a clear shift over the study period. From the 1960s through the mid-1980s, articles emphasizing



225 economic development dominated the coverage. Beginning in the 1990s, the proportion of sustainability-oriented articles gradually increased, and by the mid-2000s, this tone became more prominent within the discourse.

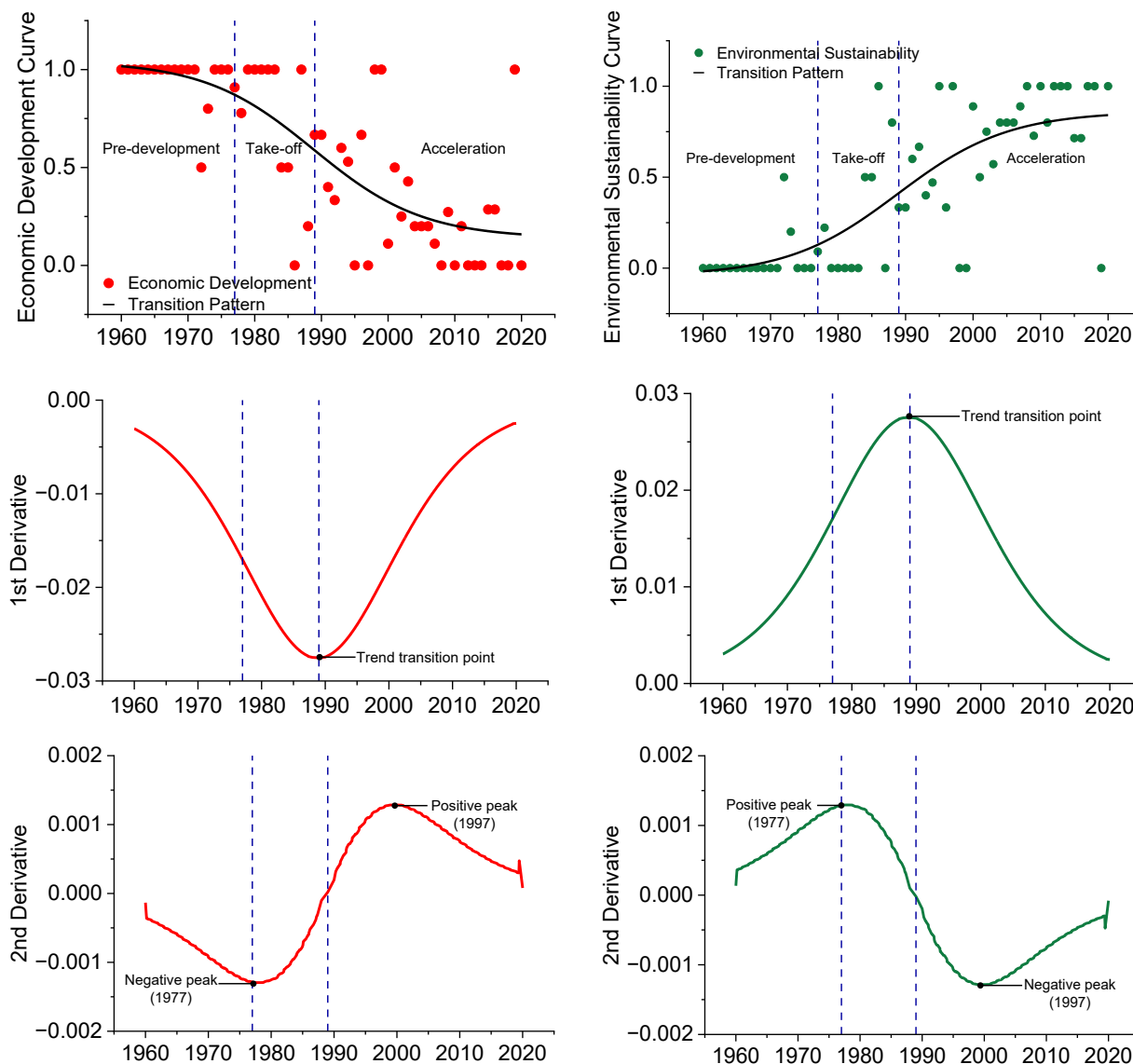
The thematic shifts closely align with changes in tone across different periods. Earlier discourse was dominated by articles framed around economic development, whereas more recent coverage shows an increasing presence of a sustainability-oriented perspective. A notable shift occurred in the 1990s, during which several widely reported water quality-related incidents, particularly those affecting drinking water sources and major rivers, coincided with heightened public concern (Kim et al., 2021; Moon, 2008; National Archives of Korea, 1991). From the mid-2000s onward, additional changes in the policy landscape surrounding the water environment and aquatic ecosystems were accompanied by a further rise in sustainability-framed articles. In particular, the enactment of the Act on the Protection of Water Quality and Aquatic Ecosystems in 2007 and subsequent institutional reforms (Kim, 2020; Kim et al., 2021) corresponded with an increase in coverage emphasizing environmental concerns, contributing to the growing proportion of sustainability-oriented content over time.

### 3.1.3 Transition pattern of the social value of water environment

The analysis indicates a long-term transition in water-related discourse framing from an economic-development orientation toward environmental sustainability over the study period (Fig. 3). Regression analysis combined with transition theory was used to examine how social values concerning the water environment in South Korea have changed from 1960 to 2020. The nonlinear regression (Eq. (1) and (2)) was statistically significant (F-test,  $p$ -value  $< 0.05$ ), indicating that the relationship is unlikely to be explained by chance and that the model meaningfully captures long-term variability. The fitted curves for the economic-development and environmental-sustainability tones, together with their first- and second-derivative profiles, constitute the basis for stage identification in transition theory (Fig. 3). These trajectories suggest a broad societal reorientation in which water issues are increasingly framed around environmental protection and long-term sustainability rather than primarily around growth and development. The derivative-based turning points further indicate that this shift is not gradual and linear, but occurs in identifiable phases, implying periods when changes in public discourse and priorities are likely intensified.

$$E_d = 0.1385 + (1.0426 - 0.1385) / \left( 1 + \left( \frac{t}{1989} \right)^{242.5339} \right) \quad (1)$$

$$E_s = 0.8615 + (-0.0426 - 0.8615) / \left( 1 + \left( \frac{t}{1989} \right)^{242.5339} \right) \quad (2)$$



**Figure 3. Regression trajectories and derivative profiles of economic development and environmental sustainability values. The plots reveal key turning points and transition phases in the evolution of water-related social priorities.**

Based on the derivative-based classification, transition theory identifies four stages in the evolution of social values. The pre-development stage (up to 1977) is characterized by minimal change in either value orientation. The take-off stage (1978-1989) marks the initial rise of sustainability framing. The acceleration stage (1990-1997) exhibits the most rapid shift, with environmental-sustainability discourse increasing and economic-development discourse declining. Notably, as of 2020, the stabilization stage has not yet been reached, indicating that the transition remains ongoing. Although the take-off (11 years) and acceleration (7 years) periods were relatively short, the onset of transition occurred later in South Korea than in China and Australia (Wei et al., 2017; Xiong et al., 2016). This continued non-stabilized pattern suggests that public framing of



260 water issues is still actively evolving, rather than having settled into a new steady state. Practically, it implies that water  
policy debates and media narratives may remain sensitive to emerging environmental events, regulatory changes, and  
societal expectations, creating both uncertainty and opportunity for adaptive, sustainability-oriented management.

## 3.2 Text mining

### 3.2.1 Keyword analysis

265 Systematic keyword tracking was used to quantify long-term shifts in societal and environmental discourse within the  
newspaper corpus. Based TF-IDF scores and term frequency metrics, the most salient keywords across all articles were  
“damage,” “construction,” “water quality,” “project,” and “flooding,” reflecting persistent attention to disaster response,  
infrastructure development, and water-quality concerns. Tone-specific keyword analysis further clarifies these distinctions.  
In articles emphasizing environmental sustainability, prominent terms such as “sewage,” “water quality,” “treatment plant,”  
270 “pollution,” and “treatment” highlight growing concern over contamination control and wastewater management. In contrast,  
articles aligned with the economic-development tone are dominated by terms such as “project,” “construction,” “sewerage,”  
“facilities,” and “development,” indicating a continued emphasis on infrastructural expansion and growth-oriented policy  
frames. These keyword profiles suggest that the two tones are not merely different sentiments toward the same issues but  
represent distinct issue framing, with one emphasizing environmental protection and public health and the other emphasizing  
275 economic growth, infrastructure development, and service expansion. This separation provides an empirical basis for  
interpreting tone as a meaningful indicator of underlying societal priorities in water-related discourse.

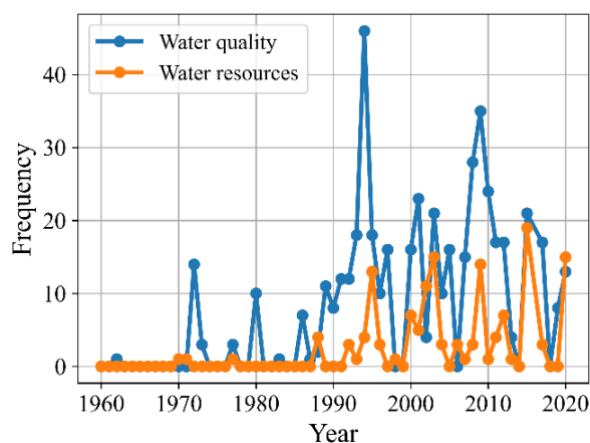
To isolate meaningful semantic change, consistently appearing background terms (e.g., facility, project, plan) were excluded.  
The remaining keywords reveal two distinct patterns. First, structurally dominant terms (e.g., water quality, development,  
sewerage) maintain both high frequency and importance over time, reflecting stable core themes. Second, emerging terms  
280 (e.g., ecology, leachate, water resources) show high TF-IDF despite lower frequency, indicating periods of rapid contextual  
shift. Among these, “water quality” and “water resources” represent key transition signals, with increasing prominence over  
time (Fig. 4a). This trend reflects a shift from infrastructure-oriented discourse toward resource management and  
environmental sustainability.

A complementary perspective is provided by co-occurrence analysis of the term “damage”, which remains consistently  
285 frequent throughout the dataset. Prior to the 2000s, “damage” is primarily associated with natural hazards such as “drought”  
and “flooding”. In contrast, after the 2000s, it increasingly co-occurs with terms such as “pollution” and “accident”. This  
shift indicates a transition from perceiving environmental damage as a natural phenomenon to framing it because of human  
activities, accompanied by a corresponding change from response-oriented to prevention- and regulation-oriented narratives  
(Fig. 4c).

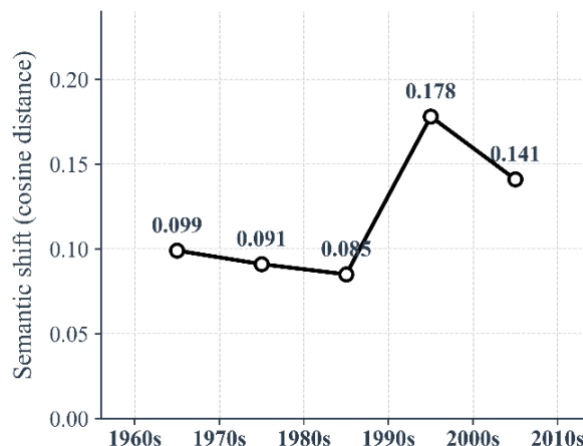
290 Semantic shift analysis based on cosine similarity further supports this transition. Among all keywords, “investigation”  
exhibits the largest cumulative semantic change (0.8056). Its meaning evolves from development-oriented surveying in the



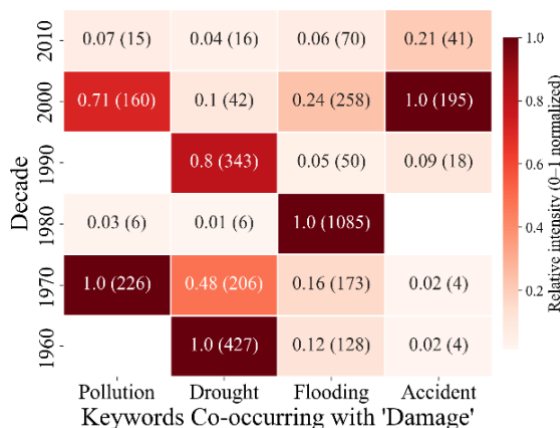
1960s, to industrial inspection in the 1970s, pollution monitoring and environmental assessment in the 1980s–1990s, contamination detection in the 2000s, and finally public health inspection in the 2010s (Figs. 4b and 4d). This progression, confirmed by changes in neighboring terms, suggests a broader transition in discourse emphasis from development and industrialization toward environmental regulation and human-centered management.



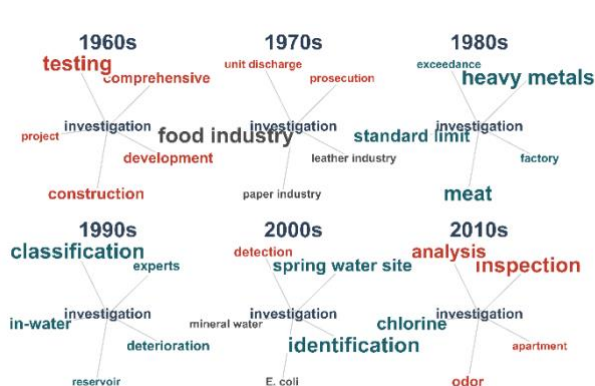
(a) Temporal frequency trend of keywords



(b) Semantic shift of 'Investigation'



(c) Co-occurrence patterns of 'Damage'



(d) Neighbor Structure of 'Investigation'

**Figure 4. Keyword dynamics and semantic evolution in water-related discourse. (a) Temporal trends in the frequency of key terms (water resources and water quality). (b) Semantic shift of investigation quantified using cosine distance across decades. (c) Decadal co-occurrence patterns between damage and selected keywords (pollution, drought, flooding, and accident), shown as a heatmap. (d) Decade-specific semantic neighborhoods of investigation, visualized as ego-networks.**

### 300 3.2.2 Temporal evolution of keywords and discourse tone

The temporal evolution of Korean environmental discourse was examined through TF-IDF analysis across three dimensions including temporal dynamics, tonal orientation, and thematic focus. The domains of natural hazards, water quality, and water policy, identified through manual coding as central narrative domains, served as interpretive anchors for examining how



305 these dimensions unfold across the corpus. This multidimensional structure enabled systematic tracking of long-term discourse transition from development-centered priorities to sustainability-oriented concerns, with each dimension providing quantifiable indicators of shifting public attention (Table 2).

**Table 2. Chronological evolution of water environment discourse: thematic analysis by decade**

Period	Theme	Major Keywords	Interpretation	Tone*
1960s– 1970s	Natural hazard	Typhoon, Flood Damage, Drought, Sewer, Water Supply, Farmers, Pumps	Response centered on minimizing economic losses and post-disaster recovery via infrastructure build-out	Mostly ED
Early 1960s– 1970s	Water policy	Release, Hydraulic Works/Irrigation Projects, Railroad, Plan, Countermeasures, Han River, Flood Control, Levee, Facilities	Emphasis on flood prevention and foundational water infrastructure expansion	Mostly ED
1970s	Water quality	Pollution, Wastewater, Steel, Regulation, Discharge, Standard Limit	Industrial wastewater problems during rapid industrialization; initial water quality standards established	Mixed ED + ES
Mid- 1980s– 2000s	Water policy	Pollution, Sewer, Water Quality, Water Supply, Environmental Agency	Shift of policy momentum toward water-quality management and pollution regulation	Turning to ES
1980s	Water quality	Industrial Complex, Damage, River, Pollution, Environment	Damages near industrial zones; emergence of compensation issues	Mainly ES
1988 onward	Natural hazard	Inundation, Loss, Fatalities, Ministry of Environment, Pollution, Recovery	Rising attention to environmental impacts (pollution, ecology, damage), beyond mere recovery	Turning to ES
1990s	Water quality	Nakdong River, Heavy Metals, Phenol, Tap Water, Standards	Tap-water safety controversies (e.g., Nakdong River phenol spill) and detection of heavy metals	ES-centric, some ED
Early 2000s	Water quality	Treatment Plant, Sewage, Water- Quality Alarm, Groundwater	Expansion of treatment infrastructure and buildout of monitoring/early- warning systems	Mostly ES
2000s onward	Water policy	Rainwater, Groundwater, Wetlands, Urban, Restoration	Expansion of sustainable urban water management and eco-centric policy orientation	ES-centric
2010s	Water	Nakdong River, Leachate, Foot-	Secondary water pollution from	ES-centric



Period	Theme	Major Keywords	Interpretation	Tone*
	quality	And-Mouth Disease, Waste, Ecology	livestock/waste; issues linked to large river projects	
Late 2010s	Water quality	Submerged Litter, Han River, Cleanup Operation	Heightened attention to riverbed contamination and underwater litter	ES-centric

\*ED: Economic development, ES: Environmental sustainability

310 For natural hazards, discourse in the 1960s and 1970s was predominantly framed within an economic-development paradigm, emphasizing disaster recovery and infrastructure reconstruction following typhoons, floods, and droughts. After 1988, however, sustainability-related terms such as “pollution” and “ecological damage” became more prominent, indicating growing recognition of environmental impacts beyond immediate disaster response. By the 2000s, keywords such as “reservoir levels,” “compensation,” and “ecosystem” reflected heightened public attention in media discourse of climate-  
315 related risks and a gradual move from centralized recovery efforts to more localized, environmentally oriented disaster management.

Water quality discourse underwent a similarly structured transformation. During the 1970s and 1980s, attention focused on industrial wastewater and pollution associated with rapid industrialization. The 1990s marked a turning point, exemplified by the Nakdong River phenol contamination, which intensified concerns over drinking water safety and strengthened  
320 monitoring and regulatory scrutiny (National Archives of Korea, 1991). In the 2000s, the expansion of wastewater treatment infrastructure signaled increasing institutional capacity, while the 2010s brought focus on more complex ecological issues, including leachate management, livestock waste treatment, and underwater debris removal. Overall, the trajectory reflects a shift from industrial regulation to ecosystem-based protection.

Water policy discourse followed a comparable progression. From the 1960s through the 1980s, national policy emphasized  
325 infrastructure development, including flood prevention, embankment construction, and irrigation projects. By the mid-1990s, keywords such as “sewage,” “water quality,” and “Environmental Agency” signaled increasing attention to pollution control and the institutionalization of environmental regulation. In the 2000s, terms like “rainwater,” “river restoration,” and “Cheonggyecheon stream” emerged, highlighting a decisive shift toward sustainable urban water management and ecological restoration.

330 The results reveal a transition from an economic development–focused narrative of the 1960s to 1980s, centered on infrastructure expansion and disaster recovery, to a sustainability-oriented discourse that began to emerge in the late 1980s. This shift is evident in the increasing prominence of keywords associated with pollution control, ecosystem restoration, and localized governance, particularly after the 2000s. Across all three analytical domains (natural hazards, water quality, and water policy), these patterns collectively demonstrate a broader transition in societal discourse in Korea toward sustainable  
335 environmental management, marked by rising public awareness and evolving policy priorities.



### 3.2.3 Topic modeling

Topic modeling was applied to the tone-classified corpus to identify dominant thematic structures and their evolution over time. Three primary topics were identified based on representative keywords. Topic 1, Hydro-infrastructure and hazard response, is characterized by terms related to infrastructure and disaster mitigation including “facility”, “drought”, “construction”, “damage”, “sewerage”, and “flooding”. Topic 2, Water development and environmental planning, captures mixed themes of development and environmental management including “project”, “water quality”, “development”, “ecology”, and “treatment plant”. Topic 3, Water quality governance and pollution control, reflects regulatory and pollution-related discourse including “pollution”, “wastewater”, “discharge”, “management”, and “standards”.

The temporal composition of topics reveals a clear structural transition (Fig. 5a). Topic 1 dominates the earlier decades (1960s–1980s), reflecting a strong emphasis on infrastructure expansion and hazard response during rapid urbanization. Its prevalence declines after the 1990s. In contrast, Topic 3 indicates a steady increase over time, becoming dominant in recent decades, indicating a growing focus on pollution control and environmental governance. Topic 2 also exhibits a gradual increase, suggesting the rising importance of integrated development and environmental planning.

Further insight is provided by topic–value coupling analysis, which links thematic content with article tone (Fig. 5b). Topic 3 consistently exhibits a strong environmental sustainability tone across all decades, highlighting its stable association with environmental concerns. Topic 1, while initially dominated by economic development-oriented narratives, shows a noticeable increase in environmental sustainability framing in the 2010s, suggesting a reinterpretation of infrastructure in the context of sustainability. Topic 2 demonstrates the most pronounced shift and is initially associated with economic development in earlier decades but transitions toward environmental sustainability after the 1980s.

To better understand this transformation, Topic 2 was examined in detail through temporal word composition analysis (Fig. 5c). In the 1970s, the topic is dominated by terms related to infrastructure expansion and resource utilization, reflecting a development-oriented perspective. By the 2000s, however, its keyword composition shifts toward water quality management and ecological planning, indicating a reorientation toward sustainability and environmental considerations. This transition highlights the evolving role of development discourse, from resource exploitation toward integrated environmental management.

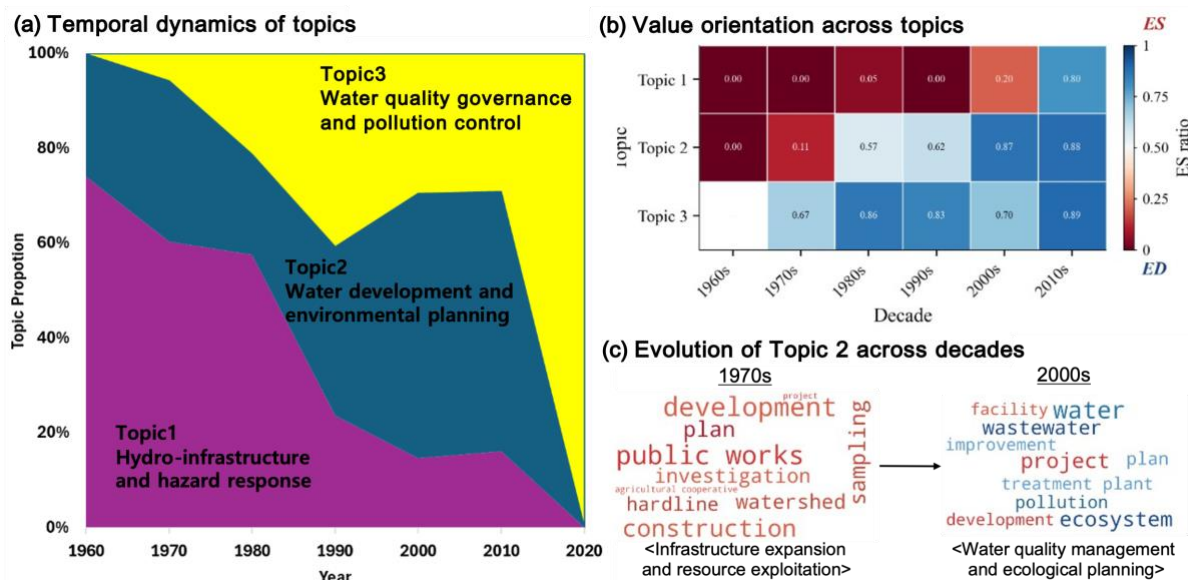


Figure 5. Topic dynamics and value shifts in water discourse. (a) Topic prevalence is shown as a stacked area plot, illustrating temporal changes in thematic composition. (b) Topic–value associations are presented using a decade-resolved heatmap of environmental sustainability (ES) versus economic development (ED) dominance. (c) Semantic evolution of Topic 2, highlighting changes in keyword composition from development-oriented to environment-oriented framing.

### 3.3 Multidimensional evolution of water-related discourse under HSS

The HSS analysis indicates that water-related media discourse in South Korea experienced a gradual structural transition between 1960 and 2020. Rather than relying on a single indicator, this transition was supported by converging evidence from manual coding, keyword trends, co-occurrence structures, semantic associations, and topic prevalence. This multidimensional evolution reflects concurrent transformations across three complementary dimensions of public discourse: (1) thematic composition, (2) the relational structure among key concepts, and (3) the broader contextual framing through which water issues were interpreted. Early discourse was strongly centered on hazard response, hydraulic infrastructure, and economic development, whereas environmental management, pollution control, and sustainability-related themes became increasingly prominent after the late 1980s (Fig 3). However, this transition did not occur through the simple replacement of earlier development-oriented narratives. Instead, multiple water-related priorities coexisted and progressively diversified over time. This pattern was consistently observed across manual coding and computational text-mining analyses. Keyword prominence, co-occurrence structures, semantic associations, and topic prevalence all indicate that water-related discourse became less concentrated on construction and disaster recovery and more connected to environmental governance, ecological management, and regulatory concerns (Figs. 4 and 5). Collectively, these results suggest that the evolution of water-related media discourse was characterized not only by thematic change, but also by broader reorganization of semantic relationships and contextual framing across decades. These empirical patterns provide the basis for the discussion that follows, which interprets the observed discourse transition in relation to social response, methodological integration, and potential use of



385 HSS indicators in sociohydrological modeling. Understanding long-term water discourse through multiple interconnected dimensions provides insight into how societal perceptions, governance priorities, and environmental concerns co-evolve over time.

## 4 Discussion

### 4.1 Long-term discourse transitions revealed by HSS

390 The theme, tone, keyword, and co-occurrence results point to a coherent long-term transition in South Korea's water discourse from hazard- and construction-centered narratives toward water-quality protection and environmental sustainability. This transition should be interpreted as a change in publicly visible discourse rather than a direct measurement of public perception. Nevertheless, because media discourse reflects how water-related issues are framed, prioritized, and circulated in society, it provides a useful proxy for examining perception-related dynamics over long periods. The high share of natural-hazard coverage from the 1960s to early 1980s is consistent with a period when floods and droughts were experienced as immediate, high-salience threats and when public discourse emphasized response and recovery (Fig. 2). Over  
395 time, however, sustained investment in water infrastructure and hazard-mitigation systems likely altered not only physical exposure to hydrologic risk, but also public discourse surrounding water-related problems. As dams, reservoirs, stormwater drainage, and flood-control measures reduced the frequency or severity of disruptive impacts in daily life, public attention appears to have gradually shifted away from acute disaster response toward broader issues of water governance and management.

400 This transition became particularly pronounced from the mid-1980s onward, as water-related discourse increasingly shifted from hydraulic development and resource allocation toward water quality, environmental management, and urban sustainability. These changes coincided with rapid urbanization and institutional expansion, which broadened public expectations surrounding water beyond protection from hydrologic extremes to include ecological integrity, public health, and quality of life. The tone trajectories further support this interpretation by showing a sustained decline in economic-  
405 development framing alongside a rise in environmental-sustainability framing. Importantly, the derivative-based turning points indicate that this transition occurred through distinct phases rather than through a gradual linear shift. The major inflection identified in the late 1980s to early 1990s coincides with periods of heightened public attention to major water-quality incidents (Kim et al., 2021; Hwang et al., 2017), implying that environmental concerns may have become increasingly prominent in social discourse before being fully institutionalized within formal policy frameworks.

410 The keyword co-occurrence results further support this interpretation. In particular, the association of "damage" shifted from drought and flooding toward pollution and accidents, indicating a broader transformation in how water-related risks were socially attributed. Whereas earlier discourse framed damage mainly as the result of natural hazards, later discourse increasingly linked water problems to anthropogenic causes and governance failures. Overall, these converging signals



support the use of HSS indicators as empirically grounded proxies for evolving social priorities that can be linked to  
415 sociohydrological feedback and decision-making contexts.

#### 4.2 Methodological implications of HSS integration

From a methodological perspective, HSS through manual coding and text mining is an effective tool for understanding changes in social systems and human responses to the water environment. The main methodological contribution is the integration of contextual interpretation and scalable pattern detection within a single workflow. This integration responds to  
420 the three methodological challenges raised in the Introduction. Text mining provides the scalability needed to trace long-term patterns across a large corpus that manual reading alone cannot cover. Manual coding preserves the contextual interpretation and evaluative meaning that fully automated analysis tends to miss. Tracking semantic shifts against a manually coded benchmark also helps separate substantive changes in water-related framing from changes that mainly reflect evolving vocabulary or journalistic conventions. By combining the deep contextual insights of manual coding with  
425 the scalability of text mining, this framework addresses the scale-validity tradeoff by overcoming both the limitations of small-scale qualitative analysis and the contextual blind spots of fully automated analysis. Rather than functioning as independent analytical procedures, manual coding and text mining operate as complementary approaches that capture different dimensions of discourse dynamics. The combined application of LDA and tone-coupling analysis reveals a gradual transition from development- and hazard-oriented discourse toward environmental governance and sustainability-oriented  
430 water management over time (Fig. 5). In addition, extracted keywords can serve as functional elements for constructing higher-level social response variables (e.g., social awareness and community sensitivity). Transforming unstructured discourse into temporally resolved indicators enables more systematic analyses of social dynamics. Furthermore, the fusion of extensive human-generated data and deep learning algorithms enables near real-time detection and predictive analysis of social phenomena, opening new avenues for HSS.

435 Our analysis demonstrates a broader methodological role for manual coding in environmental discourse analysis. Manually coded data such as thematic categories, affective labels, and interpretive criteria can serve as an interpretive reference for assessing computationally derived discourse patterns (Figs. 2 and 3, and Table 2). This is particularly important because algorithmically derived outputs may not fully correspond to how humans interpret and perceive environmental issues (Kim et al., 2026). Accordingly, manual coding within the proposed framework is not limited to an initial classification but also  
440 contributes to the interpretive evaluation and contextual grounding of computational outputs.

This complementary structure enhances both the interpretive credibility and analytical robustness of the HSS (Song et al., 2020). To improve reliability and reproducibility, future applications should incorporate multiple coders and formal intercoder agreement assessment (e.g., Krippendorff's  $\alpha$ ), and independent verification procedures (Hayes and Krippendorff, 2007; Nili et al., 2020; O'Connor and Joffe, 2020). Moreover, manual coding remains adaptable throughout the analytical  
445 process and can be re-applied to refine emerging themes, assess intermediate findings, or recalibrate computational models



as new data emerge (Compton et al., 2012; Krippendorff, 2013). This iterative interaction between human interpretation and computational analysis represents one of the key methodological strengths of the HSS framework.

### 4.3 Advancing sociohydrological modeling through HSS

The applicability of HSS in sociohydrological modeling is clearly identified. HSS bridges the gap between context-specific observation derived from data and scalable inputs required for dynamic modeling. Linking HSS data to sociohydrological models involves three primary methodological approaches, each with distinct assumptions and complexity, as illustrated in Fig. 6. These approaches collectively form a flexible toolkit that can be applied depending on the problem context and data availability. For example, the annual proportion of sustainability-framed articles could serve as a proxy for environmental awareness, while the decay of hazard-related keywords after major events could inform social memory parameters. The first approach is social parameter calibration which estimates key social parameters from data. The second approach uses proxy variables, which substitute observable indicators for abstract concepts. The third is the lumped approach which represents society as a dynamic sub-model with bi-directional feedback to physical systems. This method, while more complex, provides a powerful framework for integrated modeling. Collectively, HSS not only supports the validation and refinement of individual model components but also offers the development of simplified, yet effective, representations of complex social-environmental interactions (Mijic et al., 2024)

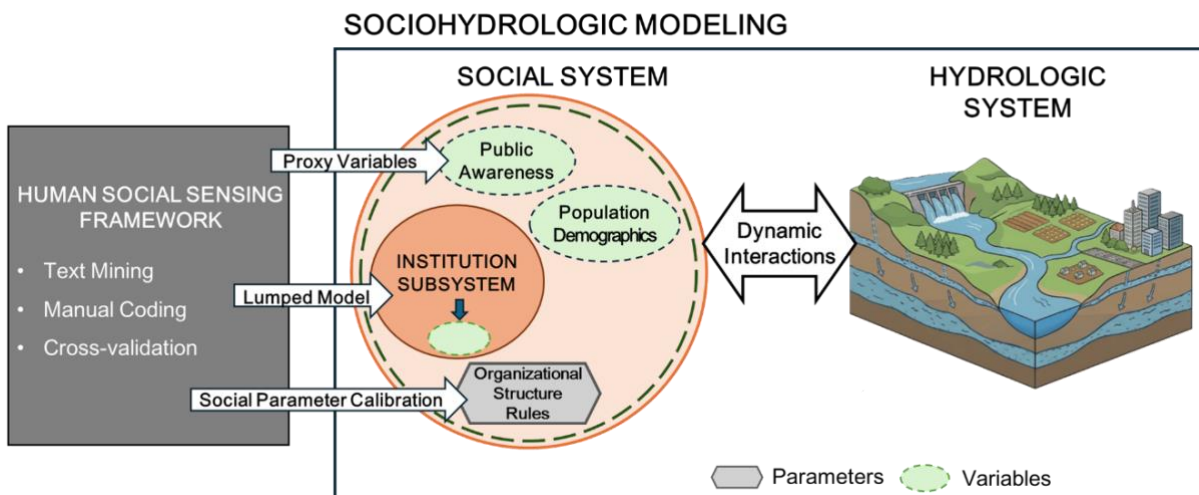


Figure 6. Methodological approaches for integrating human social sensing into sociohydrological model: social parameter calibration, proxy variables, and lumped model approach.

#### 4.3.1 Social parameter calibration

Sociohydrological models often contain parameters that describe unobserved social processes, such as the fading of disaster memory, the spread of information, and community sensitivity to flood risk (Coletta et al., 2024; Rahnema et al., 2024). These parameters are critical to model behavior but are hard to measure directly and have traditionally been set based on



subjective assumptions (Shanono and Ndiritu, 2023; S. Song et al., 2021). Social parameter calibration addresses this by using inverse estimation on social data to quantify such parameters, thereby grounding the model in empirical observations (Genova and Wei, 2023). The process improves the model's validity by keeping its theoretical structure while using real-world social response data to estimate the best parameters (Sawada and Hanazaki, 2020; S. Song et al., 2021) For example, social memory decay after disasters can be measured by analyzing disaster-related mentions in news and social media, using an exponential decay model to estimate the fading rate (Candia et al., 2019). Song et al. (2021) demonstrate that rural communities forget faster due to reliance on oral memory, while urban areas retain memories longer through media, underscoring the need for tailored policies.

#### 4.3.2 Use of proxy variables

An effective way to integrate social factors into sociohydrological models is to replace abstract, hard-to-measure concepts with observable proxy variables (Kong et al., 2025). These proxies are measurable social indicators, including the proportion of news articles on drought, the tone of media coverage, and the frequency of co-occurring disaster-related terms, which approximate underlying social perceptions such as risk awareness (Ahmad and Kam, 2024). By continuously updating these text-mined metrics from media sources, models can capture time-varying social dynamics and remain aligned with real-world trends (Veigel et al., 2025). This proxy-based approach simplifies complex social processes and enables rapid incorporation of social data into hydrological modeling (Kong et al., 2025; Liu et al., 2023). For instance, Sawada and Hanazaki (2020) used sequential data assimilation with historical population and levee data to model human–flood interactions. Adding social data, such as news and social media mentions helps the model better capture and predict social perceptions in near real-time (Bozas et al., 2024).

#### 4.3.3 Lumped model approach

The lumped model approach treats society as a single entity using aggregated social variables (like community awareness) that interact dynamically with hydrological factors (Gu et al., 2021; Vanelli et al., 2022). Within this approach, social and hydrological variables interact through feedback loops such as how drought conditions raise awareness, trigger water-saving actions, and then affect water resources (Gil-García et al., 2024). Although this approach is useful for understanding broad system behavior and guiding policy, it may overlook individual or group differences (Shrestha et al., 2022; Yu et al., 2022). Nevertheless, advances in machine learning and data driven modeling increasingly enable dynamic estimation of social response variables from historical discourse and hydrological observations, providing new opportunities for adaptive feedback representation in sociohydrological systems (Bao et al., 2025; Xu et al., 2024).

#### 4.4 Limitations and future research directions

While promising, the current HSS framework has several limitations that should be addressed in future research. These limitations do not undermine the usefulness of HSS, but they clarify the conditions under which discourse-based indicators



should be interpreted cautiously. First, newspaper data represents a mediated form of public discourse, as they reflect  
500 editorial biases, selective coverage, and urban-centric perspectives (Soroka et al., 2019). Thus, the HSS-derived indicators  
may be skewed toward issues deemed newsworthy rather than truly representative of the public. Incorporating multi-source  
data fusion, e.g., social media, public forums, or government communications, could help balance this bias and improve  
representativeness (Fan et al., 2024). Second, although topic modeling and sentiment analysis offer reproducible outputs,  
thematic labeling and interpretation still require subjective judgment (Ortu et al., 2024). Further methodological refinement,  
505 such as using supervised learning based on validated themes or applying cross-validation with expert panels, could increase  
robustness (Barberá et al., 2021). Third, while we used discourse trends as proxies for perception, the gap between expressed  
discourse and actual behavior remains a critical challenge (Borges-Tiago et al., 2024). Future studies should seek to validate  
HSS outputs through triangulation with behavioral data such as water use patterns, protest activity, and citizen science  
participation (Cesare et al., 2018). Lastly, future work could explore how perception trends can be directly coupled with  
510 sociohydrological model parameters, such as feedback strength, adaptive response rates, or institutional inertia coefficients.  
This would mark a key step toward operationalizing social memory, value shifts, and collective adaptation within physically  
based models, thus making sociohydrology not only a theoretical framework but also a practical tool for water governance.

## 5 Conclusion

This study investigated long-term changes in water-related societal discourse in South Korea from 1960 to 2020 using HSS,  
515 which integrates structural manual coding with computational text mining. Our results suggest that HSS can provide useful  
discourse-level indicators associated with long-term changes in societal framing of water-related environmental issues, as it  
characterizes long-term transitions in water-related discourse through topic trends, keyword frequencies, and their temporal  
evolution. Manual coding revealed a broad transition from development- and disaster-oriented discourse in the 1960s–1980s  
to growing attention to environmental management from the 1990s onward. Within this trajectory, text mining identified  
520 specific structural shifts, including the rising prominence of water quality, water resources, pollution control, and ecosystem  
restoration, alongside a gradual decline in hazard-centered narratives. The complementary use of both methods enhanced the  
robustness and interpretability of unstructured data analysis. Beyond complementarity, manual coding served as an  
interpretive benchmark for evaluating text mining outputs, addressing a common limitation associated with purely data-  
driven modeling approaches. This study positions HSS as a concrete, empirical framework for advancing sociohydrological  
525 modeling by enabling more realistic representations of human–water interactions. The key contribution is not that newspaper  
discourse directly measures public perception, but that it can be systematically transformed into interpretable, time-resolved  
proxy indicators of evolving societal framing. Practically, the framework can support water managers and policymakers by  
helping track long-term shifts in public priorities, thereby informing communication, planning, and policy design that are  
better aligned with societal concerns. Hydrologically, incorporating empirically derived perception dynamics can improve



530 the representation of human feedback in water systems, leading to a more realistic analysis of how social change influences water demand, risk response, and environmental management over time.

### **Code and data availability**

535 The news articles analyzed in this study were obtained from third-party news databases and are subject to copyright restrictions; therefore, the original article texts cannot be redistributed. To ensure reproducibility, the processed datasets used in this study, including the manual coding results, and data required to reproduce all figures and analyses, together with the Python code used for data processing, will be made publicly available through a Zenodo repository(10.5281/zenodo.20610482) upon publication. Information on the retrieval and selection of the original news articles is provided in the Methods section.

### 540 **Author contributions**

JL conducted the formal analysis, interpreted the results, prepared the visualizations, and wrote the original draft of the manuscript. WJ contributed to the manual coding process, data analysis, and visualization. JK contributed to validation and manuscript review. YH, DY, and RB contributed to reviewing and editing the manuscript. HJ conceived and supervised the study, contributed to the conceptualization and methodology, and assisted with writing, review, and editing. All authors  
545 discussed the results and approved the final manuscript.

### **Competing interests**

The authors declare that they have no competing interests

### **Acknowledgements**

550 Some of the manual coding results used in this study were adapted from the master's thesis of Wuseong Jeong at Seoul National University of Science and Technology. The authors used AI tools solely to support language editing and take full responsibility for the content of the manuscript.

### **Financial support**

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) (RS-2025-00557036). This research was also supported by the MSIT (Ministry of Science, ICT), Korea, under the



- 555 Global Research Support Program in the Digital Field (RS-2024-00428227) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

## References

- Ahmad, D. M. and Kam, J.: Disparity between global drought hazard and awareness, *NPJ Clean Water*, 7, <https://doi.org/10.1038/s41545-024-00373-y>, 2024.
- 560 Azadi, Y., Yaghoubi, J., Gholizadeh, H., Gholamrezai, S., and Rahimi-Feyzabad, F.: Social barriers to water scarcity adaptation: A grounded theory exploration in arid and semi-arid regions, *Agric. Water Manag.*, 309, <https://doi.org/10.1016/j.agwat.2025.109338>, 2025.
- Bao, D., George Xue, Z., Hiatt, M., Xu, K., Harris, C. K., and Trepanier, J. C.: A machine learning-based prediction-to-map framework for rapid and accurate spatial flood prediction, *npj Natural Hazards*, 2, [https://doi.org/10.1038/s44304-025-](https://doi.org/10.1038/s44304-025-00122-2)
- 565 00122-2, 2025.
- Barberá, P., Boydston, A. E., Linn, S., McMahon, R., and Nagler, J.: Automated Text Classification of News Articles: A Practical Guide, *Polit. Anal.*, 29, 19-42, <https://doi.org/10.1017/pan.2020.8>, 1 January 2021.
- Bennett, N. J.: Using perceptions as evidence to improve conservation and environmental management, *Conserv. Biol.*, 30, 582-592, <https://doi.org/10.1111/cobi.12681>, 1 June 2016.
- 570 Borges-Tiago, M. T., Almeida, A., Tiago, F. G. B., and Avelar, S. M. M.: Bridging the innovative Attitude–Behavior Gap: A dual-level analysis, *J. Innov. Knowl.*, 9, <https://doi.org/10.1016/j.jik.2024.100561>, 2024.
- Bozas, A., Andreadis, S., Chatzakou, D., Symeonidis, S., Theodosiadou, O., Kyriakidis, P., Kokkalas, A., Stathopoulos, E. A., Diplaris, S., Tsikrika, T., Gialampoukidis, I., Vrochidis, S., and Kompatsiaris, I.: From Research to Applications: What Can We Extract with Social Media Sensing?, *SN Comput. Sci.*, 5, <https://doi.org/10.1007/s42979-024-02712-9>, 2024.
- 575 de Brito, M. M., Sodoge, J., Fekete, A., Hagenlocher, M., Koks, E., Kuhlicke, C., Messori, G., de Ruiter, M., Schweizer, P. J., and Ward, P. J.: Uncovering the Dynamics of Multi-Sector Impacts of Hydrological Extremes: A Methods Overview, *Earth's Futur.*, 12, e2023EF003906, <https://doi.org/10.1029/2023EF003906>, 1 January 2024.
- Candia, C., Jara-Figueroa, C., Rodriguez-Sickert, C., Barabási, A. L., and Hidalgo, C. A.: The universal decay of collective memory and attention, *Nat. Hum. Behav.*, 3, 82–91, <https://doi.org/10.1038/s41562-018-0474-5>, 2019.
- 580 Cesare, N., Lee, H., McCormick, T., Spiro, E., and Zagheni, E.: Promises and Pitfalls of Using Digital Traces for Demographic Research, *Demography*, 55, 1979–1999, <https://doi.org/10.1007/s13524-018-0715-2>, 2018.
- Chen, X., Wang, D., Tian, F., and Sivapalan, M.: From channelization to restoration: Sociohydrologic modeling with changing community preferences in the Kissimmee River Basin, Florida, *Water Resour. Res.*, 52, 1227–1244, <https://doi.org/10.1002/2015WR018194>, 2016.



- 585 Coletta, V. R., Pagano, A., Zimmermann, N., Davies, M., Butler, A., Fratino, U., Giordano, R., and Pluchinotta, I.: Socio-hydrological modelling using participatory System Dynamics modelling for enhancing urban flood resilience through Blue-Green Infrastructure, *J. Hydrol. (Amst)*, 636, <https://doi.org/10.1016/j.jhydrol.2024.131248>, 2024.
- Compton, D., Love, T. P., and Sell, J.: Developing and Assessing Intercoder Reliability in Studies of Group Interaction, *Sociol. Methodol.*, 42, 348–364, <https://doi.org/10.1177/0081175012444860>, 2012.
- 590 Constantino, S. M., Schlüter, M., Weber, E. U., and Wijermans, N.: Cognition and behavior in context: a framework and theories to explain natural resource use decisions in social-ecological systems, *Sustain. Sci.*, 16, 1651–1671, <https://doi.org/10.1007/s11625-021-00989-w>, 2021.
- Fan, J., Peng, L., Chen, T., and Cong, G.: Mining the impact of social media information on public green consumption attitudes: a framework based on ELM and text data mining, *Humanit. Soc. Sci. Commun.*, 11, <https://doi.org/10.1057/s41599-024-02649-7>, 2024.
- 595 Galesic, M., Bruine de Bruin, W., Dalege, J., Feld, S. L., Kreuter, F., Olsson, H., Prelec, D., Stein, D. L., and van der Does, T.: Human social sensing is an untapped resource for computational social science, *Nature*, 595, 214–222, <https://doi.org/10.1038/s41586-021-03649-2>, 8 July 2021.
- Genova, P. and Wei, Y.: A socio-hydrological model for assessing water resource allocation and water environmental regulations in the Maipo River basin, *J. Hydrol. (Amst)*, 617, <https://doi.org/10.1016/j.jhydrol.2023.129159>, 2023.
- 600 Gil-García, L., Montilla-López, N. M., Gutiérrez-Martín, C., Sánchez-Daniel, Á., Saiz-Santiago, P., Polanco-Martínez, J. M., Pindado, J., and Pérez-Blanco, C. D.: Actionable human-water system modelling under uncertainty, *Hydrol. Earth Syst. Sci.*, 28, 4501–4520, <https://doi.org/10.5194/hess-28-4501-2024>, 2024.
- Grubert, J., Langlotz, T., Zollmann, S., and Regenbrecht, H.: Towards pervasive augmented reality: Context-awareness in augmented reality, *IEEE Trans. Vis. Comput. Graph.*, 23, 1706–1724, <https://doi.org/10.1109/TVCG.2016.2543720>, 2017.
- 605 Gu, J., Sun, S., Wang, Y., Li, X., Yin, Y., Sun, J., and Qi, X.: Sociohydrology: An Effective Way to Reveal the Coupled Evolution of Human and Water Systems, *Water Resour. Manage.*, <https://doi.org/10.1007/s11269-021-02984-3>, 1 November 2021.
- Hamilton, W. L., Leskovec, J., and Jurafsky, D.: Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change, in: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 1489–1501, 2016.
- 610 Han, Z., Wei, Y., and Meng, J.: Representing human agency in social-hydrological models, a water (re-) allocation perspective, *J. Hydrol. (Amst)*, 662, <https://doi.org/10.1016/j.jhydrol.2025.133944>, 2025.
- Hayes, A. F. and Krippendorff, K.: Answering the Call for a Standard Reliability Measure for Coding Data, *Commun. Methods Meas.*, 1, 77–89, <https://doi.org/10.1080/19312450709336664>, 2007.
- 615 Herrera, P. A., Marazuela, M. A., and Hofmann, T.: Parameter estimation and uncertainty analysis in hydrological modeling, *Wiley Interdiscip. Rev.: Water*, 9, e1569, <https://doi.org/10.1002/wat2.1569>, 1 January 2022.
- Hovy, D. and Spruit, S. L.: The Social Impact of Natural Language Processing, in: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 591–598, 2016.



- Javanbakht-Sheikhahmad, F., Rostami, F., Azadi, H., Veisi, H., Amiri, F., and Witlox, F.: Agricultural Water Resource Management in the Socio-Hydrology: A Framework for Using System Dynamics Simulation, *Water Resour. Manag.*, 38, 2753–2772, <https://doi.org/10.1007/s11269-024-03786-z>, 2024.
- Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., and Zhao, L.: Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey, *Multimed. Tools Appl.*, 78, 15169–15211, <https://doi.org/10.1007/s11042-018-6894-4>, 2019.
- Jurafsky, D. and Martin, J. H.: *Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*, 3rd ed., 2026.
- Kenter, J. O., Martino, S., Buckton, S. J., Waddock, S., Agarwal, B., Anger-Kraavi, A., Costanza, R., Hejnowicz, A. P., Jones, P., Lafayette, J. O., Kabubo-Mariara, J., Mukherjee, N., Pickett, K. E., Riedy, C., and Waddell, S.: Ten principles for transforming economics in a time of global crises, *Nat. Sustain.*, 8, 837–847, <https://doi.org/10.1038/s41893-025-01562-4>, 2025.
- Kim, D.: Recent trends and issues in the domestic water environment legislation, *Environ. Law Rev.*, 42, 1–26, 2020.
- Kim, H., Foster, E., and Chang, H.: Transition of water quality policies in Oregon, USA and South Korea: a historical socio-hydrological approach, *Hydrol. Sci. J.*, 66, 2117–2131, <https://doi.org/10.1080/02626667.2021.1986628>, 2021.
- Kim, J., Lim, K., Adamowski, J., and Jeong, H.: Unveiling the Performance Gap Between Deep Text Mining and Human Coders in Environmental Information Extraction, *Int. J. Hum.-Comput. Interact.*, <https://doi.org/10.1080/10447318.2026.2669061>, 2026.
- Kong, I., Seibert, J., and Purves, R. S.: Droughts and media: when and how do the newspapers talk about the droughts in England?, *Hydrol. Earth Syst. Sci.*, 29, 3795–3808, <https://doi.org/10.5194/hess-29-3795-2025>, 2025.
- Krippendorff, K.: Commentary: A Dissenting View on So-Called Paradoxes of Reliability Coefficients, *Ann. Int. Commun. Assoc.*, 1, 481–499, <https://doi.org/10.1080/23808985.2013.11679143>, 2013.
- Krippendorff, K.: *Content Analysis: An Introduction to Its Methodology*, Fourth Edition., SAGE Publications, Inc., Thousand Oaks, California, <https://doi.org/10.4135/9781071878781>, 2018.
- Kulkarni, V., Al-Rfou, R., Perozzi, B., and Skiena, S.: Statistically significant detection of linguistic change, in: *WWW 2015 - Proceedings of the 24th International Conference on World Wide Web*, 625–635, <https://doi.org/10.1145/2736277.2741627>, 2015.
- Lee, T. M., Markowitz, E. M., Howe, P. D., Ko, C. Y., and Leiserowitz, A. A.: Predictors of public climate change awareness and risk perception around the world, *Nat. Clim. Chang.*, 5, 1014–1020, <https://doi.org/10.1038/nclimate2728>, 2015.
- Li, B. and Sivapalan, M.: Long-Term Coevolution of an Urban Human-Water System Under Climate Change: Critical Role of Human Adaptive Actions, *Water Resour. Res.*, 56, <https://doi.org/10.1029/2020WR027931>, 2020.



- Liu, L., Xue, J., Mao, D., Chang, J., Wang, S., and Li, X.: An integrative socio-hydrological resilience assessment and management implications for oasis sustainability in arid regions, Northwest China, *J. Hydrol. Reg. Stud.*, 47, <https://doi.org/10.1016/j.ejrh.2023.101389>, 2023.
- 655 Malede, D. A., Elumalai, V., Andualem, T. G., Mekonnen, Y. G., Yibeltal, M., Demeke, G. G., Alamirew, T., Kassaye, S. M., Hagos, Y. G., Kassie, A. E., Kassa, M., Berihun, G. S., and Ray, R. L.: Understanding flood and drought extremes under a changing climate in the Blue Nile Basin: A review, *Environ. Sustain. Indic.*, 26, <https://doi.org/10.1016/j.indic.2025.100638>, 1 June 2025.
- Manning, C. D., Raghavan, P., and Schütze, H.: *Introduction to Information Retrieval*, Cambridge University Press, 2008.
- Miao, C., Hu, J., Moradkhani, H., and Destouni, G.: Hydrological Research Evolution: A Large Language Model-Based  
660 Analysis of 310,000 Studies Published Globally Between 1980 and 2023, *Water Resour. Res.*, 60, e2024WR038077, <https://doi.org/10.1029/2024WR038077>, 1 June 2024.
- Mijic, A., Liu, L., O’Keeffe, J., Dobson, B., and Chun, K. P.: A meta-model of socio-hydrological phenomena for sustainable water management, *Nat. Sustain.*, 7, 7–14, <https://doi.org/10.1038/s41893-023-01240-3>, 2024.
- Mikolov, T., Chen, K., Corrado, G., and Dean, J.: Efficient Estimation of Word Representations in Vector Space,  
665 arXiv:1301.3781, 2013.
- Moon, T. H.: Evolution of Environmental Policy in Korea and It’s Future Task, 1963-2007, *KOREAN POLICY STUDIES REVIEW*, 17, 131–162, 2008.
- Muñoz, R., Vaghefi, S. A., Drenkhan, F., Santos, M. J., Viviroli, D., Muccione, V., and Huggel, C.: Assessing Water Management Strategies in Data-Scarce Mountain Regions under Uncertain Climate and Socio-Economic Changes, *Water Resour. Manag.*, 38, 4083–4100, <https://doi.org/10.1007/s11269-024-03853-5>, 2024.
- National Archives of Korea: Nakdong River phenol contamination incident: Accident details and future countermeasures, 1991.
- Nili, A., Tate, M., Barros, A., and Johnstone, D.: An approach for selecting and using a method of inter-coder reliability in information management research, *Int. J. Inf. Manage.*, 54, <https://doi.org/10.1016/j.ijinfomgt.2020.102154>, 2020.
- 675 O’Connor, C. and Joffe, H.: Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines, *Int. J. Qual. Methods*, 19, <https://doi.org/10.1177/1609406919899220>, 2020.
- Ortu, M., Frigau, L., and Contu, G.: Topic based quality indexes assessment through sentiment, *Comput. Stat.*, 39, 289–311, <https://doi.org/10.1007/s00180-022-01284-7>, 2024.
- Park, E. L. and Cho, S.: KoNLPy: Korean natural language processing in Python, in: *Proceeding of the 26th Annual Conference on Human & Cognitive Language Technology*, 2014.
- 680 Qaiser, S., Utara, U., Sintok, M., Kedah, M., Ramsha, A., and Analytics, T.: Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents Text Mining, *Int. J. Comput. Appl.*, 975–8887 pp., 2018.
- Rahnama, B., Nasiri Saleh, F., Nasseri, M., and Koosheshi, M.: Developing a socio-hydrological model to assess community sensitivity to lake degradation, *Hydrol. Sci. J.*, 69, 1831–1847, <https://doi.org/10.1080/02626667.2024.2391926>, 2024.



- 685 Razavi, S., Duffy, A., Eamen, L., Jakeman, A. J., Jardine, T. D., Wheeler, H., Hunt, R. J., Maier, H. R., Abdelhamed, M. S., Ghoreishi, M., Gupta, H., Döll, P., Moallemi, E. A., Yassin, F., Strickert, G., Nabavi, E., Mai, J., Li, Y., Thériault, J. M., Wu, W., Pomeroy, J., Clark, M. P., Ferguson, G., Gober, P., Cai, X., Reed, M. G., Saltelli, A., Elshorbagy, A., Sedighkia, M., Terry, J., Lindenschmidt, K. E., Hannah, D. M., Li, K., Asadzadeh, M., Harvey, N., Moradkhani, H., and Grimm, V.: Convergent and Transdisciplinary Integration: On the Future of Integrated Modeling of Human-Water Systems, *Water Resour. Res.*, 61, e2024WR038088, <https://doi.org/10.1029/2024WR038088>, 1 February 2025.
- 690 Rotmans, J.: Societal innovations: Between dreams and reality lies complexity. Inaugural speech at Erasmus Research Institute of Management, Erasmus University Rotterdam., *J. Econ. Lit.*, 2005.
- Rotmans, J., Kemp, R., and van Asselt, M.: More evolution than revolution: transition management in public policy, *Foresight*, 3, 15–31, <https://doi.org/10.1108/14636680110803003>, 2001.
- 695 Rule, A., Cointet, J. P., and Bearman, P. S.: Lexical shifts, substantive changes, and continuity in State of the Union discourse, 1790-2014, *Proc. Natl. Acad. Sci. U. S. A.*, 112, 10837–10844, <https://doi.org/10.1073/pnas.1512221112>, 2015.
- Sawada, Y. and Hanazaki, R.: Socio-hydrological data assimilation: Analyzing human-flood interactions by model-data integration, *Hydrol. Earth Syst. Sci.*, 24, 4777–4791, <https://doi.org/10.5194/hess-24-4777-2020>, 2020.
- Schulz, C., Wolf, L. J., Martin-Ortega, J., Glenk, K., and Gischler, M.: Valuing water: A global survey of the values that underpin water decisions, *Environ. Sci. Policy*, 153, <https://doi.org/10.1016/j.envsci.2024.103685>, 2024.
- 700 Scoones, I.: The Politics of Sustainability and Development, *Annu. Rev. Environ. Resour.*, 41, 293-319, <https://doi.org/10.1146/annurev-environ-110615-090039>, 17 October 2016.
- Shanono, N. J. and Ndiritu, J.: Socio-hydrological modelling for incorporating human behaviour into storage–yield analysis, *Hydrol. Sci. J.*, 68, 2435–2448, <https://doi.org/10.1080/02626667.2023.2273413>, 2023.
- 705 Shrestha, A., Souza, F. A. A., Park, S., Cherry, C., Garcia, M., Yu, D. J., and Mendiondo, E. M.: Socio-hydrological modeling of the tradeoff between flood control and hydropower provided by the Columbia River Treaty, *Hydrol. Earth Syst. Sci.*, 26, 4893–4917, <https://doi.org/10.5194/hess-26-4893-2022>, 2022.
- Song, H., Tolochko, P., Eberl, J. M., Eisele, O., Greussing, E., Heidenreich, T., Lind, F., Galyga, S., and Boomgaarden, H. G.: In Validations We Trust? The Impact of Imperfect Human Annotations as a Gold Standard on the Quality of Validation of Automated Content Analysis, *Polit. Commun.*, 37, 550–572, <https://doi.org/10.1080/10584609.2020.1723752>, 2020.
- 710 Song, S., Wang, S., Fu, B., Dong, Y., Liu, Y., Chen, H., and Wang, Y.: Improving representation of collective memory in socio-hydrological models and new insights into flood risk management, *J. Flood Risk Manag.*, 14, <https://doi.org/10.1111/jfr3.12679>, 2021.
- Soroka, S., Fournier, P., and Nir, L.: Cross-national evidence of a negativity bias in psychophysiological reactions to news, *Proc. Natl. Acad. Sci. U. S. A.*, 116, 18888–18892, <https://doi.org/10.1073/pnas.1908369116>, 2019.
- Tallman, P. S., Piland, N. C., Villarmarzo, M., Victoria-Lacy, L., and Valdés-Velásquez, A.: Waterscapes meet socio-ecological models: A relational framework to examine water insecurity and human health and well-being, *People Nat.*, 6, 1036–1047, <https://doi.org/10.1002/pan3.10635>, 2024.



- Ulibarri, N., Scott, T. A., and Perez-Figueroa, O.: How does stakeholder involvement affect environmental impact assessment?, *Environ. Impact Assess. Rev.*, 79, <https://doi.org/10.1016/j.eiar.2019.106309>, 2019.
- Ulibarri, N., McCarty, R. J., Hamilton, M., Nesbitt, H. K., and Williamson, M. A.: Performative planning creates a values mismatch between wildfire plans and community needs, *Proc. Natl. Acad. Sci. U. S. A.*, 123, <https://doi.org/10.1073/pnas.2521536123>, 2026.
- Vanelli, F. M., Kobiyama, M., and De Brito, M. M.: To which extent are socio-hydrology studies truly integrative? the case of natural hazards and disaster research, *Hydrol. Earth Syst. Sci.*, 26, 2301–2317, <https://doi.org/10.5194/hess-26-2301-2022>, 2 May 2022.
- Veigel, N., Kreibich, H., De Bruijn, J. A., Aerts, J. C. J. H., and Cominola, A.: Content analysis of multi-annual time series of flood-related Twitter (X) data, *Nat. Hazards Earth Syst. Sci.*, 25, 879–891, <https://doi.org/10.5194/nhess-25-879-2025>, 2025.
- Wang, S., Xue, J., Chang, J., Sun, H., Li, X., and Liu, X.: An agent-based socio-hydrological modeling to identify the feedbacks between agricultural irrigation and ecological water conveyance tradeoffs in Hotan River basin, *J. Hydrol. Reg. Stud.*, 56, <https://doi.org/10.1016/j.ejrh.2024.102028>, 2024.
- Wei, J., Wei, Y., and Western, A.: 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, <https://doi.org/10.1016/j.gloenvcha.2016.12.005>, 2017.
- Xia, J., Dong, Y., and Zou, L.: Developing socio-hydrology: Research progress, opportunities and challenges, *J. Geogr. Sci.*, 32, 2131–2146, <https://doi.org/10.1007/s11442-022-2040-3>, 2022.
- Xiong, Y., Wei, Y., Zhang, Z., and Wei, J.: Evolution of China’s water issues as framed in Chinese mainstream newspaper, *Ambio*, 45, 241–253, <https://doi.org/10.1007/s13280-015-0716-y>, 2016.
- Xu, T. and Liang, F.: Machine learning for hydrologic sciences: An introductory overview, *Wiley Interdiscip. Rev.: Water*, 8, e1533, <https://doi.org/10.1002/wat2.1533>, 1 September 2021.
- Xu, W., Chen, J., Corzo, G., Xu, C. Y., Zhang, X. J., Xiong, L., Liu, D., and Xia, J.: Coupling Deep Learning and Physically Based Hydrological Models for Monthly Streamflow Predictions, *Water Resour. Res.*, 60, <https://doi.org/10.1029/2023WR035618>, 2024.
- Yang, G. and Sato, M.: The effect of culture on domestic water saving, *J. Environ. Manage.*, 373, <https://doi.org/10.1016/j.jenvman.2024.123928>, 2025.
- Yi, P., Xia, Y., and Long, Y.: Irony Detection, Reasoning and Understanding in Zero-shot Learning, *IEEE Trans. Artif. Intell.*, <https://doi.org/10.1109/TAI.2025.3579452>, 2025.
- Yu, D. J., Haeffner, M., Jeong, H., Pande, S., Dame, J., Di Baldassarre, G., Garcia-Santos, G., Hermans, L., Muneeppeerakul, R., Nardi, F., Sanderson, M. R., Tian, F., Wei, Y., Wessels, J., and Sivapalan, M.: On capturing human agency and methodological interdisciplinarity in socio-hydrology research, *Hydrol. Sci. J.*, 67, 1905–1916, <https://doi.org/10.1080/02626667.2022.2114836>, 2022.