https://doi.org/10.5194/egusphere-2026-3041
Preprint. Discussion started: 22 July 2026

Complementing drought impact databases with multilingual online sources: A European gap-filling experiment

Monika Hlavsová[1,2], Lauro Rossi[3], Kathrin Szillat[4], Kerstin Stahl[4], Mirko D'Andrea[3], Veit Blauhut[5], Gabriela Ivaňáková[6], Livia Labudová[6], and Miroslav Trnka[1,2]

[1]Department of Climate Change Impacts on Agroecosystems, Global Change Research Institute CAS, Brno, Czechia
[2]Department of Agrosystems and Bioclimatology, Mendel University in Brno, Brno, Czechia
[3]CIMA Research Foundation, Savona, Italy
[4]University of Freiburg, Faculty of Environment and Natural Resources, Freiburg, Germany
[5]Saxon State Ministry for Energy, Climate Protection, Environment and Agriculture, Dresden, Germany
[6]Slovak Hydrometeorological Institute, Bratislava, Slovakia

Correspondence to: Monika Hlavsová (hlavsova.m@czechglobe.cz)

Abstract. The use of drought impact databases has been increasingly recognized as essential for impact-based drought monitoring and risk assessment. The European Drought Impact Database (EDID) is the first continent-wide effort to systematically compile drought impact information across sectors and countries. However, the integration of existing regional and national datasets has revealed substantial spatial and temporal gaps, particularly for recent drought events and underrepresented countries. To address these deficiencies, we developed and evaluated a semiautomated retrospective gap-filling strategy that is based on multilingual online text searches, automated translation and classification, and expert validation. The workflow covered all official European Union languages, multiple drought events and nine affected sectors. It yielded several hundred new drought impact records and significantly improved coverage in selected countries. Nevertheless, the retrieval efficiency was low, and its performance varied strongly across languages and sectors. Our findings show that multilingual online data mining can complement drought impact databases but cannot replace sustained local monitoring and expert-driven data collection. Robust continental-scale drought impact monitoring, therefore, requires hybrid strategies that combine automated searches with national and regional inventories. The lessons learned from the EDID gap-filling experiment can guide future operational drought impact monitoring and similar efforts in other natural hazard domains.

1 Introduction

Drought impacts represent the tangible manifestation of water scarcity on environmental, economic and social systems, and their systematic documentation is essential for moving from hazard-based to impact- and risk-based drought management (Blauhut et al., 2016; Hagenlocher et al., 2023; Sutanto et al., 2019; Wilhite et al., 2007). Although drought hazards are routinely monitored using hydrometeorological indicators, comparable operational monitoring of drought impacts remains challenging because of the heterogeneous nature of impacts, their delayed occurrence and the lack of harmonized data collection practices (Shyrokaya et al., 2025; Stahl et al., 2016; United Nations Office for Disaster Risk Reduction, 2021).

Despite their importance, systematic drought impact data remain scarce and fragmented (Wilhite et al., 2007; Stahl et al., 2016; Hagenlocher et al., 2023). Unlike hydrometeorological observations, impacts are rarely measured directly and consistently across regions and sectors. Instead, they are documented mostly through heterogeneous sources such as media reports, governmental bulletins, expert assessments, and observer networks (Blauhut et al., 2016; Jakubínský et al., 2019; Trnka et al., 2020; Bartošová et al., 2022). These sources differ in terms of spatial coverage, temporal resolution, sectoral focus, and reporting practices, which leads to substantial inconsistencies and biases. Consequently, drought impact databases typically suffer from uneven spatial and temporal coverage, which limits their usefulness for comparative analyses, model development and operational drought risk assessment (Sutanto et al., 2019; European Commission. Joint Research Centre., 2024).

The European Drought Impact Database (EDID) was established to address this gap at the European scale by integrating drought impact information across countries, sectors, and data sources into a single operational framework (Stahl et al., 2025). Building upon earlier research databases and national monitoring efforts, the EDID provides the most comprehensive collection of drought impact records for Europe to date. By consolidating historical and recent drought impact information across countries and sectors, the EDID provides an unprecedented baseline for understanding drought impacts in Europe and for supporting impact-based drought services.

However, the initial integration phase revealed pronounced spatial and temporal gaps in impact coverage, particularly for recent drought events and countries without established national impact reporting systems. These gaps reflect not necessarily reflect the absence of drought impacts but rather differences in reporting practices, language barriers, media visibility, and data availability. Addressing such gaps is therefore a critical step toward making drought impact databases more representative and suitable for continental-scale analysis and operational use.

To improve the completeness of the EDID for operational use, a targeted gap-filling strategy was needed. Among the limited options for retrospective impact data collection at a continental scale, online text sources—such as news media and publicly available reports—offer geographically distributed, event-specific information that can potentially be accessed retrospectively and across national borders. However, the usability of online data is constrained by strong language biases, heterogeneous reporting styles, declining availability of older content and substantial noise unrelated to drought impacts (Stahl et al., 2016; Hlavsová et al., 2025). Although previous studies have successfully applied text mining approaches to drought impact detection at national or event-specific scales (Madruga De Brito et al., 2020; Sodoge et al., 2023; Wang et al., 2025), continental-scale, multilingual applications remain rare and largely untested.

In response to the identified data gaps in the EDID, a semiautomated, multilingual gap-filling strategy was developed and applied as part of the EDORA project. The approach combines sector-specific online search queries translated into all official

https://doi.org/10.5194/egusphere-2026-3041
Preprint. Discussion started: 22 July 2026

European Union languages, automated retrieval and translation of online text content, machine-assisted text classification and expert-guided manual validation. This study represents the first systematic attempt to retrieve drought impact information across multiple drought events, sectors and languages at a pan-European scale using online text sources.

By critically evaluating both the added value and the shortcomings of this approach, we aim to clarify the role of automated online searches within a broader, hybrid framework for drought impact monitoring. Rather than aiming to replace existing impact monitoring efforts, the gap-filling strategy was designed as a complementary tool to improve coverage where other data sources are lacking. Given the known limitations of online data, including reporting biases, low signal-to-noise ratios and technical barriers to automated text extraction, a critical evaluation of the performance of this method is essential. Understanding both the added value and the limitations of multilingual online gap-filling is crucial for designing robust hybrid drought impact monitoring systems that combine automated methods with sustained local and national data collection.

The objectives of this study are therefore to (i) describe the design and implementation of a multilingual, semiautomated gap-filling workflow for drought impact data in Europe; (ii) evaluate its ability to address spatial and temporal gaps in the EDID; and (iii) assess the strengths and limitations of large-scale online text mining for drought impact data collection in comparison with local and manually curated impact inventories. By doing so, this study provides methodological insights and practical guidance for future drought impact monitoring efforts in Europe and beyond.

2 Data and Methods

The gap-filling process consisted of four steps (Figure 1): a) identification of the spatial and temporal gaps in drought impact data; b) development of language- and system-specific search queries for automated online search; c) semiautomated full-text acquisition, translation and preclassification; and d) manual drought impact entry processing.

2.1 EDID as a baseline

The EDID served as the baseline dataset for the gap-filling exercise. The EDID is an operational, geospatial database that consolidates drought impact records from multiple research-based and operational sources across Europe into a harmonized structure. The impact records in the EDID are classified on the basis of the affected system, spatial reference, temporal occurrence, and severity and are designed to support both scientific analysis and operational drought monitoring. In this study, the initial pilot version of the EDID was used to identify spatial and temporal gaps in drought impact coverage. The database includes impact records derived from earlier research inventories, national monitoring systems, and targeted project-based data collections (Stahl et al., 2025). Although the EDID provides the most comprehensive drought impact dataset for Europe to date, its content reflects the availability and focus of underlying data sources, thus leading to uneven representation across countries, drought events, and impacted systems.

https://doi.org/10.5194/egusphere-2026-3041
Preprint. Discussion started: 22 July 2026

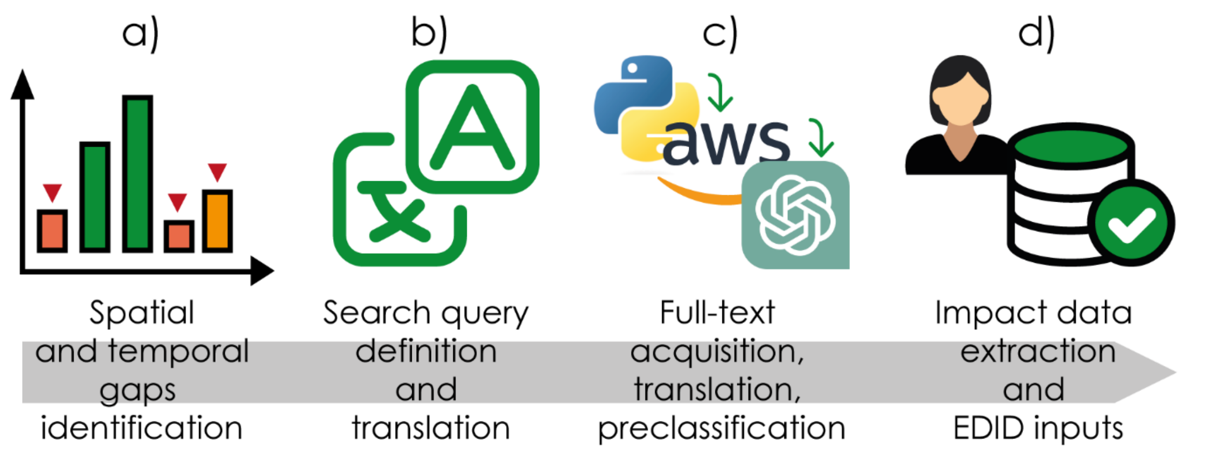

Figure 1: Gap-filling process including a) identification of the spatial and temporal gaps in drought impact data; b) development of language and system-specific search queries for automated online search; c) semiautomated full-text acquisition, translation and preclassification; and d) manual drought impact entry processing.

2.2 Identification of spatial and temporal data gaps

The first step of the gap-filling workflow consisted of identifying underrepresented drought events and regions in the EDID. For each European country, the number of drought impact records was evaluated in relation to known major drought episodes, as indicated by widely used drought indicators and indices available through the European Drought Observatory. This comparison enabled the identification of countries and periods where severe drought conditions were documented by hazard indicators but were only sparsely represented in the impact database. Attention was given primarily to drought events that occurred after the year 2010, when online media coverage become more widespread and potentially accessible for retrospective searches. The outcome of this step was a targeted selection of countries, drought periods, and affected systems for which additional impact information was likely to exist but was missing from the EDID.

2.3 System-specific multilingual search, translation and preclassification

The given structure of the EDID includes nine impacted systems (Table 1), and the detailed gap-filling process required separate searches for drought impacts that cover all of these impacted systems. For this reason, we used the initial EDID content for keyword analysis to describe each of the impacted systems with the combination of keywords with the greatest potential to align with online media text content. The respective keywords were limited to nouns with the highest frequency of occurrence in the EDID for the impacted system. The final set of keywords was appended with the word “drought” and connected by logical operators to create a full query that was usable for automated searching via Google search (Table 1). Some of the selected keywords were stripped from word endings to increase the potential of matching the other forms of these words with the online text content. After the finalization of the queries in English, they were translated into all the official European languages by a group of volunteer experts from the field. The invaluable role of native speaker experts also enabled

https://doi.org/10.5194/egusphere-2026-3041
Preprint. Discussion started: 22 July 2026

the addition of language-specific keywords and synonyms as parts of specific queries if needed. The full list of translated queries can be found in the appendix (Appendix 1). The translated queries were then tested in multiple rounds of test searches (depending on how successful the combination of a specific language and system was) and shortened or edited in the case of very poor results (no successful searches or irrelevant content).

Table 1: Definitions of system search queries in English that were used for multilingual translations and language-specific searching

System	Search query
Agriculture - annual crops	drought AND agricultur* AND (crop OR yield OR cereal OR maize OR corn OR potato)
Agriculture - permanent crops	drought AND (fruit OR orchard OR grape OR wine)
Agriculture - livestock	drought AND (fodder OR cattle OR livestock)
Ecosystems - aquatic	drought AND fish AND (die OR oxygen OR algae OR water OR wetland)
Ecosystems - terrestrial	drought AND forest AND (tree OR die OR leaf OR stress)
Energy - hydropower	drought AND energ* AND (water OR hydro OR reserv* OR river OR electr*)
Energy - thermal	drought AND energ* AND (nuclear OR plant OR cool)
Inland navigation	drought AND transport AND (river OR naviga* OR ship OR cargo OR level)
Public water supply	drought AND drink AND (water OR suppl* OR ban OR shortage OR lack OR source)

The multilingual search queries were used to conduct automated online searches using Python-based scripts that emulate manual search behavior through the Google search engine interface using an alternating method developed for the global drought impact report near-real-time monitoring tools (Hlavsová et al., 2025). The searches were constrained by language, country and period according to the identified data gaps. For each query, search results were retrieved from the first ten pages of results, which reflected typical user relevance rankings. For each identified search result, the uniform resource locator (URL) was stored for further processing. Full-text extraction of the corresponding webpages was attempted using the Trafilatura Python package (Barbaresi, 2024). Owing to technical limitations such as paywalls, advertisements, dynamic content and access restrictions, full-text extraction was not successful for all URLs. In such cases, articles were flagged for potential manual inspection during later processing steps.

The extracted text content was translated into English using an Amazon AWS Translation service (Amazon Web Services, n.d.) via an application programming interface (API). Translation into a common language was necessary to enable consistent downstream processing and expert evaluation across all the countries and languages. Following translation, the texts were subjected to automated preclassification using a large language model (GPT 3.5) API. A structured prompt (Figure 2) was

https://doi.org/10.5194/egusphere-2026-3041
Preprint. Discussion started: 22 July 2026

designed to extract concise summaries of drought-related impacts and to assign them to the relevant affected systems, together with basic spatial and temporal information. The output format was constrained to a tabular structure to facilitate further processing.

> "You are a data extractor system that given a news article generates a table of impacts of drought for each of the following sectors: “agriculture", "inland navigation", "ecosystem", "public water supply", "energy". INSTRUCTIONS: - If the article does not contain informations on a sector, the corresponding row in its entirety must be skipped. - Output only the table no comments or other text.- Keep the description under 80 words. ARTICLE: "{title}" ({date}) TABLE HEADER: "sector, location, period, description" {text}"

Figure 2: Final version of the prompt used to extract concise summaries

The automated classification step was used exclusively as a filtering and summarization tool. No impact records were included in the EDID solely on the basis of automated classification results without expert review. Domain experts reviewed all automatically generated summaries and classifications manually. During this step, irrelevant or duplicate entries were discarded, ambiguous cases were clarified through inspection of the original sources, and relevant drought impact reports were reformulated to comply with the data model of the EDID. Validated impact records were assigned spatial references, temporal information, and affected systems in accordance with EDID guidelines. Where necessary, severity levels were assigned on the basis of qualitative descriptions and predefined classification rules. Only records that met the minimum quality and completeness requirements of the EDID were integrated into the database. This supervised validation ensured consistency with existing EDID content while preserving traceability to original sources.

2.4 Evaluation metrics and analysis

The performance of the gap-filling strategy was evaluated using several descriptive metrics, including (i) the number of retrieved search results versus the number of validated drought impact records; (ii) the contribution of gap-filled records to the total EDID content by country; and (iii) differences in success rates across languages and affected systems (Table 2 and Figure 3). These metrics were used to assess both the effectiveness and limitations of the multilingual online gap-filling approach. No attempt was made to interpret the resulting impact counts as quantitative measures of drought severity or frequency; instead, the analysis focused on methodological performance and data coverage.

3 Results

3.1 Evaluation of the integration of the gap-filling results

The semiautomated multilingual gap-filling workflow generated a large volume of candidate material from online sources. Across all the countries, systems and drought periods that were included in the targeted searches, more than 61,000 individual search results (online articles) were retrieved, whereas the search for the selected underrepresented drought events (year and country combinations) was limited. After automated text extraction, translation, classification, and subsequent expert validation, 739 drought impact records (after nonrelevant articles, duplicates, articles that report on other events or countries, etc., were removed) were identified as relevant and suitable for integration into the EDID (Table 2). Finally, the gap-filling process provided 1090 new entries for the EDID. In addition to EDID inputs for preselected underrepresented events, the gap-filling procedure also yielded new entries for countries (Germany and Malta) that were not included initially in the underrepresented list but were reported by media articles from other countries.

Table 2: Results of gap-filling efforts for the identified underrepresented drought events. The underrepresented event years , which are highlighted in red represent the years with the most EDID entries

Country	Underrepresented drought events	Web crawling search results	Web crawling relevant results	EDID entries
Albania	2020, 2022	0	0	0
Austria	**2022**	0	0	20
Belgium	2011, 2014, 2015, 2016, 2020, **2022**	0	0	16
Bulgaria	2012, 2016, 2022	3291	11	13
Croatia	2020, **2022**	842	28	30
Cyprus	2008, **2018**	0	0	5
Czech Republic	2011	5937	154	155
Denmark	2011, 2012	4567	5	9
Estonia	2011, **2018**	2292	47	45
Finland	2013, **2018**, 2019, 2020	2651	56	48
France	2021, **2022**	4845	106	166
Germany	-	29	2	20
Greece	2012, 2013, 2020, **2021**, 2022	5290	27	25
Hungary	2011, 2012, 2016, **2022**	2462	70	73
Ireland	2020, **2022**	229	0	13
Italy	2016, **2018**	4368	23	43
Latvia	2011, **2018**	2804	16	34

https://doi.org/10.5194/egusphere-2026-3041
Preprint. Discussion started: 22 July 2026

Country	Underrepresented drought events	Web crawling search results	Web crawling relevant results	EDID entries
Lithuania	2011, **2018**, 2019	2530	13	6
Luxemburg	2015, 2016, 2020, **2022**	0	0	10
Malta	-	720	6	7
Montenegro	2020	0	0	1
Netherlands	2014, 2015, 2016, **2018**, 2020, 2022	4545	52	62
North Macedonia	2020	0	0	1
Norway	2013, 2018, 2020, 2021	0	0	10
Poland	2014, 2018, 2019, 2020	2221	29	63
Portugal	2015, 2016, 2020, **2022**	0	0	20
Romania	**2022**	0	0	26
Serbia	**2022**	0	0	6
Slovakia	2011, 2015, 2018	562	11	11
Slovenia	2011, **2022**	0	0	9
Spain	2014, 2015, 2019, 2020, **2022**	4875	68	94
Sweden	2013, **2018**, 2019, 2020	0	0	18
Switzerland	2016	0	0	4
Turkey	2013, 2014, 2021	0	0	1
United Kingdom	2020, **2022**	5940	15	26

The system most impacted at the country level clearly varies across Europe (Figure 3a). In Western and Northern Europe, the gap-filling dataset most frequently identifies impacts on public water supply system. In contrast, agriculture - particularly annual crops - emerges as the prevailing impacted system across Eastern and Southern Europe. Other systems, including agriculture (livestock and permanent crops), hydropower production, inland water navigation, and terrestrial ecosystems, each of which dominates specific regions, play more localized but still notable role. The spatial distribution of the most frequent impact year highlights the strong dominance of recent drought events across Europe (Figure 3b). The year 2022 was the most frequently reported impact year in the majority of countries, particularly across Western, Central, and parts of Eastern Europe. Earlier events, such as those in 2018 and 2020, were more prominent in Northern Europe and selected southeastern regions,

https://doi.org/10.5194/egusphere-2026-3041
Preprint. Discussion started: 22 July 2026

whereas those in 2015 were identified only in the case of Poland. Overall, the pattern reflects a clear concentration of reported impacts in the most recent major drought episodes.

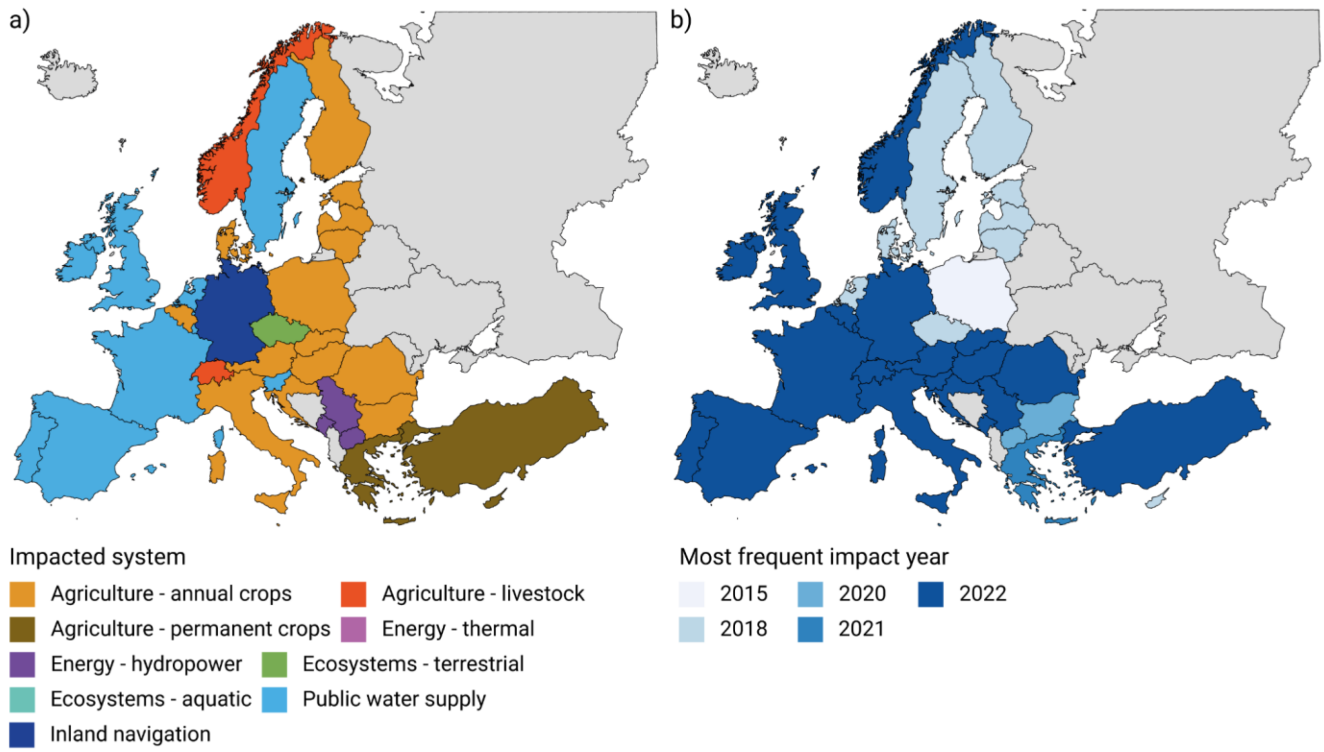

Figure 3: Data obtained using the gap-filling strategy that indcates a) the most frequently impacted system in the gap-filling results for each country and b) the most frequent year of drought impact in the gap-filling results

The gap-filling strategy provided data for only limited set of years, including mostly the period of 2010–2023 (1084 impact records), with only 6 records reporting on impacts from years 1976, 1988, 2003 and 2008. A comparison of the yearly numbers of impact records in the gap-filling dataset (Figure 4b) and existing EDID content (Figure 4a) also reveals differences in the impact report severity classes. Most of the impact records in the existing EDID content (Figure 4a) are classified as severe, whereas the gap-filling dataset (Figure 4b) has a greater share of moderate impacts. A comparison of the impacted systems across the datasets revealed similarities in the most frequently impacted systems of agriculture-annual crops and public water supply. The existing EDID content (Figure 4c) shows a greater share of impact reports on systems ecosystems-aquatic and inland navigation. The gap-filled dataset (Figure 4d) shows higher shares of impact records that represent systems agriculture-livestock and ecosystems-terrestrial. For both datasets, the least frequent system is energy-thermal.

https://doi.org/10.5194/egusphere-2026-3041
Preprint. Discussion started: 22 July 2026

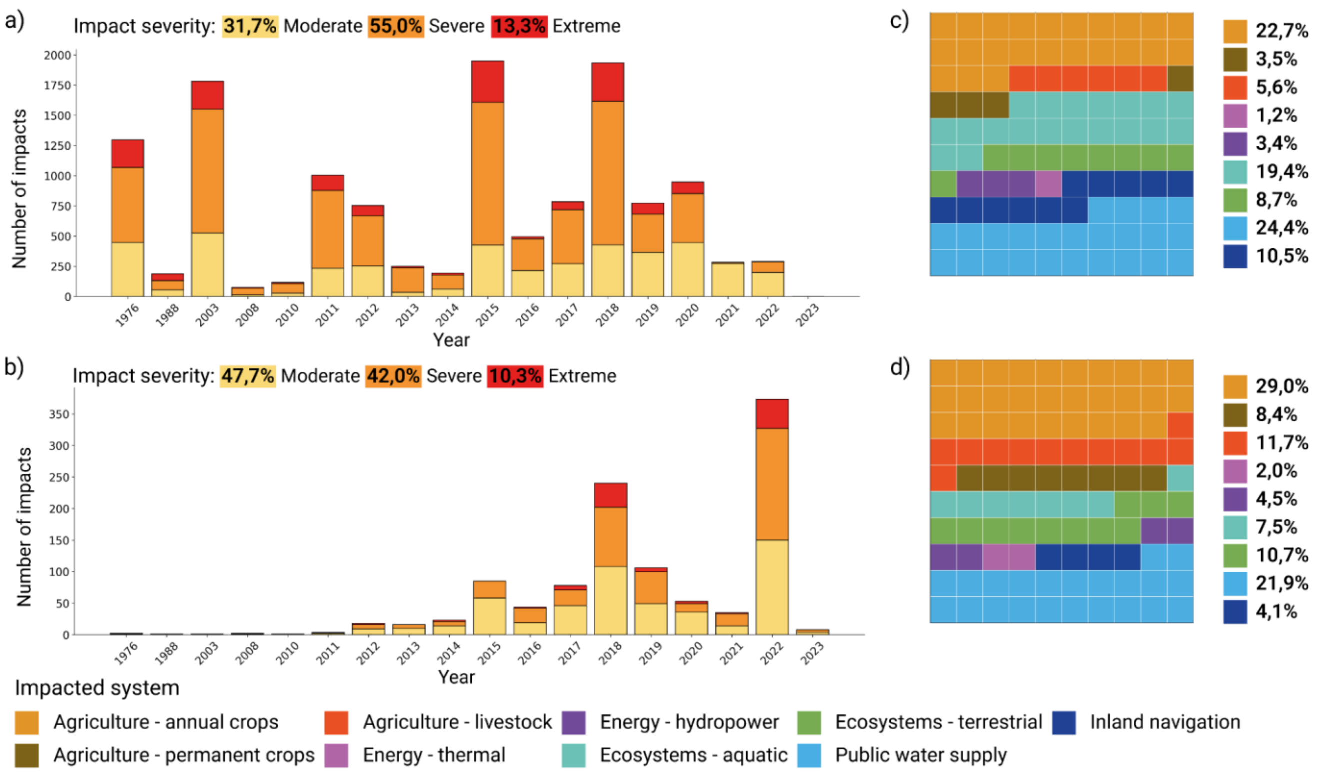

Figure 4: Comparison of existing records in the European Drought Impact Database (EDID) and those obtained through the gap-filling strategy. Panels (a) and (b) show the annual numbers of impact records by severity class for the original EDID and the gap-filled dataset, respectively. Panels (c) and (d) present the corresponding shares of impacted systems in the EDID and gap-filled dataset, respectively.

Finally, the efficiency and results of the gap-filling workflow were evaluated. The number of successful and relevant hits per 100 searches varies across countries (Figure 5a), with higher numbers (more than two new impact report EDID entries per 100 search hits) in Czechia, Hungary, Croatia, France, and Finland. In contrast, the lowest success rates were recorded for the United Kingdom, Ireland, and Denmark (Figure 5a) (with under 0.33 new impact report EDID entries per 100 search hits).

https://doi.org/10.5194/egusphere-2026-3041
Preprint. Discussion started: 22 July 2026

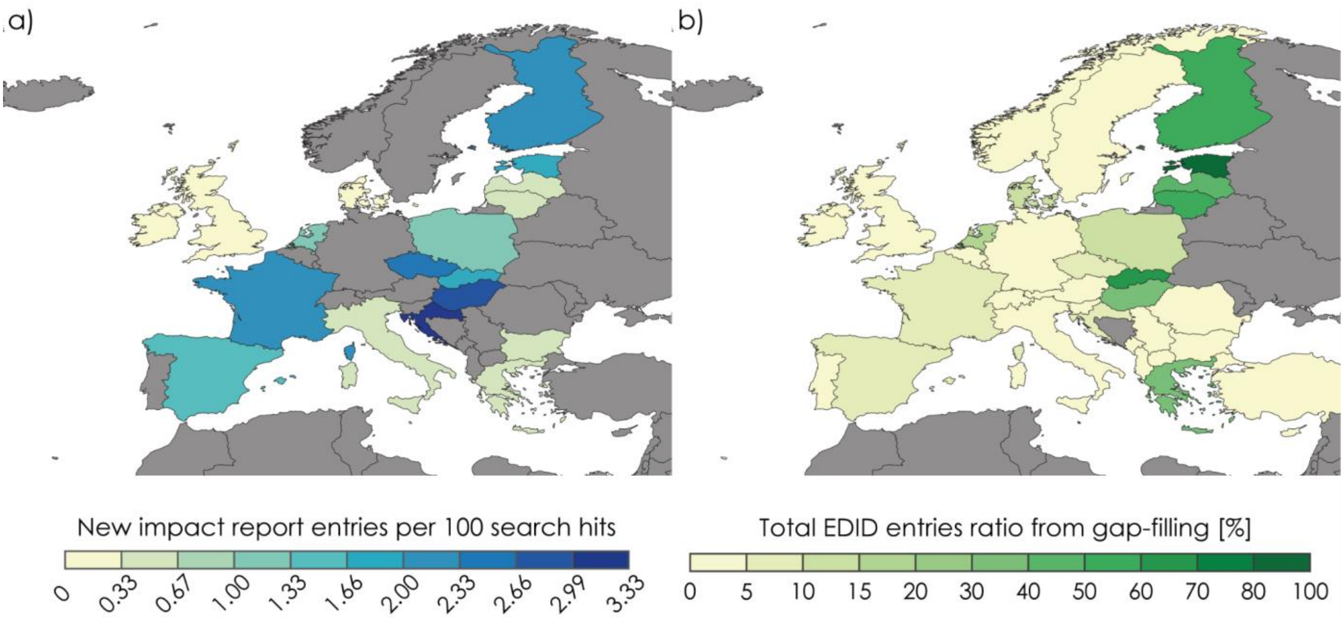

Figure 5: Results of semiautomated gap-filling by country, including a) the success rate, which is represented by the number of new drought impact reports per 100 search hits, and b) the total share of EDID drought impact reports per country imported as a result of the gap-filling workflow

For some of the countries that were represented by fewer entries in the first version of the EDID, gap-filling proved to be an effective tool for populating the database (Figure 5b). For Lithuania, Finland, Slovakia, Malta, and Estonia, more than 50 % of the total entries in the EDID originated from the gap-filling process.

In addition to retrieval performance, the human validation effort required to process the preclassified search results was quantified. A total of 37 hours of expert review were logged during manual evaluation of the automatically preprocessed entries. During this period, approximately 23,948 candidate entries were assessed, which resulted in 434 validated drought impact records suitable for integration into the EDID. Although this process also included the elimination of nonrelevant or duplicate articles, these figures provide an operational estimate of the human effort required to convert multilingual automated search outputs into quality-controlled drought impact records.

3.2 National sectoral impact databases versus semiautomated gap-filling

The summer of 2018 was characterized by prolonged meteorological and soil moisture drought conditions across Slovakia, with indices indicating extreme deficits in spring and early summer and localized severity (Pecho et al., 2018). This period was part of a broader multiyear hot–dry episode that affected much of Central and Northern Europe and has been documented in continental synthesis studies of the 2018 drought (Blauhut et al., 2022). The gap-filling procedure for Slovakia yielded 11 new EDID records across three drought events (Table 2), thereby improving impact data coverage in the pilot version of the

https://doi.org/10.5194/egusphere-2026-3041
Preprint. Discussion started: 22 July 2026

database (Figure 3, b)). Only two of these records referred to the 2018 drought event, both of which were classified under Agriculture – annual crops system, with reported impacts on cereal and maize yields. Although the search was restricted to the underrepresented drought events of 2011, 2015 and 2018, most retrieved records of drought impacts were related to more recent years (2022 and 2023) and, in many cases, lacked precise spatial information (Table 3).

Table 3: Summary of the gap-filling results for Slovakia

Year	Starting month	NUTS3	Location	EDID system	Impacted commodity
2017	7	SK021	Trnavský kraj	Agriculture - permanent crops	apricot, walnut
2018	8	SK032	Banskobystricky kraj	Agriculture - annual crops	wheat, corn
2018	7		Slovakia	Agriculture - annual crops	wheat, oil seed rape
2022	8		Slovakia	Agriculture - annual crops	alfalfa, sugar beet, soy, corn
2022	8		Slovakia	Agriculture - livestock	cow
2022	6		Slovakia	Agriculture - annual crops	corn
2022	8		Slovakia	Agriculture - livestock	cow, sheep
2022	8	SK022	Trenčínský kraj	Agriculture - annual crops	alfalfa
2022	7	SK023	Nitranský kraj	Ecosystems - aquatic	
2022	8		Slovakia	Ecosystems - aquatic	
2023	3		Slovakia	Agriculture - annual crops	onion

In contrast, the national drought monitoring network that is operated in Slovakia provides substantially more detailed and spatially explicit information on drought impacts (Figure 6). The system integrates regular reports from farmers, foresters and other agricultural experts who assess both local drought conditions and system-specific impacts at the district level (Trnka et al., 2020; Bartošová et al., 2022). The results from the monitoring network for 2018 illustrate the distribution of drought impacts on crop yield records across districts and reveal the prevailing occurrence of minor to moderate impacts, as captured by the monitoring network (Figure 6a). For most of the regions, impacts on forestry, wine growing or fruit growing were also reported.

https://doi.org/10.5194/egusphere-2026-3041
Preprint. Discussion started: 22 July 2026

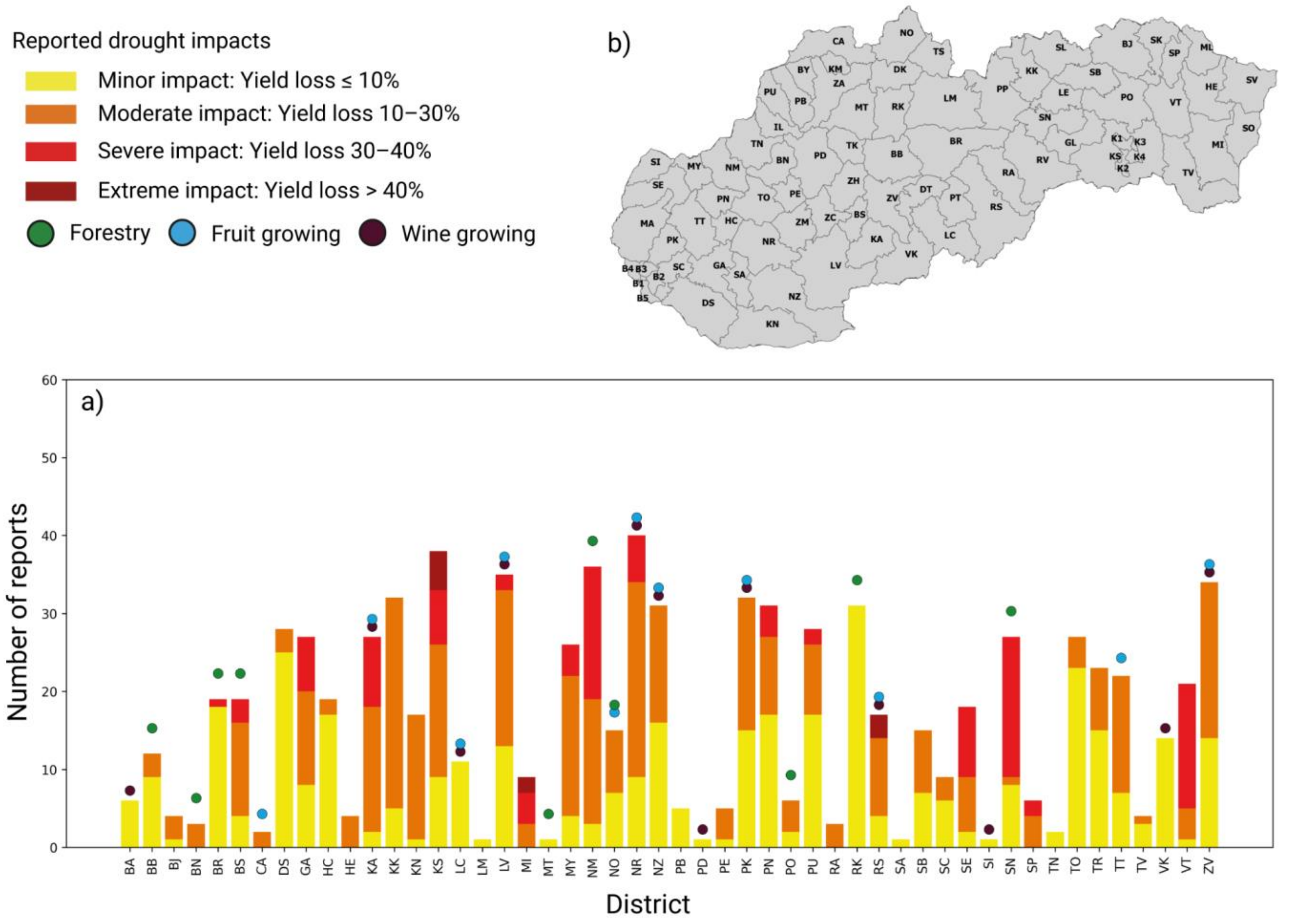

Figure 6: Impacts of the 2018 drought in the districts of Slovakia as reported by the national drought impact monitoring network of agricultural and forestry experts: a) A stacked bar chart that shows the distribution of impacts in districts, including impacts on yield, forestry, wine growing and fruit growing, and b) a map of the districts of Slovakia.

4 Discussion

This study set out to design and test a multilingual, semiautomated gap-filling workflow for drought impact data in Europe to improve the coverage achievable through standard approaches to impact data collection. We aimed to evaluate its effectiveness in addressing spatial and temporal gaps in the European Drought Impact Database (EDID), and to assess the strengths and limitations of large-scale online text mining in comparison with local and manually curated impact inventories. The results demonstrate that the workflow successfully increased coverage in underrepresented regions and periods; its efficiency varied strongly across languages and systems and required substantial expert validation, and it provided more meaningful results only for the most recent drought episodes, which was attributable to the fading memory of internet sources. These findings provide both methodological insights and practical guidance for the future development of hybrid drought impact monitoring systems at the continental scale.

https://doi.org/10.5194/egusphere-2026-3041
Preprint. Discussion started: 22 July 2026

A major contribution of this study is the elaboration and testing of multilingual search terms for drought impacts in the EU context. Previous studies have often focused on one country (Madruga De Brito et al., 2020; Sodoge et al., 2023; Jobbová et al., 2024), and international efforts have relied heavily on either English-language reports or the human capital of collaborative research projects, which has resulted in biases toward the partner countries involved (Stahl et al., 2016; Jakubínský et al., 2019; Stephan et al., 2021). Multilingual online gap-filling proved capable of making meaningful contributions to drought impact databases at the continental scale, particularly by improving coverage in countries and periods that are otherwise weakly represented (Figure 5b). Despite the low overall retrieval efficiency, the gap-filling workflow successfully identified 1090 new EDID entries that would not have been captured through the integration of existing research inventories and national monitoring systems alone. These records represent a substantive addition to the EDID and demonstrate the feasibility of retrospective multilingual impact discovery at the continental scale. However, they further illustrate that a key strength of the approach lies in its ability to access locally reported information across linguistic boundaries.

Many of the validated impact records originated from local-language sources without English equivalents, which underscores the importance of multilingual searches for the collection of pan-European drought impact data. These findings align with those of previous national and event-focused studies (Stahl et al., 2016; Jakubínský et al., 2019; Madruga De Brito et al., 2020; Wang et al., 2025) while extending their scope to a continental, multisectoral context. As such, multilingual online gap-filling provides a valuable mechanism for discovering otherwise inaccessible impact information (Hlavsová et al., 2025) and increasing the diversity of reported impacts in large-scale databases. The search terms now provide other researchers and practitioners with a set of terms to continue their own efforts and, perhaps over time, develop national monitoring with regular text report EDID inputs. These lessons are important for all multilanguage regions of the world and indicate that although English constitutes a lingua Franca, its reach is far from universal, at least in Europe.

Moreover, the results highlight the fundamental limitations of online data as a primary source of drought impact information. The high volume of nonrelevant material encountered during the search process illustrates the difficulty of extracting relevant, high-quality impact information from the open online environment (De Bruijn et al., 2019), as was also revealed by the low retrieval efficiency and high volumes of nonrelevant results (Figure 5a). This inefficiency reflects the structural characteristics of online content rather than being a technical shortcoming of the workflow itself. Drought-related terminology is frequently used in nonimpact contexts (e.g., the use of the term “drought” in the context of sports, as in the article “Michail Antonio regretting changing game amid drought” (BBC, 2023)), and relevant reports are embedded within large volumes of unrelated material. The extraction workflow also encountered technical bottlenecks related to automated content scraping and data handling, including interference from advertisements, dynamic web elements, and unrelated content.

Moreover, online impact reporting is subject to strong and uneven biases. Media attention varies substantially across events, countries, cultures, and systems, and is influenced by political, economic, and societal factors that are largely independent of

https://doi.org/10.5194/egusphere-2026-3041
Preprint. Discussion started: 22 July 2026

drought severity (Houston et al., 2015; Kong and Purves, 2026). Even within the scope of natural hazards, drought events tend to be underreported (Sufi and Alsulami, 2025). Similarly, the timing of an event and its compounding with other more pressing issues lead to underreporting bias. Consequently, the absence of online impact reports cannot be interpreted as evidence of absent or minor drought impacts.

The key limitations, which are connected to the use of online content, include is the limited temporal accessibility of online content, with older reports often inaccessible or no longer retrievable or not properly linked (Hansen and Paul, 2015), which constrains the applicability of retrospective searches beyond the past two decades (Hansen and Paul, 2015; Chapekis et al., 2024). This constraint was reflected in the gap-filling results, where the collected data mostly represented the latest drought episodes (Table 2 and Figure 4b), despite the targeted search for earlier events, and pivots further online content mining efforts toward operational monitoring rather than large-scale retrospective data search (Hlavsová et al., 2025). The differences in online-extracted content also include a shift in the report severity pattern that was already noted for the full version of the EDID (Stahl et al., 2025). Compared with data sourced from traditional newspapers and other resources, the impact report data sourced from online resources and media outputs show greater share of the lower severity class. This finding was confirmed by the gap-filling dataset, which was extracted in a semiautomated manner, solely on the basis of available online content (Figure 4b).

The aforementioned limitations become particularly evident when the gap-filling results are compared with those of the national drought monitoring network in Slovakia during the 2018 drought event. Although the retrospective multilingual search yielded only sparse references to 2018 and was dominated by more recent years (2022–2023) (Table 3), the national sectoral monitoring network documented substantial and spatially explicit impacts (Figure 6a). The Slovak system integrates structured reports from farmers and agricultural experts, including quantitative yield reduction estimates and qualitative severity assessments, thereby enabling near-real-time documentation of impacts. Compared with the limited number of gap-filled records, the national monitoring network captures a broader spectrum of impacts with higher spatial precision and clearer attribution to specific drought episodes. This comparison confirms that retrospective online searches alone are insufficient for comprehensive impact data collection and once again highlights the critical role of locally embedded monitoring systems.

Taken together, these findings indicate that multilingual semiautomated gap-filling is best understood not as a replacement for structured monitoring systems but as a complementary discovery tool. Its primary value lies in identifying previously undocumented impacts, particularly in regions that lack institutionalized monitoring frameworks, and in expanding the linguistic and thematic breadth of continental impact databases. With respect to targeted retrospective analyses of specific events, system-specific searches within national or local media sources may offer greater efficiency (Jakubínský et al. 2019; Madruga De Brito et al. 2020). Large-scale impact monitoring frameworks benefit most from integrating such automated discovery approaches with already operational and robust monitoring systems that are based on observer networks, institutional

https://doi.org/10.5194/egusphere-2026-3041
Preprint. Discussion started: 22 July 2026

reporting and systematic impact assessment (Trnka et al., 2020; Stahl et al., 2025; VigiEau, 2026), as well as regular online scanning to reduce long-term information loss (Hlavsová et al., 2025; Drought Impact Reporter). With continued improvements in text classification and semantic language modeling, these methods may increasingly support preselected, system- and language-specific sources, thus reducing the manual workload while maintaining quality control.

5 Conclusions

In this study, evaluated the potential and limitations of multilingual, semiautomated gap-filling for continental-scale drought impact monitoring were evaluated. By analyzing its ability to reduce spatial and temporal gaps in the EDID and comparing it with structured national monitoring systems, we provide a critical assessment of large-scale online text mining as a data collection strategy. The results confirm that semiautomated searching and text analysis represent a viable approach for retrospective identification of drought impact information in online-accessible sources, particularly where resources, time or access to established monitoring networks and loss reports are limited. However, the method is subject to important constraints, including limited long-term accessibility of online content, reporting biases and persistent technical bottlenecks in automated text extraction and processing. Although multilingual online gap-filling can meaningfully complement existing impact databases, it cannot replace sustained local monitoring and expert-driven data collection. Robust continental-scale drought impact monitoring therefore requires hybrid strategies that combine automated discovery tools with structured national and regional impact inventories.

https://doi.org/10.5194/egusphere-2026-3041
Preprint. Discussion started: 22 July 2026

Appendix 1 Full list of translated search querries

Table A1: Full list of search query translations for the Agriculture - annual crops system

Language	Translated search query
BG	(засушаване OR суша) AND земеделие OR селско стопанство AND (земеделск култур OR добив OR зърнени OR царевица OR картоф OR зеленчук)
CZ	sucho AND zeměděl AND (plodin OR výnos OR obil OR kukuř OR brambor)
DA	tørke AND landbrug AND (afgrød OR høst OR korn OR hvede OR kartoffel OR grøntsag)
DE	(Trocken OR Dürre) AND Landwirtschaft AND (Ernte OR Ertr OR Getreide OR Mais OR Kartoffel OR Gemüse)
EN	drought AND agricultur AND (crop OR yield OR cereal OR maize OR corn OR potato)
EL	ξηρασία AND γεωργία AND (καλλιέργεια OR απόδοση παραγωγής OR δημητριακό OR καλαμπόκι OR πατάτα OR λαχανικό)
ES	sequía AND agricultur OR regad OR riego OR secano AND (cultiv OR cosech OR cereal OR maíz OR hortic)
ET	(põud OR põua) AND (teravili OR vili OR vilja OR kartul OR juurvili)
FI	(kuivuus OR kuivakausi) AND (maatalou OR maanviljely) AND (kasvi OR sato OR vilja OR maissi OR peruna OR vihannes)
FR	sécheresse AND agriculture AND (culture OR rendement OR céréale OR maïs OR maraich)
GA	triomach AND talmhaíocht AND (barr OR táirgeacht OR grán OR grán buí OR práta OR glasra OR plandúil)
HR	suša AND poljoprivreda AND (usjev OR prinos OR žitarice OR kukuruz OR krumpir OR povrće)
HU	aszály AND mezőgazdaság AND (termény OR termés OR gabona OR kukorica OR burgonya OR zöldség)
IT	siccità AND agricol AND (coltiv OR seminativ OR produzion OR cereal OR mais OR frumento OR grano OR riso OR orzo patat OR ortaggi OR orticol OR resa OR rese)
LT	sausr AND žemdirbyst AND (pasėl OR derl OR javai OR kukurūz OR bulv OR daržov OR raps)
LV	sausums AND (zemkopība OR lauksaimniecība) AND (labība OR raža OR graudaugi OR kukurūza OR kartupeļi OR dārzeņi)
MT	nixfa AND agrikoltura AND (għelejjel OR rendiment OR ċereali OR qamħirrun OR patata OR ħxejjex)
NL	(droogte OR watertekort OR watergebrek) AND landbouw AND (gewas OR opbrengst OR graan OR mais OR aardappel OR groente)
PL	susza AND rolnicza AND (zbiory OR plony OR zboże OR kukurydza OR ziemniaki OR warzywa gruntowe)
pt-PT	seca AND agricultura AND (culturas OR produção OR cereais OR milho OR batata OR legumes OR produtos hortícolas)
RO	secetă AND agriculture AND (recoltă OR producție OR cereală OR ***grâu OR porumb OR cartof OR legume)

SI	suša AND kmetijstvo AND (pridelek OR pridelek OR žita OR koruza OR krompir OR zelenjava)
SK	sucho AND poľnohospodár AND (plodin OR výnos OR obilnin OR kukuric OR zemiak OR zelenin)
SV	(torka OR torr) AND jordbruk AND (gröda OR skörd OR spannmål OR majs OR potatis OR grönsak)

Table A2: Full list of search query translations for the Agriculture - permanent crops system

Language	Translated search query
BG	(засушаване OR суша) AND (плод OR овощна градина OR грозде OR вино)
CZ	sucho AND (ovoc OR sad OR hrozny OR vin OR zelenin)
DA	tørke AND (frugt OR plantage OR drue OR vin)
DE	(Trocken OR Dürre) AND (Obst OR Früchte OR Wein)
EN	drought AND (fruit OR orchard OR grape OR wine)
EL	ξηρασία AND (φρούτα OR οπωρώνα OR σταφύλι OR κρασί)
ES	sequía AND agricultur OR regad OR riego OR secan AND (uva OR viñedo OR oliv OR frutal OR hortic)
ET	(põud OR põua) AND (puuvili OR puuviljaaed OR viinamari OR vein)
FI	(kuivuus OR kuivakausi) AND (hedelm OR hedelmätarha OR omena OR viinirypäle OR viini)
FR	sécheresse AND (fruit OR verger OR vigne OR vin OR raisin)
GA	triomach AND (toradh OR torthaí OR úllord OR fíonchaor OR fíon)
HR	suša AND (voće OR voćnjak OR grožđe OR vino OR maslina)
HU	aszály AND (gyümölcs OR gyümölcsös OR szőlő OR bor)
IT	siccità AND (frutt OR vigneto OR vite OR vign OR vin OR oliv OR olio)
LT	sausums AND (augļi OR augļu dārzs OR vīnogas OR vīns)
LV	sausr AND (vais OR sodinink OR obuol)
MT	nixfa AND (frott OR ġardina OR għenba OR inbid)
NL	(droogte OR watertekort OR watergebrek OR captatieverbod) AND (fruit OR teelt OR boomgaard OR druif OR druiven OR wijn)
PL	susza AND (drzewa owocowe OR krzewy owocowe OR sad OR plantacja)
pt-PT	seca AND (fruta OR pomares OR uva OR vinha)
RO	secetă AND (fruct OR livadă OR strugure OR vin)

https://doi.org/10.5194/egusphere-2026-3041
Preprint. Discussion started: 22 July 2026

SI	suša AND (sadje OR sadovnjak OR grozdje OR vino)
SK	sucho AND (ovoc OR sad OR hrozno OR víno)
SV	(torka OR torr) AND (frukt OR odling OR druv OR vin)

Table A3: Full list of search query translations for the Agriculture - livestock system

Language	Translated search query
BG	(засушаване OR суша) AND (фураж OR говеда OR добитък OR животновъдство OR стадо OR фураж OR животно)
CZ	sucho AND (píce OR dobytek OR stád OR krmiv)
DA	tørke AND (foder OR kvæg OR flok OR fodr OR dyr)
DE	(Trocken OR Dürre) AND (Futter OR Vieh OR Herde OR Bestand OR Tier)
EN	drought AND (fodder OR cattle OR livestock)
EL	ξηρασία AND (ζωοτροφές OR βοοειδή OR αγέλη OR ζωοτροφές OR ζώα)
ES	sequía AND (pasto OR pienso OR ganad OR ovin OR bovin OR caprin OR porcin OR avicol OR cabaña ganadera)
ET	(põud OR põua) AND (sööt OR sööda OR kari OR karja OR loom)
FI	(kuivuus OR kuivakausi) AND (rehu OR karja OR lauma OR rehuvilja OR kotieläin)
FR	sécheresse AND (fourrage OR bétail OR troupeau OR aliment OR animal)
GA	triomach AND (fodar OR farae OR eallach OR caora OR caoirigh OR tread OR beatha OR ainmhí)
HR	suša AND (krma OR stoka OR stado OR opskrba OR životinja)
HU	aszály AND (takarmány OR szarvasmarha OR állatállomány OR takarmány OR állat)
IT	siccità AND (foragg OR pascol OR mangim OR bestiame OR allevament)
LT	sausums AND (lopbarība OR liellopi OR mājlopi OR ganāmpulks OR barība OR ganīšanās OR ganības OR dzīvnieki)
LV	sausr AND (pašar OR galvij OR ganym OR girdy OR gyvul OR ganykl)
MT	nixfa AND (foraġġ OR baqar OR merħla OR għalf OR annimal)
NL	(droogte OR watertekort OR watergebrek) AND (voeder OR voer OR vee OR stapel OR hoevedier)
PL	susza AND (pasza OR karma OR bydło OR pogłowie OR hodowla)
pt-PT	seca AND (forragem OR gado OR manada OR alimento OR animais)
RO	secetă AND (furaj OR fân OR vite OR bovine OR cireadă OR turmă OR șeptel OR efectiv OR hrană OR animal)

https://doi.org/10.5194/egusphere-2026-3041
Preprint. Discussion started: 22 July 2026

SI	suša AND (krma OR govedo OR čreda OR krma OR žival)
SK	sucho AND (krmiv OR dobytok OR stádo OR kŕmi OR zviera)
SV	(torka OR torr) AND (foder OR kreatur OR boskap OR bete OR djur)

351

352 **Table A4: Full list of search query translations for the Ecosystems - aquatic system**

Language	Translated search query
BG	(засушаване OR суша) AND риба AND (умира ORзагива OR кислород OR водорасли OR вода OR влажна зона)
CZ	sucho AND ryb AND (úhyn OR kysl OR řas OR vod)
DA	tørke AND fisk AND (dø OR ilt OR alg OR water OR mose)
DE	(Trocken OR Dürre) AND Fisch AND (sterben OR Sauerstoff OR Alge OR Wasser OR Sumpf OR Feuchgebiet)
EN	drought AND fish AND (die OR oxygen OR algae OR water OR wetland)
EL	ξηρασία AND ψάρια AND (θάνατος OR οξυγόνο OR άλγη OR νερό OR υγρότοπος)
ES	sequía AND peces OR pisci OR icto AND (mortandad OR oxígeno OR alga OR agua OR humedal)
ET	(põud OR põua) AND kala AND (surm OR hapnik OR vetikad OR vesi OR vee OR märgala)
FI	(kuivuus OR kuivakausi) AND kala AND (kuol OR happi OR levä OR vesi OR kosteikko)
FR	(sécheresse OR étiage) AND poissons AND (mort OR oxygèn OR eau douce OR zone humide)
GA	triomach AND iasc AND (bás a fháil OR dul in éag OR ocsaigin OR algaí OR uisce OR bogach)
HR	suša AND riba AND (ugibanje OR kisik OR alge OR voda OR močvar)
HU	aszály AND hal AND (elpusztul OR oxigén OR alga OR víz OR vizes élőhely)
IT	siccità AND pesc AND (mort OR moria OR ossigeno OR alghe OR acqu OR eutrofizzazione OR anossia OR zon umid OR palud OR lagun OR acquitrin)
LT	sausums AND zivis AND (mirst OR skābeklis OR aļģes OR ūdens OR mitrājs)
LV	sausr AND žuv AND (miršta OR deguoni OR dumbli OR vand OR pelk OR ežer OR upė)
MT	nixfa AND ħut AND (imut OR ossiġnu OR alka OR ilma OR art mistagħdra)
NL	(droogte OR watertekort OR watergebrek) AND vis AND (sterven OR sterfte OR zuurstof OR alg OR water OR drasland OR beemd OR broek OR natte weilanden OR natte grasslanden)
PL	susza AND ryby AND (śnięcie OR natlenienie OR zakwity alg OR mokradła)
pt-PT	seca AND peixe AND (mortos OR oxigénio OR algas OR água OR zona húmida)

RO	secetă AND pește AND (moare OR oxygen OR alge OR apă OR zonă umedă)
SI	suša AND ribe AND (poginejo OR kisik OR alge OR voda OR mokrišče)
SK	sucho AND ryb AND (hynúť OR kyslík OR rias OR vod OR mokraď)
SV	(torka OR torr) AND fisk AND (dö OR syre OR alg OR vatten OR våtmark)

Table A5: Full list of search query translations for the Ecosystems - terrestrial system

Language	Translated search query
BG	(засушаване OR суша) AND гора AND (дърво OR изсъхнал OR листа OR стрес)
CZ	sucho AND les AND (strom OR úhyn OR list OR stres)
DA	tørke AND skov AND (træ OR dø OR blad OR stress)
DE	(Trocken OR Dürre) AND Wald AND (Baum OR sterben OR Laub OR Stre)
EN	drought AND forest AND (tree OR die OR leaf OR stress)
EL	ξηρασία AND δάσος AND (δέντρο OR πεθάνει OR φύλλο OR άγχος)
ES	sequía AND bosqu OR forestal AND (arbol OR mortandad OR incendio OR estrés)
ET	(põud OR põua) AND mets AND (puu OR surm OR leht OR stress)
FI	(kuivuus OR kuivakausi) AND mets AND (puu OR kuol OR lehti OR stressi)
FR	sécheresse AND (forêt OR forest) AND (arbre OR mort OR feuill OR stress)
GA	triomach AND foraois AND (crann OR duilleog bás a fháil OR dul in éag OR strus)
HR	suša AND šuma AND (drvo OR odumiranje OR list OR stres)
HU	aszály AND erdő AND (fa OR elhal OR levél OR stressz)
IT	siccità AND forest OR bosc AND (alber OR mort OR moria OR fogl OR stress)
LT	sausums AND mežs AND (koks OR mirst OR lapa OR stress)
LV	sausr AND mišk AND (medž OR džiūst OR lap OR stres)
MT	nixfa AND foresta AND (siġra OR imut OR werqa OR stress)
NL	(droogte OR watertekort OR watergebrek) AND (bos or woud) AND (boom OR bomen OR sterf OR sterven OR blad OR stress)
PL	susza AND (lasy OR ekosystemy leśne) AND (drzewa OR obumieranie OR usychanie OR liście OR choroby OR szkodniki)

pt-PT	seca AND Floresta AND (árvore OR mortas OR folha OR stress)
RO	secetă AND pădure AND (pom OR moare OR frunză OR stress)
SI	suša AND gozd AND (drevo OR umre OR list OR stres)
SK	sucho AND les AND (strom OR uhyn OR list Or stres)
SV	(torka OR torr) AND skog AND (träd OR dö OR blad or stress)

Table A6: Full list of search query translations for the Energy - hydropower system

Language	Translated search query
BG	(засушаване OR суша) AND енергия AND (вода OR хидро OR язовир OR резервоар OR река OR електричество OR електроенергия)
CZ	sucho AND energi AND (vod OR hydro OR přehrad OR řek OR elekt)
DA	tørke AND energ AND (water OR hydro OR reserv OR flod OR elektr)
DE	(Trocken OR Dürre) AND Energie AND (Wasserkraft OR Wasser OR Reservoir OR Stausee OR Talsperre OR Elektro)
EN	drought AND energ AND (water OR hydro OR reserv OR river OR electr)
EL	ξηρασία AND ενέργεια AND (νερό OR υδροηλεκτρικό OR ταμιευτήρας OR ποτάμι OR ηλεκτρική ενέργεια)
ES	sequía AND energ AND (water OR hidro OR embals OR río OR presa)
ET	(põud OR põua) AND energ AND (vesi OR vee OR hüdro OR veehoidla OR jõgi OR jõe OR elektr)
FI	(kuivuus OR kuivakausi) AND energia AND (vesi OR vesivoima OR tekojärvi OR joki OR sähkö)
FR	(sécheresse OR étiage) AND energ AND (eau OR hydro OR réserv OR rivière OR fleuv OR fluv OR electr)
GA	triomach AND fuinneamh AND (hidrea OR uisce OR taiscumar OR abhainn OR leictreachas)
HR	suša AND energija AND (voda OR hidro OR akumul OR rijeka OR električna)
HU	aszály AND energia AND (víz OR víztározó OR folyó OR elektromos)
IT	siccità AND energy AND (acqu OR idro OR bacin OR invas OR fium OR fluvial OR elettric)
LT	sausums AND enerģija AND (ūdens OR hidroenerģija OR ūdenskrātuve OR upe OR elektrība)
LV	sausr AND energeti AND (vand OR hidro OR tvenkin OR upė OR elektra)
MT	nixfa AND enerġija AND (ilma OR idro OR riżerva OR xmara OR elettriku)
NL	(droogte OR watertekort OR watergebrek) AND energie AND (water OR waterkracht OR hydro OR reservoir OR bekken OR rivier OR elektris)

PL	susza AND energia AND (wodna OR zbiorniki wodne OR elektryczna)
pt-PT	seca AND energia AND (água OR hidro OR reserve OR electric)
RO	secetă AND energ AND (apă OR hidro OR rezerv OR bazin OR acumulare OR râu OR electr)
SI	suša AND energija AND (voda OR hidro OR rezervoar OR reka OR elektrika)
SK	sucho AND energ AND (vod OR hydro OR rezerv OR rie OR elektri)
SV	(torka OR torr) AND energ AND (vatten OR OR reservoar OR vattendrag OR el)

357

358 **Table A7: Full list of search query translations for the Energy - thermal system**

Language	Translated search query
BG	(Засушаване OR суша) AND енергия AND (енергиен OR атомна OR инсталация, централа OR охлаждане)
CZ	sucho AND energi AND (nuklear OR jadern OR chlazen)
DA	tørke AND energ AND (kern OR kraftværk OR kold)
DE	(Trocken OR Dürre) AND energ AND (Nuklear OR Kraftwerk OR Kühl)
EN	drought AND energ AND (nuclear OR plant OR cool)
EL	ξηρασία AND ενέργεια AND (πυρηνικό OR φυτό OR ψυχρό)
ES	sequía AND energ AND (nuclear OR planta OR enfria)
ET	(põud OR põua) AND energy AND (tuuma OR elektrijaam OR jahutus)
FI	(kuivuus OR kuivakausi) AND energia AND (ydin OR voimala OR jäähdy)
FR	(sécheresse OR étiage) AND energ AND (nucléair OR centrale OR refroidi)
GA	triomach AND fuinneamh AND (núicléach OR stáisiún OR gléasra OR fuaraigh)
HR	suša AND energija AND (nuklearna OR elektrana OR hlađenje)
HU	aszály AND energia AND (nukleáris OR erőmű OR hűtés)
IT	siccità AND energy AND (nuclear OR impiant OR central OR raffreddament OR termo)
LT	sausums AND enerģija AND (kodolenerģija OR atomelektrostacija OR kodolspēkstacija OR dzesēšana)
LV	sausr AND energeti AND (branduolin OR šilumin OR aušinim)
MT	nixfa AND enerġija AND (nukleari OR impjant OR frisk)
NL	(droogte OR watertekort OR watergebrek) AND energie AND (nucleair OR kern OR centrale OR koel)

https://doi.org/10.5194/egusphere-2026-3041
Preprint. Discussion started: 22 July 2026

PL	susza AND energia AND (elektrownia OR atomowa OR elektrociepłownia OR chłodzenie)
pt-PT	seca AND energia AND (nuclear OR central OR carvão)
RO	secetă AND ernerg AND (nuclear OR central OR răcire)
SI	suša AND energija AND (jedrska OR elektrarna OR hlajenje)
SK	sucho AND energ AND (jadrov OR zariadeni OR chladi)
SV	(torka OR torr) AND energ AND (kärnkraft OR kraftverk OR kyla)

Table A8: Full list of search query translations for the Inland navigation system

Language	Translated search query
BG	(засушаване OR суша) AND (транспорт OR река OR речен OR навигация OR кораб OR карго OR ниво)
CZ	sucho AND doprav AND (řek OR splavn OR lod OR náklad OR hladin)
DA	tørke AND transport AND (flod OR naviga OR skib OR last OR niveau)
DE	(Trocken OR Dürre OR Niedrigwasser) AND transport AND (Fluss OR naviga Or Schiff OR Fracht OR wasserstand)
EN	drought AND transport AND (river OR naviga OR ship OR cargo OR level)
EL	ξηρασία AND μεταφορά AND (ποτάμι OR ναυσιπλοΐα OR πλοίο OR φορτίο OR επίπεδο)
ES	sequía AND transport AND (naveg OR recreac)
ET	(põud OR põua) AND transport AND (jõgi OR jõe OR naviga OR laev OR last OR tase)
FI	(kuivuus OR kuivakausi) AND kuljet AND (joki OR navigointi OR laiva OR rahti OR korkeus)
FR	(sécheresse OR étiage) AND transport AND (rivière OR fleuv OR fluv OR navi OR péniche OR niveau)
GA	triomach AND iompar AND (abhainn OR loingseoireacht OR long OR lastas OR leibhéal)
HR	suša AND transport AND (rijeka OR plovi OR brod OR teret OR vodostaj)
HU	aszály AND szállítás AND (folyó OR hajózás OR rakomány OR vízszint)
IT	siccità AND transport AND (fium OR fluvial OR navig OR nav OR chiatt OR livell)
LT	sausums AND transports AND (upe OR navigācija OR kuģis OR krava OR ūdens līmenis)
LV	sausr AND transport AND (upe OR navigacija OR laiv OR krovin OR lygis)
MT	nixfa AND trasport AND (xmara OR navigazzjoni OR vapur OR merkanzija OR livell)
NL	(droogte OR watertekort OR watergebrek) AND (transport OR vervoer) AND (vaart OR kanaal OR rivier OR naviga OR schip OR schepen OR cargo OR peil OR niveau OR hoogte)

PL	susza AND (transport wodny OR żegluga) AND (rzeka OR statek OR barka OR ładunek OR stan wody)
pt-PT	(seca AND transporte AND (rio OR navega OR navio OR eclusa OR carga OR nível)
RO	secetă AND transport AND (râu OR naviga OR barcă OR cargo OR nivel)
SI	suša AND transport AND (reka OR plovba OR ladja OR tovor OR nivo)
SK	sucho AND doprav AND (rie OR naviga OR lod OR náklad OR úrov)
SV	(torka OR torr) AND transport AND (vattendrag OR navig OR fartyg OR skepp OR last OR nivå)

Table A9: Full list of search query translations for the Public water supply system

Language	Translated search query
BG	(засушаване OR суша) AND (питейна вода OR вода OR доставена OR водоснабдяване OR забран OR недостатъчно количество OR недостиг OR липса OR източник)
CZ	sucho AND pitn AND (vod OR dodávk OR zákaz OR nedostat OR zdroj)
DA	tørke AND drik AND (water OR suppl OR forbud OR mangel OR mangl OR kilde)
DE	(Trocken OR Dürre) AND Trinkwasser AND (Wasser OR Versorgung OR Verbot OR Knappheit OR Mangel OR Quelle)
EN	drought AND drink AND (water OR suppl OR ban OR shortage OR lack OR source)
EL	ξηρασία AND ποτό AND (νερό OR παροχή OR απαγόρευση OR έλλειψη OR έλλειψη OR πηγή)
ES	sequía AND potabl OR abastec AND (agua OR restrin OR restricc OR cort OR falta OR fuente)
ET	(põud OR põua) AND (jook OR joogi) AND (veevaru OR keeld OR puudus OR allika)
FI	(kuivuus OR kuivakausi) AND juoma AND (vesi OR jakelu OR kielto OR vaje OR puute OR lähde)
FR	(sécheresse OR étiage) AND potable AND (eau OR approvis OR restrict OR pénurie OR manque OR source)
GA	triomach AND (uisce OR soláthar OR cosc OR ganntanas OR easpa OR foinse)
HR	suša AND piće AND (voda OR opskrba OR zabrana OR manjak OR nedostatak OR izvor)
HU	aszály AND inni AND (víz OR ellátás OR tilalom OR hiány OR szűkös OR forrás)
IT	siccità AND potabil AND (acqu OR fornitur OR diviet OR razionament OR limitazion OR mancanz OR carenz OR approvvigionament)
LT	sausums AND dzert AND (ūdens OR apgāde OR piegāde OR ierobežojums OR trūkums OR deficīts OR trūkums OR avots)
LV	sausr AND geriamas AND (vanduo OR tiekim OR draudim OR trūkum OR vandentiek)

https://doi.org/10.5194/egusphere-2026-3041
Preprint. Discussion started: 22 July 2026

MT	nixfa AND ixrob AND (ilma OR provvista OR projbizzjoni OR nuqqas OR nuqqas OR sors)
NL	(droogte OR watertekort OR watergebrek) AND drink AND (water OR voorzien OR ban OR dichtdraaien OR dichgedraaid OR stop OR tekort OR bron OR verbod)
PL	susza AND woda pitna AND (woda OR zasoby wodne OR zakaz OR niedobór OR niedostatek OR brak wody OR źródło wody)
pt-PT	seca AND beber AND (água potável OR abastec OR racionamento OR falta OR fonte OR nascente OR poço, armazenamento)
RO	secetă AND potabil AND (apă OR supl OR interzis OR deficit OR lipsă OR sursă)
SI	suša AND pijača AND (voda OR dobava OR prepoved OR pomanjkanje OR pomanjkanje OR vir)
SK	sucho AND piť AND (vod OR zásob OR dodáv OR zákaz OR nedostat OR zdroj)
SV	(torka OR torr) AND drick AND (vatten OR försörj OR restriktion OR begräns OR förbud OR brist OR resurs)

Code and data availability

All codes and data used for the statistical analysis can be made available upon individual request.

Supplement link

The link to the supplement will be included by Copernicus, if applicable.

Author contributions

Conceptualisation: MH, LR, VB, KS, KSZ, MD'A, MT; data curation: GI, KSZ, LL, LR, MD'A; formal analysis: MH, KSZ; visualisation: MH; coordination and management: MH, VB, LR, KS; writing (original draft preparation): MH; writing (review and editing): GI, MH, LL, LR, KSZ, KS, VB, MT

Competing interests

The authors declare that they have no conflict of interest.

Disclaimer

Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements

EDORA project (No 09200200.A092005/2021/862347/ENV.C.1—Lot 1) funded by the European Commission Directorate General Environment (DG ENV). MH was supported by from AdAgriF—Advanced methods of greenhouse gases emission reduction and sequestration in agriculture and forest landscape for climate change mitigation (CZ.02.01.01/00/22_008/0004635).

We gratefully acknowledge our international colleagues who supported the multilingual component of this study by reviewing and refining search queries and assisting with the evaluation of selected search results in their respective native languages. Their expertise was essential for ensuring linguistic accuracy and contextual interpretation across countries.

The authors acknowledge the use of AI-based language editing tools during manuscript preparation. All scientific content, analysis, interpretation and conclusions remain the sole responsibility of the authors. Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 license

Financial support

MH was supported by from AdAgriF—Advanced methods of greenhouse gases emission reduction and sequestration in agriculture and forest landscape for climate change mitigation (CZ.02.01.01/00/22_008/0004635).

References

Amazon Web Services: Amazon Translate, n.d.

VigiEau: https://vigieau.gouv.fr/, last access: 22 January 2026.

Barbaresi, A.: Trafilatura, 2024.

Bartošová, L., Fischer, M., Balek, J., Bláhová, M., Kudláčková, L., Chuchma, F., Hlavinka, P., Možný, M., Zahradníček, P., Wall, N., Hayes, M., Hain, C., Anderson, M., Wagner, W., Žalud, Z., and Trnka, M.: Validity and reliability of drought reporters in estimating soil water content and drought impacts in central Europe, Agric. For. Meteorol., 315, 108808, https://doi.org/10.1016/j.agrformet.2022.108808, 2022.

BBC, S.: Antonio regretting changing his game, , 23rd February, 2023.

Blauhut, V., Stahl, K., Stagge, J. H., Tallaksen, L. M., De Stefano, L., and Vogt, J.: Estimating drought risk across Europe from reported drought impacts, drought indices, and vulnerability factors, Hydrol. Earth Syst. Sci., 20, 2779–2800, https://doi.org/10.5194/hess-20-2779-2016, 2016.

Blauhut, V., Stoelzle, M., Ahopelto, L., Brunner, M. I., Teutschbein, C., Wendt, D. E., Akstinas, V., Bakke, S. J., Barker, L. J., Bartošová, L., Briede, A., Cammalleri, C., Kalin, K. C., De Stefano, L., Fendeková, M., Finger, D. C., Huysmans, M., Ivanov, M., Jaagus, J., Jakubínský, J., Krakovska, S., Laaha, G., Lakatos, M., Manevski, K., Neumann Andersen, M., Nikolova, N., Osuch, M., Van Oel, P., Radeva, K., Romanowicz, R. J., Toth, E., Trnka, M., Urošev, M., Urquijo Reguera, J., Sauquet, E., Stevkov, A., Tallaksen, L. M., Trofimova, I., Van Loon, A. F., Van Vliet, M. T. H., Vidal, J.-P., Wanders, N., Werner, M., Willems, P., and Živković, N.: Lessons from the 2018–2019 European droughts: a collective need for unifying drought risk management, Nat. Hazards Earth Syst. Sci., 22, 2201–2217, https://doi.org/10.5194/nhess-22-2201-2022, 2022.

Chapekis, A., Bestvater, S., Remy, E., and Rivero, G.: When Online Content Disappears, 2024.

De Bruijn, J. A., De Moel, H., Jongman, B., De Ruiter, M. C., Wagemaker, J., and Aerts, J. C. J. H.: A global database of historic and real-time flood events based on social media, Sci. Data, 6, 311, https://doi.org/10.1038/s41597-019-0326-9, 2019.

European Commission. Joint Research Centre.: World drought atlas., Publications Office, LU, 2024.

Hagenlocher, M., Naumann, G., Meza, I., Blauhut, V., Cotti, D., Döll, P., Ehlert, K., Gaupp, F., Van Loon, A. F., Marengo, J. A., Rossi, L., Sabino Siemons, A. S., Siebert, S., Tsehayu, A. T., Toreti, A., Tsegai, D., Vera, C., Vogt, J., and Wens, M.: Tackling Growing Drought Risks—The Need for a Systemic Perspective, Earths Future, 11, e2023EF003857, https://doi.org/10.1029/2023EF003857, 2023.

Hansen, K. A. and Paul, N.: Newspaper archives reveal major gaps in digital age, Newsp. Res. J., 36, 290–298, https://doi.org/10.1177/0739532915600745, 2015.

Hlavsová, M., Poděbradská, M., Fischer, M., Žalud, Z., and Trnka, M.: Global drought impacts in near-real-time: a media-based database for drought impact reporting, Environ. Res. Lett., 20, 084027, https://doi.org/10.1088/1748-9326/ade819, 2025.

Houston, J. B., Hawthorne, J., Perreault, M. F., Park, E. H., Goldstein Hode, M., Halliwell, M. R., Turner McGowen, S. E., Davis, R., Vaid, S., McElderry, J. A., and Griffith, S. A.: Social media and disasters: a functional framework for social media use in disaster planning, response, and research, Disasters, 39, 1–22, https://doi.org/10.1111/disa.12092, 2015.

Jakubínský, J., Bláhová, M., Bartošová, L., Steinerová, K., Balek, J., Dížková, P., Semerádová, D., Alexandru, D., Bardarska, G., Bokal, S., Borojević, G., Bucur, A., Kalin, K. C., Barbu, A. C., Debre, B., Đorđević, M., Đurić, I., Mircea, B. F., Gatarić, S., Gregorič, G., Hasenauer, S., Ivanov, M., Kircsi, A., Labudová, L., Turňa, M., Marinović, I., Marković, M., Mateescu, E., Oblišar, G., Popescu, A., Srđević, Z., Savić-Šljivić, T., Supić, D., Sušnik, A., Pazin, N., Drljevic, M., Kuc, T., Mitrovic, L., Micev, S., Wagner, W., Eitzinger, J., Daneu, V., Blauhut, V., Stahl, K., and Trnka, M.: Repository of Drought Event Impacts Across the Danube Catchment Countries Between 1981 and 2016 Using Publicly Available Sources, Acta Univ. Agric. Silvic. Mendel. Brun., 67, 925–938, https://doi.org/10.11118/actaun201967040925, 2019.

Jobbová, E., Crampsie, A., Murphy, C., Ludlow, F., McLeman, R., Horvath, C., Seifert, N., Myslinski, T., and Sente, L.: The Irish drought impacts database: A 287-year database of drought impacts derived from newspaper archives, Geosci. Data J., gdj3.272, https://doi.org/10.1002/gdj3.272, 2024.

https://doi.org/10.5194/egusphere-2026-3041
Preprint. Discussion started: 22 July 2026

Kong, I. and Purves, R. S.: Analyzing Geographic Bias of Newspaper Articles Reporting Global Climate Disasters, Ann. Am. Assoc. Geogr., 116, 270–288, https://doi.org/10.1080/24694452.2025.2564220, 2026.

Madruga De Brito, M., Kuhlicke, C., and Marx, A.: Near-real-time drought impact assessment: a text mining approach on the 2018/19 drought in Germany, Environ. Res. Lett., 15, 1040a9, https://doi.org/10.1088/1748-9326/aba4ca, 2020.

Drought Impact Reporter: https://droughtreporter.unl.edu/.

Pecho, J., Turňa, M., Faško, P., Labudová, L., Ivaňáková, G., Kajaba, P., Markovič, L., and Bochníček, O.: Sucho na Slovensku v roku 2018, Aktual. SHMÚ, 2018.

Shyrokaya, A., Pappenberger, F., Messori, G., Pechlivanidis, I., Cloke, H., and Di Baldassarre, G.: How good is my drought index? Evaluating predictability and ability to estimate impacts across Europe, Environ. Res. Lett., 20, 034051, https://doi.org/10.1088/1748-9326/adb869, 2025.

Sodoge, J., Kuhlicke, C., and De Brito, M. M.: Automatized spatio-temporal detection of drought impacts from newspaper articles using natural language processing and machine learning, Weather Clim. Extrem., 41, 100574, https://doi.org/10.1016/j.wace.2023.100574, 2023.

Stahl, K., Kohn, I., Blauhut, V., Urquijo, J., De Stefano, L., Acácio, V., Dias, S., Stagge, J. H., Tallaksen, L. M., Kampragou, E., Van Loon, A. F., Barker, L. J., Melsen, L. A., Bifulco, C., Musolino, D., de Carli, A., Massarutto, A., Assimacopoulos, D., and Van Lanen, H. A. J.: Impacts of European drought events: insights from an international database of text-based reports, Nat. Hazards Earth Syst. Sci., 16, 801–819, https://doi.org/10.5194/nhess-16-801-2016, 2016.

Stahl, K., Szillat, K., Blauhut, V., Hlavsova, M., Rossi, L., Masante, D., and Toreti, A.: Towards an operational European Drought Impacts Database (EDID), https://doi.org/10.5194/egusphere-2025-4806, 3 November 2025.

Stephan, R., Erfurt, M., Terzi, S., Žun, M., Kristan, B., Haslinger, K., and Stahl, K.: An inventory of Alpine drought impact reports to explore past droughts in a mountain region, Nat. Hazards Earth Syst. Sci., 21, 2485–2501, https://doi.org/10.5194/nhess-21-2485-2021, 2021.

Sufi, F. and Alsulami, M.: Unmasking Media Bias, Economic Resilience, and the Hidden Patterns of Global Catastrophes, Sustainability, 17, 3951, https://doi.org/10.3390/su17093951, 2025.

Sutanto, S. J., Van Der Weert, M., Wanders, N., Blauhut, V., and Van Lanen, H. A. J.: Moving from drought hazard to impact forecasts, Nat. Commun., 10, 4945, https://doi.org/10.1038/s41467-019-12840-z, 2019.

Trnka, M., Hlavinka, P., Možný, M., Semerádová, D., Štěpánek, P., Balek, J., Bartošová, L., Zahradníček, P., Bláhová, M., Skalák, P., Farda, A., Hayes, M., Svoboda, M., Wagner, W., Eitzinger, J., Fischer, M., and Žalud, Z.: Czech Drought Monitor System for monitoring and forecasting agricultural drought and drought impacts, Int. J. Climatol., 40, 5941–5958, https://doi.org/10.1002/joc.6557, 2020.

United Nations Office for Disaster Risk Reduction (Ed.): Special report on drought 2021, United Nations, Geneva, 173 pp., 2021.

Wang, J., Castelletti, A., Madruga De Brito, M., and Pernici, B.: Drought perceived impacts via text mining of social media, Environ. Res. Water, 1, 045007, https://doi.org/10.1088/3033-4942/ae2e37, 2025.

Wilhite, D. A., Svoboda, M. D., and Hayes, M. J.: Understanding the complex impacts of drought: A key to enhancing drought mitigation and preparedness, Water Resour. Manag., 21, 763–774, https://doi.org/10.1007/s11269-006-9076-5, 2007.