https://doi.org/10.5194/nhess-2023-228
Preprint. Discussion started: 21 December 2023

Natural Hazards and Earth System Sciences
Discussions

Text-mining uncovers the unique dynamics of socio-economic impacts during multi-year drought

Jan Sodoge[1,2], Christian Kuhlicke[1,2], Miguel D. Mahecha[3,4,5], Mariana Madruga de Brito[1]

[1] Department of Urban and Environmental Sociology, UFZ-Helmholtz Centre for Environmental Research, 04318, Leipzig, Germany
[2] Institute of Environmental Science and Geography, University of Potsdam, 14476, Potsdam-Golm, Germany
[3] Remote Sensing Centre for Earth System Research, Leipzig University, 04103, Leipzig, Germany
[4] Department of Remote Sensing, Helmholtz Centre for Environmental Research—UFZ, Leipzig, Germany
[5] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena Leipzig, Leipzig, Germany

Correspondence to: Jan Sodoge (jan.sodoge@ufz.de)

Abstract: Droughts often lead to cross-sectoral and interconnected socio-economic impacts, affecting human well-being, ecosystems, and economic development. During extended drought periods, such as the 2018-2022 event in Germany, these impacts are amplified due to temporal carry-over effects. Yet, our understanding of drought impact dynamics during increasingly frequent multi-year drought periods is still in its infancy. In this study, we analyze the socio-economic impacts of the 2018-2022 multi-year drought in Germany and compare them to previous single-year events. Leveraging text-mining tools, we derive a dataset covering impacts reported by 260 newspapers on agriculture, forestry, livestock, waterways, aquaculture, fire, and social impacts spanning 2000 to 2022. We introduce the concept of drought impact profiles (DIPs) to describe spatio-temporal patterns of the reported co-occurrences of impacts. We employ a clustering algorithm to detect these DIPs and then use sequence mining, visualization techniques statistical tests to analyze spatio-temporal trends. Our results reveal that the 2018-2022 multi-year drought event had distinct impact patterns compared to prior single-year droughts regarding their spatial extent, impact diversity, and prevalent impact types. For the multi-year drought period, we identify shifts in how impacts have been perceived regionally, especially focusing on legacy and cascading effects on forestry and social activities. Also, we show how regional differences in relevant impacts are controlled by different land-cover types. Our findings enhance the understanding of the dynamic nature of drought impacts, highlighting the potential of text-mining techniques to study drought impact dynamics. The insights gained underscore the need for different strategies in managing multi-year droughts compared to single-year events.

1. Introduction

Droughts challenge human well-being, ecosystems, and economic development worldwide. Their impacts spread across multiple socio-economic sectors such as agriculture, livestock, and waterways navigation (Stahl et al., 2016). They can occur concomitantly (i.e., compounding) or spread from one economic sector to another (i.e., cascading) (Erian et al., 2021; de Brito, 2021; Lawrence et al., 2020; Garrick et al., 2018). For instance, drought-related harvest failures in Russia in 2010, combined with an export ban, led to a global spike in cereal prices. This shortage is assumed to have amplified the food security risk in other countries (Challinor et al., 2018, 2017). Another example is the 2018 summer in Germany, where low soil moisture values caused crop failures, leading to feeding shortages and consequent livestock reductions (de Brito, 2021).

The socio-economic impacts of droughts are not only driven by the biophysical severity of the drought itself but are also shaped by factors such as societal exposure, vulnerability, and adaptation responses (Damian et al., 2023; Blauhut et al., 2015; Lindner et al., 2010; Simpson et al., 2021). Also, impacts influence each other, forming a complex network of cascading and compounding patterns (Chen et al., 2022; de Brito, 2021; Erian et al., 2021). As a result, the socio-economic impacts of droughts are spatiotemporally dynamic and not directly proportional to the biophysical occurrence of drought hazards. This complexity becomes especially salient during multi-year droughts, which are characterized by an extended duration of low precipitation and water scarcity typically leading to regional biophysical feedbacks that exacerbate the hazardous conditions (Miralles et al., 2019). Here, the effects of vulnerability and exposure tend to build up over time during prolonged droughts (Kim et al., 2021; De Silva and Kawasaki, 2018). Consequently, the impacts of multi-year droughts are not static and consistent; rather, they evolve and change continuously.

The increasing incidence of multi-year drought periods in several regions worldwide (Rakovec et al., 2022; Moravec et al., 2021; Fischer et al., 2021) underscores the need to comprehend how these extended drought events impact society. Previous studies have shown that the duration of drought is linked to the emergence of new socio-economic impact types (Yu et al., 2018; Tijdeman et al., 2022; Chen et al., 2022). An intuitive example of the effect of drought duration is the dieback of trees in Australian and California forests due to the extended and intense droughts (Stephenson et al., 2018; Matusick et al., 2018). Therefore, research on the distinct spatio-temporal impact patterns during multi-year droughts is needed for designing and implementing robust adaptation measures (Liguori et al., 2021; Rakovec et al., 2022).

Over the past years, substantial advancements have been made in studying patterns of socio-economic drought impacts (Niggli et al., 2022; Erfurt et al., 2020; Dahlmann et al., 2022; Liguori et al., 2021; de Brito, 2021). However, a majority of these studies exhibit a limited scope, both spatially and temporally. Their concentration on isolated incidents undermines the potential for broader generalization and it often is unclear whether patterns observed during a particular drought event are representative of other periods not covered by the study. Also, very few studies consider multiple sectors impacted by droughts, and a focus on singular sectors such as agriculture or forestry prevails (Stahl et al., 2016; de Brito et al., 2020; Sutanto et al., 2019). Overall, there is a clear need for a systematic approach that incorporates the multi-sectoral effects of drought during extended periods and geographic regions.

In this paper, we study the spatio-temporal patterns of reported socio-economic drought impacts of both multi-year and single-year drought periods in Germany from 2000 to 2022. Germany is selected as a case study because of its recent history of significant droughts (2003, 2015, 2018-2022) with widespread impacts on agriculture, forestry, livestock, and waterways navigation, among others (Peña-Angulo et al., 2022; Rakovec et al., 2022; Tijdeman et al., 2022; de Brito et al., 2020). The assessment of reported impacts supports a focal point on their human perception. Specifically, we focus on the multi-year drought period between 2018-2022, which is considered a new benchmark in terms of duration and intensity (Rakovec et al., 2022). With this, we aim to understand (a) how single-year and multi-year drought events differ, (b) how drought impact patterns change or persist over the years, and (c) how land-cover are related to these impact patterns.

2. Methods

In this study, we used newspaper texts to create a German-wide drought impact dataset covering multiple sectors. We introduced the concept of a drought impact profile (DIP) to construct a typology summarizing co-occurring drought impact types at a certain time and region. Based on the developed DIPs, we investigated patterns of drought impact occurrence throughout 3 levels of analysis (see Fig. 1). First, we compared the DIPs of multiple drought events to understand how single-year and multi-year drought events differ. Second, we examined how the DIPs change or persist within each district using graphical and sequence mining methods. Third, we used land-cover data to demonstrate how external data on exposure and vulnerability can be linked to the DIPs to understand what controls their occurrences.

2.1 Data

We developed a dataset covering 7 commonly observed drought impact types in Germany between 2000 and 2022. These include impacts on agriculture (including crop yield losses), livestock (i.e. impacts on livestock farming and animal populations), waterways (i.e. impacts on shipping and navigation), forestry (i.e. impacts on trees and forest ecosystems), aquaculture (i.e. impacts on fishing-related activities), social (i.e. impacts on places and activities used for recreation, tourism, leisure), and fire (i.e. fire in forests or other areas due to drought conditions); for a detailed description of each impact class see Table A.1.

To create this dataset, we leveraged the text-mining approach proposed by Sodoge et al. (2023) for detecting and classifying the drought impacts and their geographic location from newspaper articles. We considered ~50,000 German newspaper articles mentioning drought-related keywords published between 2000 and 2022. We first removed duplicate and non-relevant articles. Then, we classified the impact types using lasso logistic regression models that were trained and evaluated on a sample of 1,800 annotated newspaper articles. The models achieved an 89% median accuracy when compared to the manually annotated data (see Table A.2). In a final step, we estimated the impact location on the district level following the nomenclature of territorial units for statistics (NUTS-3 units).

The resulting dataset described the frequency of drought impact statements (DIS) by year and district. A DIS documents a specific type of reported impact, its estimated date of occurrence, and its location. For example, a DIS could describe the reported impact on agriculture in Berlin on 16.8.2022. We aggregated the DIS per year and district. The aggregation by year followed natural breakpoints, as shown in Fig. A.2. Most impacts were reported in summer and continuously decreased towards winter. To assess how well our DIS dataset corresponds to external data, we correlated it against multiple empirical indicators: precipitation deficit (DWD, 2023), Google trends data reflecting public awareness (Google, 2023), forest fire statistics (BZL, 2020), and agricultural yield losses (RDB, 2023). The validation results showed that the DIS and these empirical indicators were correlated, suggesting that our estimates are accurate (see Fig. A.1). Detailed descriptions of the proposed method, validation procedure, and results can be found in Sodoge et al. (2023).

https://doi.org/10.5194/nhess-2023-228
Preprint. Discussion started: 21 December 2023

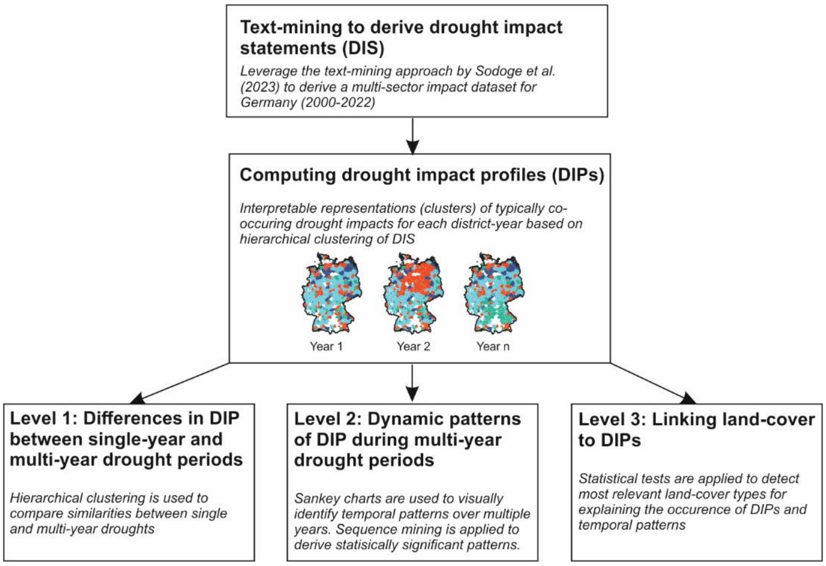


Figure 1: Overview of the methods used to compute drought impact profiles and analyze their patterns.

2.2 Computing and analyzing drought impact profiles (DIPs)

Drought and their impacts are known to be power-law distributed (Zscheischler et al., 2014; Mahecha et al., 2017). Accordingly, we find that most observations in our DIS database have few or no impacts reported, while a few observations contained the majority of reported impacts. This intrinsic imbalance hinders the construction of DIPs because they disrupt clustering by co-occurrence patterns. Hence, we used the following transformations to the DIS data to ensure that the resulting DIPs primarily reflect co-occurrence patterns rather than the severity of droughts. To this end, we only consider observations from years with severe drought impacts. These were selected based on previous studies (Peña-Angulo et al., 2022; Tijdeman et al., 2022; Rakovec et al., 2022) and the annual magnitude of the DIS. Specifically, we focused on the droughts of 2003, 2015, and the 2018-2022 multi-year drought period. We excluded 2021 from the analysis because few impacts were reported, which could skew the computation of DIPs. Furthermore, we did not treat 2022 as a single-year drought because its effects were still reminiscent of the preceding 2018-2020 drought. We then grouped the DIS data by year and district and re-scaled from 0 to 1, where 0 means the minimum DIS value within the grouping, and 1 is the maximum (see Fig A.3). This rescaling allowed us to assess variations in the relative significance of impacts across different regions (e.g. north vs. south) and years.

After transforming the data, we created the DIPs by clustering similar observations. To this end, we computed their Euclidean distances, where a small distance reflects similar impacts, whereas a larger distance indicates distinct ones. Based on these distances, we clustered the observations using an agglomerative hierarchical clustering algorithm called Ward's linkage (Sharma et al., 2019). We selected this algorithm because it is known to provide robust results when dealing with continuous data by minimizing the variance between clusters. This method initially labels each observation as an individual cluster and iteratively merges them into larger clusters based on the identified distances (Husson et al., 2010). The ideal number of clusters (k) was determined using both quantitative measures (e.g. Elbow method, silhouette

Natural Hazards and Earth System Sciences
Open Access
EGU
Discussions

coefficient) (Ketchen and Shook, 1996; Thorndike, 1953) and domain knowledge. For the latter, we considered existing information about compounding and cascading impacts in Germany (de Brito, 2021) as well as co-occurrence patterns within our DIS dataset (Fig. A..4).

2.2.1 Differences in impact patterns between single-year and multi-year drought events

To compare the impact patterns of single-year and multi-year drought events, (analysis level 1 in Fig. 1), we used a similarity measure and hierarchical clustering. We computed the similarity between the two years by counting the number of districts with identical DIP in both years. For instance, if 140 districts exhibited the same DIP in 2003 and 2015, the similarity measure would also be 140. Subsequently, we applied hierarchical clustering with Ward's linkage to visualize these pairwise similarities in a dendrogram.

To further explore the differences between single-year and multi-year droughts, we also considered the diversity of occurring DIPs. To this end, we calculated the Shannon index (H) (Spellerberg and Fedor, 2003) for each year by summing the products of the relative abundance of each category (pi) and the natural logarithm of that category's relative abundance (ln(pi)) (see Eq. 1). A higher H value suggests that there are many different types of DIPs across the analyzed districts, and these are evenly distributed. In contrast, a lower H value indicates fewer distinct types of DIPs, and some may dominate.

$$H = -\sum_i p_i * \ln(p_i) \quad (1)$$

2.2.2 Dynamic patterns of impacts during multi-year drought periods

To investigate how the DIP patterns evolved during the 2018-2022 multi-year drought period (analysis level 2 in Fig. 1), we employed two distinct yet complementary approaches: a graphical analysis using alluvial diagrams and statistical sequence mining. Both approaches aimed at identifying temporal sequences that describe DIP's characteristic shifts (or persistence). Alluvial charts served to effectively visualize sequences, presenting them in proportion to the number of affected districts. Sequence mining assisted as a quantitatively complementary approach to identify statistically significant sequences of DIPs in consecutive years. We employed the CSPADE (Sequential Pattern Discovery using Equivalence classes) algorithm, a widely used sequence mining implementation (Zaki, 2001; Wright et al., 2015) (see Fig. 2). To apply the CSPADE algorithm, we created a transactional database with the antecedent and consequent DIPs in each district during the 2018-2022 drought period. The extracted sequences were evaluated on 3 measures: support, confidence, and lift. Support corresponds to how often the particular sequence appeared within the data (see Eq. 2). Confidence measures how often the DIP occurred together relative to all observations with the antecedent (see Eq. 3). Lift measures how often antecedent and consequent DIPs were observed together relative to how often they were expected to be observed (see Eq. 4). The obtained sequences with high lift can be interpreted as the most prevalent ones.

$$Support(A) = \frac{Number\ of\ sequences\ containing\ pattern\ A}{Total\ number\ of\ sequences} \quad (2)$$

$$Confidence\ (A \rightarrow B) = \frac{Support(A+B)}{Support\ (A)} \quad (3)$$

$$Lift\ (A \rightarrow B) = \frac{Support\ (A+B)}{Support\ (A) * Support\ (B)} \quad (4)$$

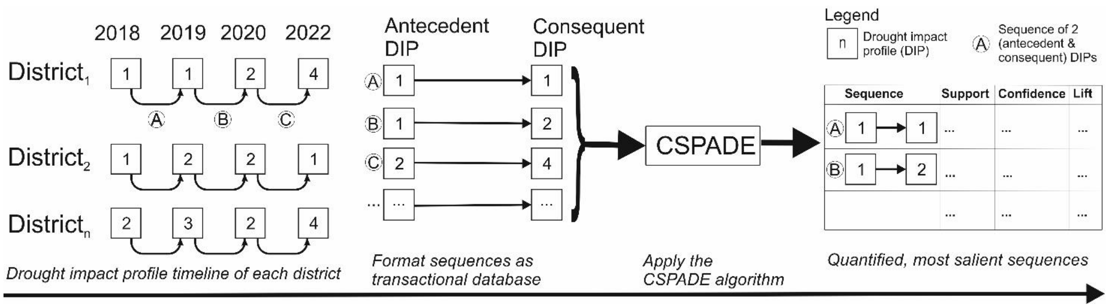


Figure 2: Process of extracting DIP sequences with CSPADE algorithm. For each district, a timeline was created representing the DIP observed during the multi-year drought period. These were split into a transactional database representing the transitions between the DIPs of two consecutive years. By using the CSPADE algorithm, we extracted the most frequent sequences and quantified them based on lift, support, and confidence metrics.

2.2.3 Linking land-cover to multi-year drought impact profiles (DIPs)

To demonstrate the analytical capabilities of the DIPs typology, we leveraged non-parametric statistical tests to search for significant associations between the DIPs and land-cover data (analysis level 3 in Fig. 1). For this analysis, we focused on the observations during the multi-year drought period as we aim to disentangle the patterns within this special period. Land-cover has been found to control the effect of drought on ecology (Flach et al., 2021) and the socio-economic impacts of drought (Sutanto et al., 2019; Blauhut et al., 2016). We considered the 10 most prevalent types of land-cover in Germany using the CORINE (Coordination of Information on the Environment, Land Cover) dataset (Büttner et al., 2004). For each district, we calculated the relative share of each land-cover type. To detect statistically significant associations between the DIPs and the share of each land-cover type, we applied a one-sided Mann-Whitney U test. This test was chosen because of its suitability in detecting significant differences between two data samples without requiring specific data distributions. It compares two samples through rank transformation and subsequent comparison of these ranks (details can be found in McKnight and Najab (2010)). The resulting p-value indicates whether there are significant differences between the two analyzed samples. For this study, we performed two types of comparisons. First, we compared the land-cover types of districts affected to those unaffected by a particular DIP. A significant p-value indicates that districts impacted by a specific DIP exhibit a greater proportion of a particular land-cover type. Second, we compared the land-cover types of districts experiencing a particular DIP sequence to those without that sequence. Here, we select the most prevalent sequences from the sequence mining application (see Section 2.2.2). Using sequences can provide insights into what factors drive regions to switch from one DIP to another.

3. Results

3.1 Socio-economic drought impact dataset

The text-mining-based drought impact dataset for Germany comprises 31,370 DIS along 7 impact types reported by newspapers between 2000 and 2022 (see Fig. 3). Notably, the period from 2018-2022 (excluding 2021) accounts for 42 % of all DIS. Throughout this period, we observe a varied and diverse distribution of the DIS across time and space. For example, northeastern Germany’s agriculture and livestock sectors were particularly affected. Conversely, impact types such as ‘social’, ‘forestry’, or ‘fire’ exhibit a more widespread occurrence. The severe drought events of 2003 and 2018 have caused the highest number of impacts across all the impact types we analyzed. However, there are variations in their temporal trends. For instance, ‘agriculture’ impacts peaked during the 2018 drought. Instead, ‘forestry’ impacts were less pronounced in 2018 and peaked in 2019 and 2020.

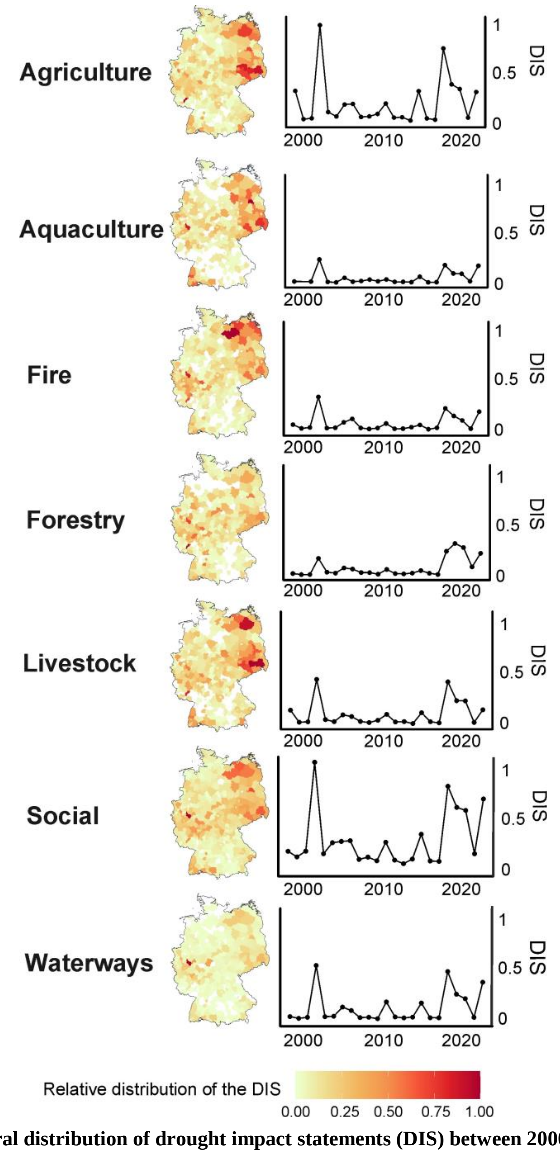
3.2 Drought impact profiles

As a result of the hierarchical clustering, we identified 4 clusters of observations with similar co-occurring impact types, referred to here as drought impact profiles (DIPs) (Fig. 4). Overall, both quantitative evaluation metrics (i.e. silhouette coefficient and dendrogram inspection) and qualitative inspection of the DIPs, confirm the distinctiveness of these 4 clusters (see Fig. A.5). With an emphasis on interpretability, the derived DIPs showcase unique characteristics which closely mirror co-occurrence patterns from correlation analysis results (see Fig. A.4). The silhouette coefficient, measuring 0.22, suggests a moderate degree of separation and discernible structure within the data. In light of the exploratory nature of this study, the moderate clustering results can be considered suitable as they uphold interpretability and align with domain knowledge.

Each DIP is enriched by characteristic impact types and has a varying spatial and temporal distribution. These are used here as a reference point for subsequent analysis. For example, DIP 1 predominantly features ‘agriculture’ and ‘livestock’ impacts and is particularly prevalent in eastern Germany. The prevalence of DIP 1 declined during the multi-year drought period. Meanwhile, it was dominant during 2018 and 2003. The second DIP is enriched by water ecosystem consequences, including ‘waterways’ and ‘aquaculture’ impacts. In 2018 and 2022, DIP 2 reached its peak when compound heat and drought events affected aquaculture and led to low flows, limiting waterway transportation on major water courses (Conradt et al., 2023; Free et al., 2023). As such, DIP 2 is prevalent in districts with major water courses, such as the Rhine River in western Germany and the Oder River on the Polish border. DIP 3, on the other hand, is composed mainly of ‘forestry’ impacts and is spread across Germany, especially in forestry ecosystems that experienced notable drought effects: the Harz region, Saxon Switzerland mountains, and Alsace (Holzwarth et al., 2020; Erfurt et al., 2020). While DIP 3 hardly occurred during the single-year drought events, we observed an increase in 2019. Lastly, DIP 4 is characterized by the interplay between ‘fire’ and ‘social’ impacts. The occurrence of forest fires, or a high likelihood of them, limits the functioning of recreational zones, such as parks and forests. While this DIP is dispersed heterogeneously across Germany, it reflects hot spots of past forest fires such as (north-)eastern Germany (Thonfeld et al., 2022). Notably, we observe an increasing occurrence of this DIP over the last 20 years.

https://doi.org/10.5194/nhess-2023-228
Preprint. Discussion started: 21 December 2023

Natural Hazards and Earth System Sciences
Open Access
EGU
Discussions



225 **Figure 3: Spatial and temporal distribution of drought impact statements (DIS) between 2000-2022. Each map displays the relative distribution for a particular impact type, where 1 corresponds to the DIS type's national maximum and 0 to the minimum. Each time series displays the magnitude of DIS, where 1 corresponds to the maximum DIS among all impact types, and 0 is the minimum.**

3.3 Comparison of the DIPs between single-year and multi-year drought events

The comparison of the DIPs across the drought events shows the distinctiveness of the multi-year drought period compared to prior single-year events (Fig. 5). The droughts of 2003 and 2015 display the highest similarities despite being more than a decade apart. Both share a high prevalence of DPI 1 (enriched in agriculture and livestock impacts), particularly in eastern Germany and many districts without any impact. The dominance of 'agriculture ' and 'livestock ' impacts can be attributed to the importance and vulnerability of the agricultural sector in (northeastern) Germany, as well as the societal significance of the resulting crop yield losses (Zink et al., 2016; Schmitt et al., 2022; Reyer et al., 2012).

https://doi.org/10.5194/nhess-2023-228
Preprint. Discussion started: 21 December 2023

Natural Hazards and Earth System Sciences
Open Access
EGU
Discussions

Both years exhibit the lowest DIP diversity scores, corroborating the hypothesis that single-year droughts tend to have more homogeneous impacts.

The number of districts being affected (thus having a DIP) sets the single-year from the multi-year drought events apart. In single-year droughts, an average of 60% of all districts in Germany were affected, whereas during the 2018-2022 drought, 92% of the districts had at least one reported impact each year. The widespread impacts of the 2018-2022 drought can be linked to the severe biophysical drought conditions and their extensive reach, which positioned the multi-year drought as an unprecedented event (Rakovec et al., 2022). For instance, while 2018 is similar to 2003 and 2015 concerning dominating DIP 1, the spatial extent of the impacts in 2018 is strongly different. Only 2020, where 21% of the districts did not report any DIS, displays higher similarity scores to the single-year events. During the multi-year drought period, the varying similarities between each year indicate some evolving differences. A striking finding is the low similarity between 2018 and 2019.

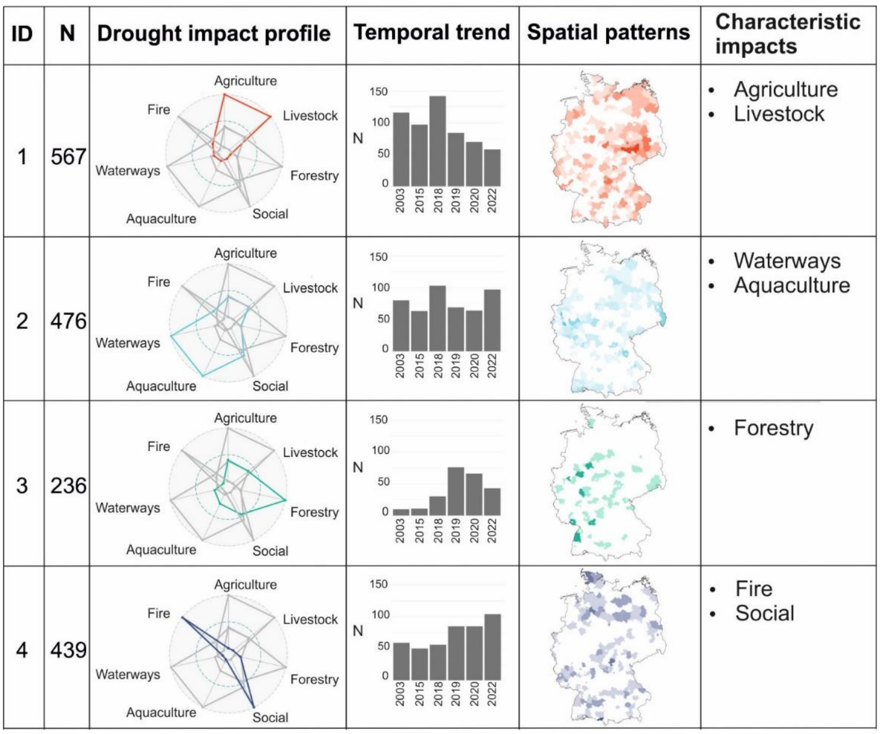


Figure 4: Overview of drought impact profiles (DIPs) derived from hierarchical clustering (Fig. A.5). Each DIP describes a characteristic combination of co-occurrences among the 7 DIS categories. For each impact type in the radar chart, the maximum and minimum correspond to the maximum/minimum of the particular impact type in the DIS dataset. For the spatial patterns, DIPs are aggregated for the analyzed years.

3.4 Dynamic patterns of impacts during multi-year drought periods

Throughout the multi-year drought period, we observe distinct patterns of how the DIPs change over time. The probability that a district remains with the identical DIP for two years is only 26 %. Yet, the DIPs do not change in random order and instead follow identifiable patterns that cause shifts in the dominating DIP. By examining the results from both the alluvial chart analysis and sequence mining, we identify 4 major trends (Fig. 6 and 7).

https://doi.org/10.5194/nhess-2023-228
Preprint. Discussion started: 21 December 2023

First, we recognize a legacy effect driving a delayed emergence of the ’forestry’ DIP from 2019 onwards. From 2018 to 2019, 53 (13%) districts shifted from the ’aquaculture/waterways’ and ’livestock/agriculture’ DIPs to the forestry DIP. Sequence mining also revealed similar sequences (DIP 1 to 3, support=0.20 in Fig. 6). This underlines the escalating significance of the forestry sector in 2019. The increased effects on the forestry sector are documented and attributed to an increased vulnerability of the trees after the 2018 drought and exacerbated by bark beetle pest causing higher tree mortality (Bastos et al., 2020; Schuldt et al., 2020; Kannenberg et al., 2020). After 2019, the prevalence of the ’forestry ’ DIP slowly declines yet remains at higher levels compared to 2018.

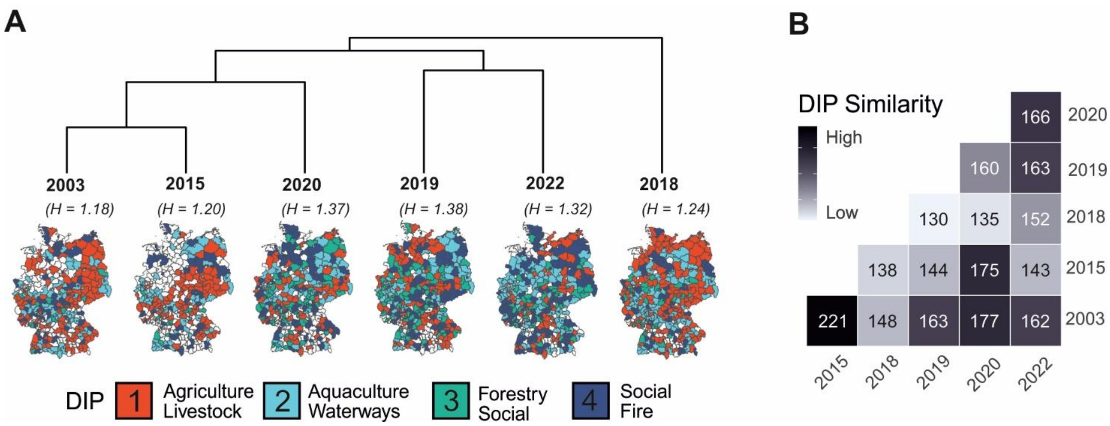


Figure 5: Comparison of annual events based on DIPs within each district. A dendrogram of hierarchical clustering where a structure of similar years emerges. H indicates the calculated diversity index. B similarity matrix with the number of identical DIPs between individual years and is used to perform the hierarchical clustering in panel A.

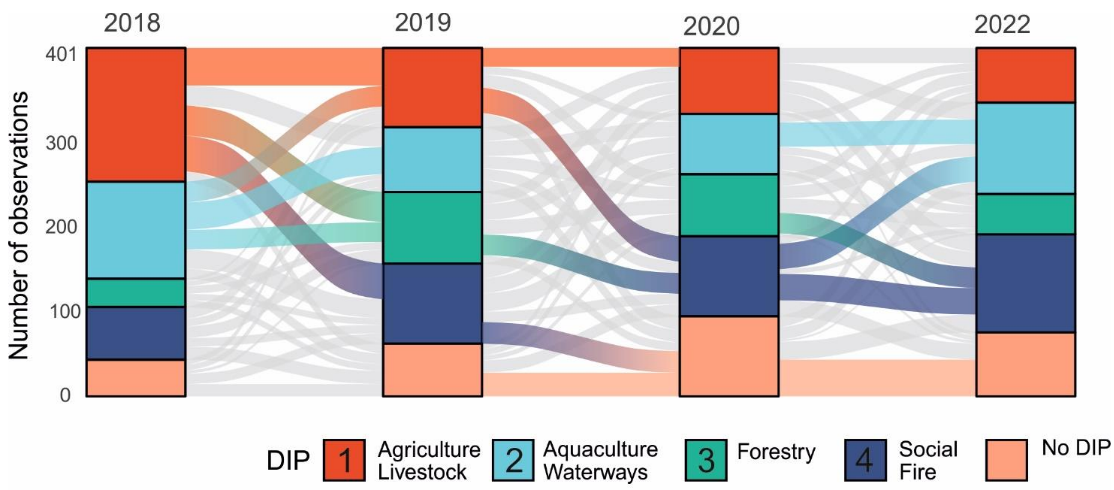


Figure 6: Transitions among DIPs present in each district during the multi-year drought period. Flows of at least 20 districts between two DIPs are highlighted.

Second, we identify an increasing prevalence of the ’social/fire’ DIP, which was present in 13% of the districts in 2018 and increased to 26% in 2022. Within this context, 65 districts affected by ’agriculture/livestock’ DIPs in 2018 and 2019 shifted to ’social/fire’ DIPs. Additionally, 44 districts associated with the ’forestry’ DIP in 2019 and 2020 shifted to the

https://doi.org/10.5194/nhess-2023-228
Preprint. Discussion started: 21 December 2023

’social/fire’ DIP in 2022. Here, we hypothesize that severe and long-lasting forest damages reported in the prior 2 years had resulted in a loss of forest function for recreation or made forests more vulnerable to fire. Then, the shift towards ’social/fire’ DIP would directly result from the multiple years of drought that have damaged forest ecosystems. Such patterns are documented in regions like the Harz mountain region (Hahne et al., 2009; Schütte and Plothe, 2022).

Third, the prior two trends are underpinned by a steadily decreasing relevance of the ’agriculture/livestock’ DIP and a more even distribution of the DIPs in the consecutive years. In 2018, 142 districts were linked to the ’agriculture/livestock’ DIP, while in 2022 only 58 were affected. This decreasing relevance results in a more even representation of the DIPs in the following years, which is visible in the measured DIP diversity (see Fig. 5). Concurrently, a more fragmented geographic distribution of the DIPs emerges. For instance, northeastern Germany is less dominated by ’agriculture/livestock’ impact.

Fourth, we found that districts affected by the ’waterways/aquaculture’ DIP exhibit a higher degree of persistence, meaning that they are less likely to transition to other DIPs. The sequence mining highlights a sequence where districts remain with the ’waterways/aquaculture’ DIP for two years (DIP 2 → DIP2, support = 0.246 in Fig. 6) . This persistence can be attributed to the importance of waterbodies for specific regions, exemplified by the vital role of waterbodies like the Rhine River. Meanwhile, a less frequent sequence was identified where districts shift from the ’waterways/aquaculture’ to the ’social/fire’ DIP. Here, we posit that districts with affected waterbodies may have closer ties to recreational impacts, thus contributing to the association between these specific DIPs (Wieland and Martinis, 2020; Erfurt-Cooper and others, 2009).

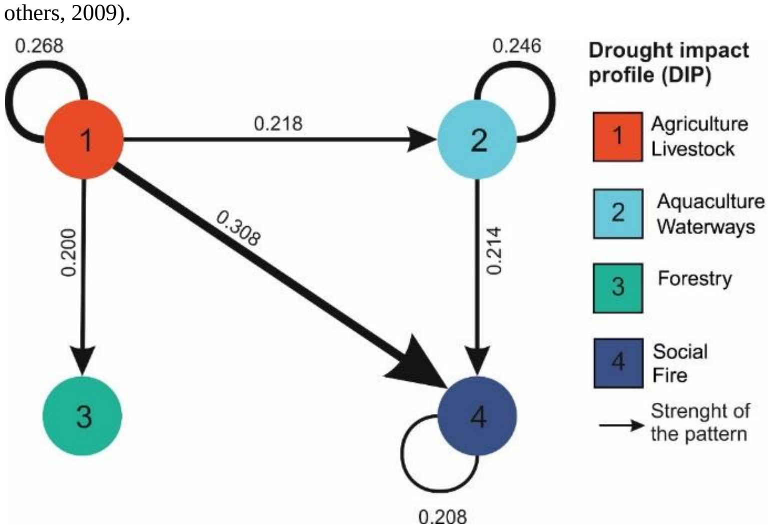


Figure 7: Sequences of DIPs during multi-year drought period discovered with CSPADE sequence mining algorithm. All sequences with a minimum support measure of 0.2 are displayed and labeled accordingly Full evaluation metrics are provided in Table A.3.

3.5 Linking land-cover and multi-year drought impact patterns

To investigate the exposure factors contributing to drought impacts, we linked the DIPs with distinct land-cover types (Fig. 8). Our analysis revealed key associations between DIP categories and land-cover types. DIP 1, representing ‘agriculture’ and ‘livestock’ impacts, is more prevalent in districts with non-irrigated, arable land than those without this

https://doi.org/10.5194/nhess-2023-228
Preprint. Discussion started: 21 December 2023

DIP (p-value 0.00). At the same time, districts with agricultural land cover are more likely to experience ’agriculture’ and ’livestock’ impacts. DIP 2 (’aquaculture/waterways’) is significantly linked to a higher presence of watercourses and water bodies (p = 0.001; 0.003, respectively). Districts impacted by the ’forestry’ DIP exhibit elevated levels of broad-leaved and mixed forest land-cover (p=0; 0.04), while those influenced by the ’social/fire’ DIP show greater proportions of mixed and coniferous forests (p=0;0.02). Here, we note a particular differentiation: coniferous forests are significantly linked to the DIP ’fire/social’ DIP, whereas broad-leaved forests with the ’forestry’ DIP. This distinction points to a higher susceptibility of coniferous forests to ’fire’ impacts, while broad-leaved forests appear to be more affected by factors such as tree mortality. This observation aligns with prior research, which highlights the heightened susceptibility of coniferous forests to fires, especially in eastern Germany (Gnilke and Sanders (2021). Other significant associations were also found. For instance, the ’forestry’ DIP is linked to the discontinuous urban fabric and commercial units’ land-cover. While an intuitive linkage cannot explain these findings, these might stem from (i) multi-collinearity among the land-cover types, (ii) unknown characteristics of affected districts or impacts, or (iii) driven by special events.

To further understand what land-cover types drive districts to shift DIPs from one to another, we identify land-cover types that match the sectors affected by the temporal sequences. For example, districts sticking to the ’forestry’ DIP within two consecutive years show significantly higher broad-leaved forest land-cover. This additional analysis adds additional depth to the characteristics of the districts. For instance, districts affected by the ’agriculture/livestock’ DIP within two consecutive years display higher shares of agricultural land-cover. Instead, districts that shift from ’agriculture/livestock’ DIP to the ’social/fire’ DIP have no significantly higher agricultural land-cover and instead higher coniferous forest land-cover. These differences indicate that districts remaining impacted by dominating ’agriculture/livestock’ impacts possess different land-cover characteristics to those shifting towards other DIPs.

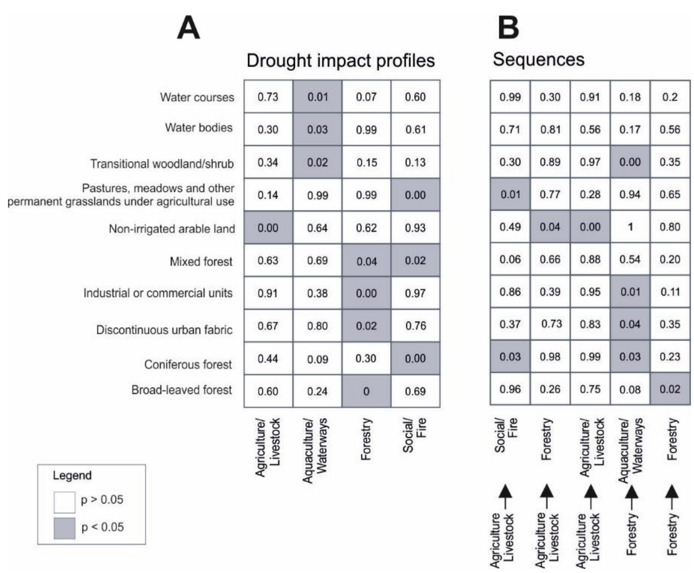


Figure 8: Testing associations between land-cover types and DIP occurrences using the one-sided Mann-Whitney U test. A significant p-value indicates that districts where a particular DIP (sequence) indicates have a higher share of respective land-cover types.

https://doi.org/10.5194/nhess-2023-228
Preprint. Discussion started: 21 December 2023

Natural Hazards and Earth System Sciences Discussions
Open Access

3. Discussion

Multi-year drought periods are becoming increasingly likely and thus require special attention for developing effective adaptation measures (Rakovec et al., 2022; van der Wiel et al., 2023). Against this background, we investigated the impact patterns during the recent multi-year drought period from 2018-2022 in Germany and compared those with patterns observed in single year droughts. Using a text-mining-based socio-economic impact dataset, our study provides insights into (1) differences between the multi-year drought and single-year drought events, (2) dynamic patterns during multi-year drought periods, and (3) linkages between land-cover and impact patterns during the multi-year drought period.

Using text-mining to obtain socio-economic drought impact data, we demonstrated how natural language processing can support the assessment of impacts. This can, in turn, empower scientists to study drought patterns over long timescales and with broad geographical coverage. Prior research on drought impact patterns has often been challenged by the lack of multi-sectoral and large-scale impact datasets and thus used smaller spatio-temporal scopes. While studies highlight the advantages of using newspaper articles for natural hazard impact, reports can miss or overemphasize impact data (Noone et al., 2017; Engelmann, 2010; Llasat et al., 2009; de Brito et al., 2020). However, our empirical validations highlighted that impacts' spatial and temporal distribution correlated with external indicators (Sodoge et al., 2023). Still, for future research, it is necessary to improve the evaluation of reporting biases to enhance the accuracy of resultant impact data and enable more insightful interpretations.

For examining multi-sectoral and spatio-temporal drought impact patterns, this study illustrates the effectiveness of combining multiple pattern mining methods for both visual and statistical examination. By using clustering algorithms to create a typology of co-occurring impact types that match patterns of cascading and compounding impacts in Germany, we advanced the representation of multi-sectoral impact patterns. Prior work has used dyadic conceptualization of impact interactions (i.e. the relationships between 2 linked impacts) through forms of network analysis for studying multi-sectoral patterns (de Brito, 2021; Chen et al., 2022). Meanwhile, clustering approaches have already been used for hydrological characteristics of droughts yet not for socio-economic impacts (Kim et al., 2021; Arabzadeh et al., 2016; Hao and Singh, 2015). Our approach has the potential to facilitate a multi-sectoral perspective on drought impact patterns as it can incorporate patterns of cascading and compounding impacts.

In addition to these methodological contributions, our work also adds to empirical knowledge on droughts in Germany. Concerning the differences between single-year and multi-year drought events, we showed distinct patterns in the multi-year drought event compared to single-year events. The lower spatial extent and diversity of impacts separated the single-year drought events from the multi-year drought period. Agriculture and livestock impacts dominated during the single-year events, while the multi-year drought period displayed a more diverse distribution of impacts. Several studies identified differing hydrological characteristics and effects on ecosystems between multi-year and single-year drought periods (Rakovec et al., 2022; Moravec et al., 2021; Tijdeman et al., 2022; Tsakiris et al., 2010). Specifically for southwestern Germany, Tijdeman et al. (2022) confirmed similar trends for 2003, 2015, 2018, and 2019, which they linked to changing biophysical conditions and the severity of the droughts. Next to the previously identified biophysical

https://doi.org/10.5194/nhess-2023-228
Preprint. Discussion started: 21 December 2023

Natural Hazards and Earth System Sciences Discussions

differences, our study thus makes a significant contribution by pointing out the differentiating factors concerning socio-economic impacts.

During the multi-year drought period, we discovered dynamically changing DIPs that led to an increasingly diverse landscape of impacts. Specifically, we found that an initial dominance of agriculture/livestock impacts was increasingly replaced by forestry impacts and, subsequently social/fire impacts. The emergence of impacts that increasingly gain relevance during multi-year drought periods reflects evidence from several studies (Tijdeman et al., 2022; Chen et al., 2022; Al-Faraj and Tigkas, 2016). For example, Chen et al. (2022) showed that during a multi-year drought period in 1920s China, cascading effects led to unprecedented effects such as growing food prices, dietary changes, and declining health conditions following agricultural losses. Concerning the multi-year drought period under investigation here, particularly the delayed effects on the forestry ecosystem from 2019 onwards, were pointed out by other studies. Repeated stress exposure caused tree damage that became evident throughout Central Europe (Schuldt et al., 2020; Buras et al., 2020; Kannenberg et al., 2020). Here, we advanced existing knowledge by showing the consequent effects on districts affected in the forestry sector, which later shifted to social impacts as visible in the Harz region (Hahne et al., 2009; Schütte and Plothe, 2022). Next to such sequential patterns, our longitudinal coverage of the multi-year drought period also revealed the sudden effects of extreme events. For instance, the high shares of water-related impacts in 2018 and 2022 were fostered by compounding drought and heat waves (Zscheischler and Fischer, 2020; Wieland and Martinis, 2020). By using a multi-sectoral perspective, we were able to detect such overarching trends that shaped the impact patterns across Germany and connected various sectors. Future research can leverage these identified trends to conduct more in-depth investigations into the mechanisms that underpin these dynamic shifts.

Our results also demonstrated that distinct land-cover types, such as forest or agricultural land, control the occurrence of impact patterns. We found intuitive connections between land-cover types and the DIPs. For instance, regions with high shares of agricultural land-cover were more likely to experience impacts on agriculture and livestock. We also unveiled subtler effects, demonstrating that coniferous forest land-cover heightened fire-related impacts, which aligns with research findings on German forests (Gnilke and Sanders, 2021). Instead, broad-leaved forests did not exhibit such an association. Identifying factors controlling impact patterns (such as exposure and vulnerability) is necessary to effectively design adaptation measures (Tijdeman et al., 2022; Bachmair et al., 2017; Rannow et al., 2010). Various case studies have demonstrated significant effects of land-cover (and land-use) when assessing drought risk and predicting impacts (Blauhut et al., 2016; Ihinegbu and Ogunwumi, 2022). For instance, Blauhut et al. (2016) found diverse land-cover types relevant for predicting drought risk across Europe. Yet, there remains a scarcity of publications addressing the relationships between multi-sectoral impacts and land-cover while researchers have delved into more nuanced distinctions within specific sectors like agriculture (Brown et al., 2011; Taiwo et al., 2023; Carter et al., 2013). Therefore, future research should advance both the exploration of additional variables (Knutson et al., 1998) and methods for linking these to impacts. As mentioned earlier, such progress will require impact datasets of greater spatio-temporal scope.

4. Conclusion

In this study, we analyzed the patterns of socio-economic drought impacts during both single-year and multi-year drought events in Germany. We found that during the multi-year drought period, an increasingly diverse landscape of drought impacts emerged that replaced dominating agriculture and livestock impacts. We noted distinct regional variances in impact patterns, characterized by shifts towards social and forestry-related consequences in some areas and relatively stable agriculture and livestock impacts in others. These findings underscore the need for localized and context-specific approaches to drought management that consider droughts' duration and cumulative effects. Finally, we demonstrated how these impact patterns are controlled by land-cover types, providing insights into the underlying exposure factors that drive them. Expanding on attributing the impact patterns in future research, we could design more targeted and effective drought adaptation strategies. Overall, our research provides an improved understanding of the unique shifts in socio-economic impacts during a multi-year drought period and highlights the potential of text- and pattern-mining methods to analyze complex drought impact patterns.

Code availability: The code for generating the impact dataset is available at https://github.com/jansodoge/drought-impact-text-mining, and the code for the analysis conducted here is provided at https://github.com/jansodoge/drought_impact_profiles_paper

Data availability: The newspaper corpus cannot be available due to licensing/copyright reasons. The impact dataset is available at https://github.com/jansodoge/drought_impact_profiles_paper

Author contribution:
Jan Sodoge: Conceptualization, Methodology, Software, Investigation, Data Curation, Formal analysis, Writing-Original draft preparation, Writing - Review & Editing, Visualization.
Mariana Madruga de Brito: Conceptualization, Investigation, Writing - Review & Editing, Visualization, Supervision, Funding acquisition.
Christian Kuhlicke: Conceptualization, Writing - Review & Editing, Project Administration, Supervision, Funding acquisition.
Miguel Mahecha: Methodology, Writing - Review & Editing

Competing interests: The authors declare that they have no conflict of interest.

Acknowledgments: Uwe Ehret is thanked for his support and consultation in integrating potential information theoretical methods into this research.

References

Al-Faraj, F. A. and Tigkas, D.: Impacts of multi-year droughts and upstream human-induced activities on the development of a semi-arid transboundary basin, Water Resour. Manag., 30, 5131–5143, 2016.

Arabzadeh, R., Kholoosi, M. M., and Bazrafshan, J.: Regional hydrological drought monitoring using principal components analysis, J. Irrig. Drain. Eng., 142, 04015029, 2016.

Bachmair, S., Svensson, C., Prosdocimi, I., Hannaford, J., and Stahl, K.: Developing drought impact functions for drought risk management, Nat. Hazards Earth Syst. Sci., 17, 1947–1960, 2017.

Bastos, A., Ciais, P., Friedlingstein, P., Sitch, S., Pongratz, J., Fan, L., Wigneron, J.-P., Weber, U., Reichstein, M., Fu, Z., and others: Direct and seasonal legacy effects of the 2018 heat wave and drought on

European ecosystem productivity, Sci. Adv., 6, eaba2724, 2020.

Blauhut, V., Gudmundsson, L., and Stahl, K.: Towards pan-European drought risk maps: quantifying the link between drought indices and reported drought impacts, Environ. Res. Lett., 10, 014008, 2015.

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

de Brito, M. M.: Compound and cascading drought impacts do not happen by chance: A proposal to quantify their relationships, Sci. Total Environ., 778, 146236, 2021.

de Brito, M. 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, 2020.

Brown, I., Poggio, L., Gimona, A., and Castellazzi, M.: Climate change, drought risk and land capability for agriculture: implications for land use in Scotland, Reg. Environ. Change, 11, 503–518, 2011.

Buras, A., Rammig, A., and Zang, C. S.: Quantifying impacts of the 2018 drought on European ecosystems in comparison to 2003, Biogeosciences, 17, 1655–1672, 2020.

Büttner, G., Feranec, J., Jaffrain, G., Mari, L., Maucha, G., and Soukup, T.: The CORINE land cover 2000 project, EARSeL EProceedings, 3, 331–346, 2004.

BZL: Waldbrandstatistik der Bundesrepublik Deutschland für das Jahr 2019, 2020.

Carter, D. R., Fahey, R. T., and Bialecki, M. B.: Tree growth and resilience to extreme drought across an urban land-use gradient, 2013.

Challinor, A. J., Adger, W. N., and Benton, T. G.: Climate risks across borders and scales, Nat. Clim. Change, 7, 621–623, 2017.

Challinor, A. J., Adger, W. N., Benton, T. G., Conway, D., Joshi, M., and Frame, D.: Transmission of climate risks across sectors and borders, Philos. Trans. R. Soc. Math. Phys. Eng. Sci., 376, 20170301, 2018.

Chen, X., Tian, F., and Su, Y.: How did the late 1920s drought affect northern Chinese society?, Weather Clim. Extrem., 36, 100451, 2022.

Conradt, T., Engelhardt, H., Menz, C., Vicente-Serrano, S. M., Farizo, B. A., Peña-Angulo, D., Domínguez-Castro, F., Eklundh, L., Jin, H., Boincean, B., and others: Cross-sectoral impacts of the 2018–2019 Central European drought and climate resilience in the German part of the Elbe River basin, Reg. Environ. Change, 23, 32, 2023.

Dahlmann, H., Stephan, R., and Stahl, K.: Upstream-downstream asymmetries of drought impacts in major river basins of the European Alps, Front. Water, 4, 1061991, 2022.

Damian, N., Mitrică, B., Mocanu, I., Grigorescu, I., and Dumitraşcu, M.: An index-based approach to assess the vulnerability of socio-ecological systems to aridity and drought in the Danube Delta, Romania, Environ. Dev., 45, 100799, 2023.

De Silva, M. and Kawasaki, A.: Socioeconomic vulnerability to disaster risk: a case study of flood and drought impact in a rural Sri Lankan community, Ecol. Econ., 152, 131–140, 2018.

DWD: Climate Data Center (CDC)—German Meteorological Service (DWD). 2020, 2023.

https://doi.org/10.5194/nhess-2023-228
Preprint. Discussion started: 21 December 2023

Natural Hazards and Earth System Sciences
Open Access
EGU
Discussions

Engelmann, I.: Journalistische Instrumentalisierung von nachrichtenfaktoren. einflüsse journalistischer einstellungen auf simulierte issue-, quellen-und statement-entscheidungen, MK Medien Kommun., 58, 525–543, 2010.

Erfurt, M., Skiadaresis, G., Tijdeman, E., Blauhut, V., Bauhus, J., Glaser, R., Schwarz, J., Tegel, W., and Stahl, K.: A multidisciplinary drought catalogue for southwestern Germany dating back to 1801, Nat. Hazards Earth Syst. Sci., 20, 2979–2995, 2020.

Erfurt-Cooper, P. and others: European waterways as a source of leisure and recreation, River Tour., 95–116, 2009.

Erian, W., Pulwarty, R., Vogt, J., AbuZeid, K., Bert, F., Bruntrup, M., El-Askary, H., de Estrada, M., Gaupp, F., Grundy, M., and others: GAR Special Report on Drought 2021, 2021.

Fischer, E., Sippel, S., and Knutti, R.: Increasing probability of record-shattering climate extremes, Nat. Clim. Change, 11, 689–695, 2021.

Flach, M., Brenning, A., Gans, F., Reichstein, M., Sippel, S., and Mahecha, M. D.: Vegetation modulates the impact of climate extremes on gross primary production, Biogeosciences, 18, 39–53, 2021.

Free, G., Van de Bund, W., Gawlik, B., Van Wijk, L., Wood, M., Guagnini, E., Koutelos, K., Annunziato, A., Grizzetti, B., Vigiak, O., and others: An EU analysis of the ecological disaster in the Oder River of 2022, 2023.

Garrick, D. E., Schlager, E., De Stefano, L., and Villamayor-Tomas, S.: Managing the cascading risks of droughts: Institutional adaptation in transboundary river basins, Earths Future, 6, 809–827, 2018.

Gnilke, A. and Sanders, T.: Forest fire history in Germany (2001-2020), Eberswalde Thünen Inst. For. Ecosyst., 2, 2021.

Google: Google Trends, 2023.

Hahne, U., Adams, C., and von Kampen, D.-I. S.: Tourismusdestination Nordhessen im Klimawandel: Betroffenheit und Chancen durch den Klimawandel, Arbeitspapier „Klimawandel Anpassung Tour., 2009.

Hao, Z. and Singh, V. P.: Drought characterization from a multivariate perspective: A review, J. Hydrol., 527, 668–678, 2015.

Holzwarth, S., Thonfeld, F., Abdullahi, S., Asam, S., Da Ponte Canova, E., Gessner, U., Huth, J., Kraus, T., Leutner, B., and Kuenzer, C.: Earth observation based monitoring of forests in Germany: A review, Remote Sens., 12, 3570, 2020.

Husson, F., Josse, J., and Pages, J.: Principal component methods-hierarchical clustering-partitional clustering: why would we need to choose for visualizing data, Appl. Math. Dep., 17, 2010.

Ihinegbu, C. and Ogunwumi, T.: Multi-criteria modelling of drought: a study of Brandenburg Federal State, Germany, Model. Earth Syst. Environ., 8, 2035–2049, 2022.

Kannenberg, S. A., Schwalm, C. R., and Anderegg, W. R.: Ghosts of the past: how drought legacy effects shape forest functioning and carbon cycling, Ecol. Lett., 23, 891–901, 2020.

Ketchen, D. J. and Shook, C. L.: The application of cluster analysis in strategic management research: an analysis and critique, Strateg. Manag. J., 17, 441–458, 1996.

https://doi.org/10.5194/nhess-2023-228
Preprint. Discussion started: 21 December 2023

Kim, J. E., Yu, J., Ryu, J.-H., Lee, J.-H., and Kim, T.-W.: Assessment of regional drought vulnerability and risk using principal component analysis and a Gaussian mixture model, Nat. Hazards, 109, 707–724, 2021.

Knutson, C., Hayes, M., and Phillips, T.: How to reduce drought risk, 1998.

Lawrence, J., Blackett, P., and Cradock-Henry, N. A.: Cascading climate change impacts and implications, Clim. Risk Manag., 29, 100234, 2020.

Liguori, A., McEwen, L., Blake, J., and Wilson, M.: Towards 'creative participatory science': exploring future scenarios through specialist drought science and community storytelling, Front. Environ. Sci., 8, 589856, 2021.

Lindner, M., Maroschek, M., Netherer, S., Kremer, A., Barbati, A., Garcia-Gonzalo, J., Seidl, R., Delzon, S., Corona, P., Kolström, M., and others: Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems, For. Ecol. Manag., 259, 698–709, 2010.

Llasat, M., Llasat-Botija, M., and López, L.: A press database on natural risks and its application in the study of floods in Northeastern Spain, Nat. Hazards Earth Syst. Sci., 9, 2049–2061, 2009.

Mahecha, M. D., Gans, F., Sippel, S., Donges, J. F., Kaminski, T., Metzger, S., Migliavacca, M., Papale, D., Rammig, A., and Zscheischler, J.: Detecting impacts of extreme events with ecological in situ monitoring networks, Biogeosciences, 14, 4255–4277, 2017.

Matusick, G., Ruthrof, K. X., Kala, J., Brouwers, N. C., Breshears, D. D., and Hardy, G. E. S. J.: Chronic historical drought legacy exacerbates tree mortality and crown dieback during acute heatwave-compounded drought, Environ. Res. Lett., 13, 095002, 2018.

McKnight, P. E. and Najab, J.: Mann-Whitney U Test, Corsini Encycl. Psychol., 1–1, 2010.

Miralles, D. G., Gentine, P., Seneviratne, S. I., and Teuling, A. J.: Land–atmospheric feedbacks during droughts and heatwaves: state of the science and current challenges, Ann. N. Y. Acad. Sci., 1436, 19–35, 2019.

Moravec, V., Markonis, Y., Rakovec, O., Svoboda, M., Trnka, M., Kumar, R., and Hanel, M.: Europe under multi-year droughts: how severe was the 2014–2018 drought period?, Environ. Res. Lett., 16, 034062, 2021.

Niggli, L., Huggel, C., Muccione, V., Neukom, R., and Salzmann, N.: Towards improved understanding of cascading and interconnected risks from concurrent weather extremes: Analysis of historical heat and drought extreme events, PLOS Clim., 1, e0000057, 2022.

Noone, S., Broderick, C., Duffy, C., Matthews, T., Wilby, R. L., and Murphy, C.: A 250-year drought catalogue for the island of Ireland (1765–2015), Int. J. Climatol., 37, 239–254, 2017.

Peña-Angulo, D., Vicente-Serrano, S., Domínguez-Castro, F., Lorenzo-Lacruz, J., Murphy, C., Hannaford, J., Allan, R. P., Tramblay, Y., Reig-Gracia, F., and El Kenawy, A.: The complex and spatially diverse patterns of hydrological droughts across Europe, Water Resour. Res., 58, e2022WR031976, 2022.

Rakovec, O., Samaniego, L., Hari, V., Markonis, Y., Moravec, V., Thober, S., Hanel, M., and Kumar, R.: The 2018–2020 Multi-Year Drought Sets a New Benchmark in Europe, Earths Future, 10, e2021EF002394, 2022.

Rannow, S., Loibl, W., Greiving, S., Gruehn, D., and Meyer, B. C.: Potential impacts of climate change in Germany—identifying regional priorities for adaptation activities in spatial planning, Landsc. Urban Plan., 98, 160–171, 2010.

Reyer, C., Bachinger, J., Bloch, R., Hattermann, F. F., Ibisch, P. L., Kreft, S., Lasch, P., Lucht, W., Nowicki, C., Spathelf, P., and others: Climate change adaptation and sustainable regional development: a case study for the Federal State of Brandenburg, Germany, Reg. Environ. Change, 12, 523–542, 2012.

Schmitt, J., Offermann, F., Söder, M., Frühauf, C., and Finger, R.: Extreme weather events cause significant crop yield losses at the farm level in German agriculture, Food Policy, 112, 102359, 2022.

Schuldt, B., Buras, A., Arend, M., Vitasse, Y., Beierkuhnlein, C., Damm, A., Gharun, M., Grams, T. E., Hauck, M., and Hajek, P.: A first assessment of the impact of the extreme 2018 summer drought on Central European forests, Basic Appl. Ecol., 45, 86–103, 2020.

Schütte, A. and Plothe, M.: Nachhaltige Forstwirtschaft im Zeichen des Klimawandels, in: Klimaschutz und Energiewende in Deutschland: Herausforderungen–Lösungsbeiträge–Zukunftsperspektiven, Springer, 767–794, 2022.

Sharma, S., Batra, N., and others: Comparative study of single linkage, complete linkage, and ward method of agglomerative clustering, in: 2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon), 568–573, 2019.

Simpson, N. P., Mach, K. J., Constable, A., Hess, J., Hogarth, R., Howden, M., Lawrence, J., Lempert, R. J., Muccione, V., Mackey, B., and others: A framework for complex climate change risk assessment, One Earth, 4, 489–501, 2021.

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., 100574, 2023.

Spellerberg, I. F. and Fedor, P. J.: A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon–Wiener'Index, Glob. Ecol. Biogeogr., 12, 177–179, 2003.

Stahl, K., Kohn, I., Blauhut, V., Urquijo, J., De Stefano, L., Acácio, V., Dias, S., Stagge, J. H., Tallaksen, L. M., Kampragou, E., and others: Impacts of European drought events: insights from an international database of text-based reports, Nat. Hazards Earth Syst. Sci., 16, 801–819, 2016.

Stephenson, N. L., Das, A. J., Ampersee, N. J., Cahill, K. G., Caprio, A. C., Sanders, J. E., and Williams, A. P.: Patterns and correlates of giant sequoia foliage dieback during California's 2012–2016 hotter drought, For. Ecol. Manag., 419, 268–278, 2018.

Sutanto, S. J., van der Weert, M., Wanders, N., Blauhut, V., and Van Lanen, H. A.: Moving from drought hazard to impact forecasts, Nat. Commun., 10, 1–7, 2019.

Taiwo, B. E., Kafy, A.-A., Samuel, A. A., Rahaman, Z. A., Ayowole, O. E., Shahrier, M., Duti, B. M., Rahman, M. T., Peter, O. T., and Abosede, O. O.: Monitoring and predicting the influences of land use/land cover change on cropland characteristics and drought severity using remote sensing techniques, Environ. Sustain. Indic., 18, 100248, 2023.

RDB: Regionalstatistik Datenbank Deutschland, 2023.

Thonfeld, F., Gessner, U., Holzwarth, S., Kriese, J., Da Ponte, E., Huth, J., and Kuenzer, C.: A First Assessment of Canopy Cover Loss in Germany's Forests after the 2018–2020 Drought Years, Remote Sens.,

https://doi.org/10.5194/nhess-2023-228
Preprint. Discussion started: 21 December 2023

Natural Hazards and Earth System Sciences
Discussions

14, 562, 2022.

Thorndike, R.: Who belongs in the family?, Psychometrika, 18, 267–276, 1953.

Tijdeman, E., Blauhut, V., Stoelzle, M., Menzel, L., and Stahl, K.: Different drought types and the spatial variability in their hazard, impact, and propagation characteristics, Nat. Hazards Earth Syst. Sci., 22, 2099–2116, 2022.

Tsakiris, G., Vangelis, H., and Tigkas, D.: Assessing water system vulnerability to multi-year droughts, Eur Water, 29, 21–29, 2010.

van der Wiel, K., Batelaan, T. J., and Wanders, N.: Large increases of multi-year droughts in north-western Europe in a warmer climate, Clim. Dyn., 60, 1781–1800, 2023.

Wieland, M. and Martinis, S.: Large-scale surface water change observed by Sentinel-2 during the 2018 drought in Germany, Int. J. Remote Sens., 41, 4742–4756, 2020.

Wright, A. P., Wright, A. T., McCoy, A. B., and Sittig, D. F.: The use of sequential pattern mining to predict next prescribed medications, J. Biomed. Inform., 53, 73–80, 2015.

Yu, H., Zhang, Q., Sun, P., and Song, C.: Impact of droughts on winter wheat yield in different growth stages during 2001–2016 in Eastern China, Int. J. Disaster Risk Sci., 9, 376–391, 2018.

Zaki, M. J.: SPADE: An efficient algorithm for mining frequent sequences, Mach. Learn., 42, 31–60, 2001.

Zink, M., Samaniego, L., Kumar, R., Thober, S., Mai, J., Schäfer, D., and Marx, A.: The German drought monitor, Environ. Res. Lett., 11, 074002, 2016.

Zscheischler, J. and Fischer, E. M.: The record-breaking compound hot and dry 2018 growing season in Germany, Weather Clim. Extrem., 29, 100270, 2020.

Zscheischler, J., Reichstein, M., Harmeling, S., Rammig, A., Tomelleri, E., and Mahecha, M. D.: Extreme events in gross primary production: a characterization across continents, Biogeosciences, 11, 2909–2924, 2014.

Appendix A

Table A.1: Definition of impact classes following de Brito et al. (2020)

Impact class	Definition
Agriculture	Impacts within the agricultural sector including the following sub-categories: reduced productivity of crops, early harvesting, increased need for irrigation, economic losses.
Livestock	Impacts within the livestock sector including the following sub-categories: reduced productivity of livestock farming, forced reduction of stock, shortage of feed for livestock, general impacts to animals (including e.g. insect mortality), economic losses for livestock farming
Social	Impacts within the social sector including the following sub-categories: parks, tourism, recreation areas and activities affected
Forestry	Reduces tree growth or vitality, water stress on trees, decrease in forestry products, increase in pest and disease attacks on trees, increased dieback of trees, economic losses for forestry

https://doi.org/10.5194/nhess-2023-228
Preprint. Discussion started: 21 December 2023

Aquaculture	Commercial and non-commercial fishing and aquaculture activities
Waterways	Impaired navigability of streams (reduction of load, increased need for interim storage transportation of goods at ports)
Fire	Occurrence of forest and wildfires

Table A.2: Performance of classification models to detect reported drought impacts in newspaper articles

Impact class	Recall	Precision	F-score	Accuracy	Sensitivity
Livestock	0.92	0.93	0.93	0.88	0.92
Fires	0.97	0.95	0.96	9.93	0.97
Forestry	0.94	0.9	0.92	0.89	0.94
Waterways	0.99	0.96	0.98	0.96	0.99
Aquaculture	0.85	0.93	0.83	0.74	074
Social	0.74	0.93	0.83	0.74	0.74
Agriculture	0.92	0.94	0.93	0.89	0.92

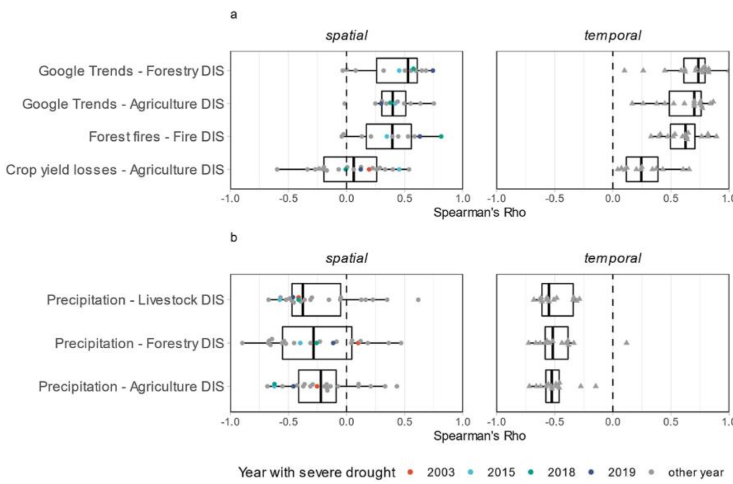


Figure A.1: Correlation of DIS with external validation indicators from Sodoge et al. (2023). For spatial correlations, each dot represents a year. For temporal correlations, each triangle represents a NUTS-1 unit. Subfigure a) describes correlation analysis in which an ideal explanation corresponds to Spearmans Rho = 1. Subfigure b) describes correlation analysis in which an ideal explanation corresponds to Spearmans Rho = −1.

https://doi.org/10.5194/nhess-2023-228
Preprint. Discussion started: 21 December 2023

640

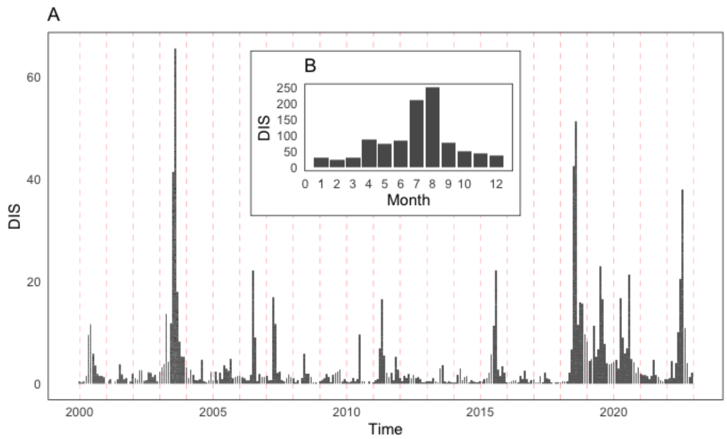
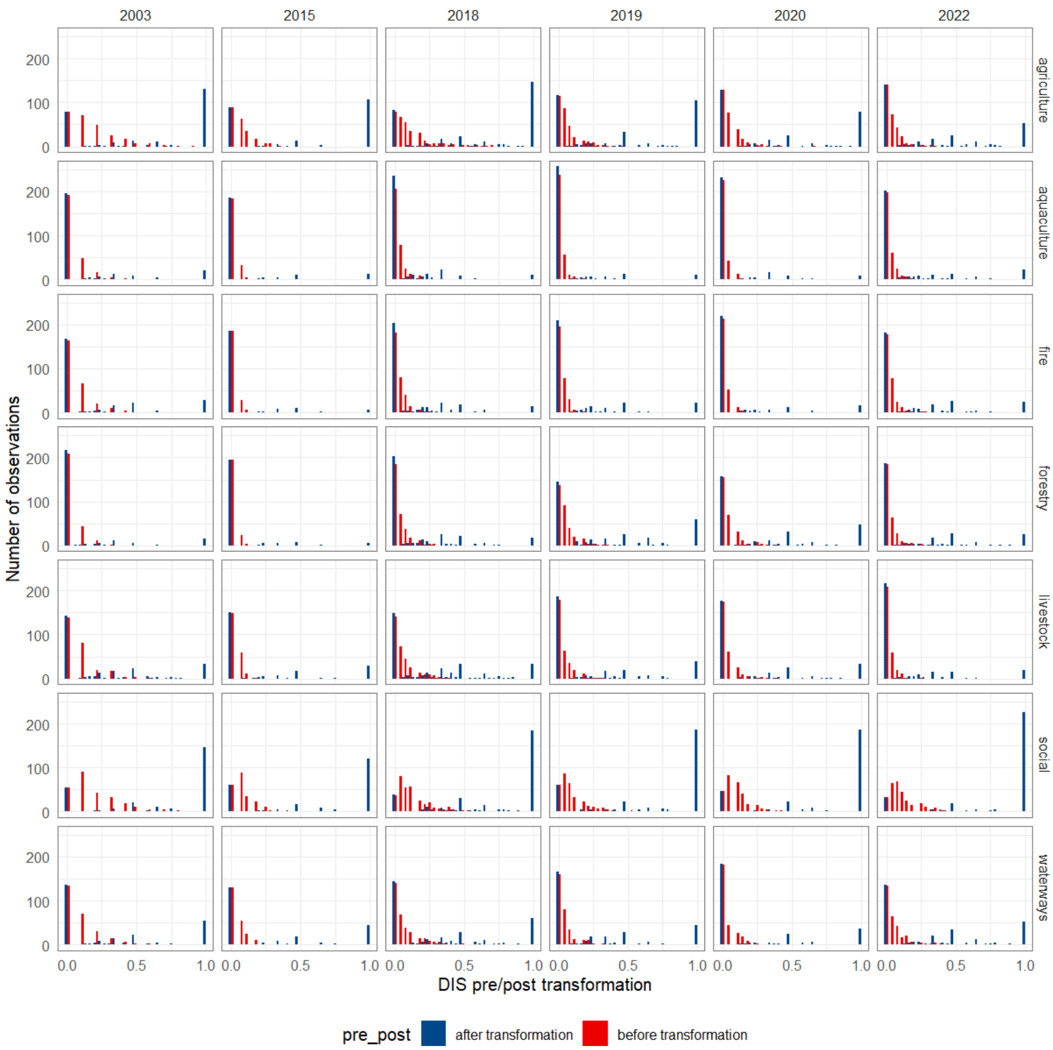


Figure A.2: Temporal distribution of DIS. (a) Temporal distribution for the entire period studied. Clear peaks exist for studied drought events. (b) total number of DIS per month. A normal distribution with peaks in July and only a few impacts reported during winter months.

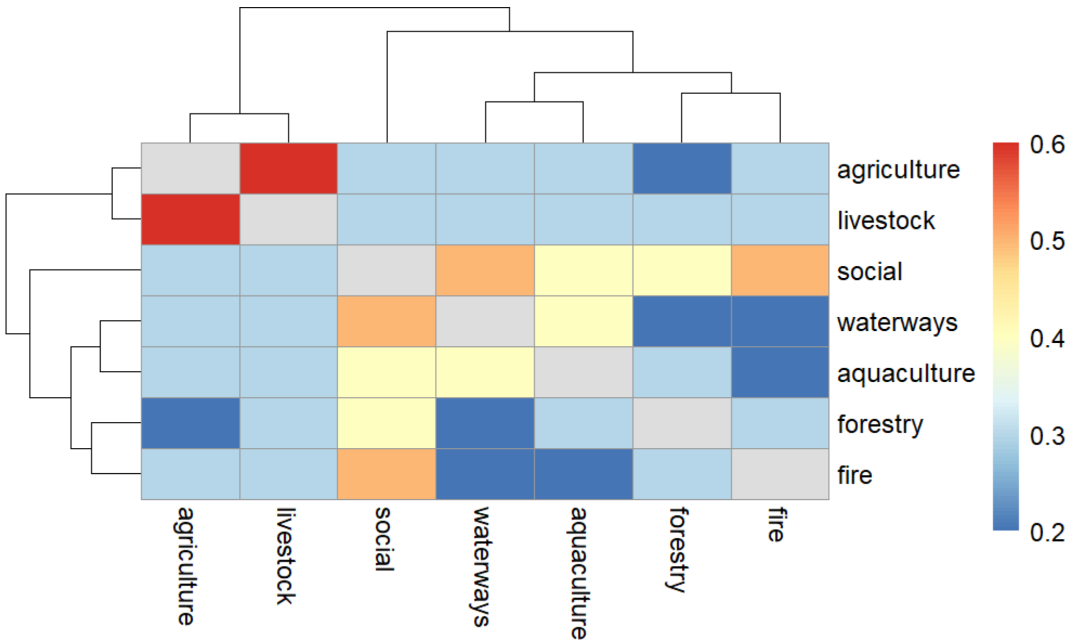
https://doi.org/10.5194/nhess-2023-228
Preprint. Discussion started: 21 December 2023



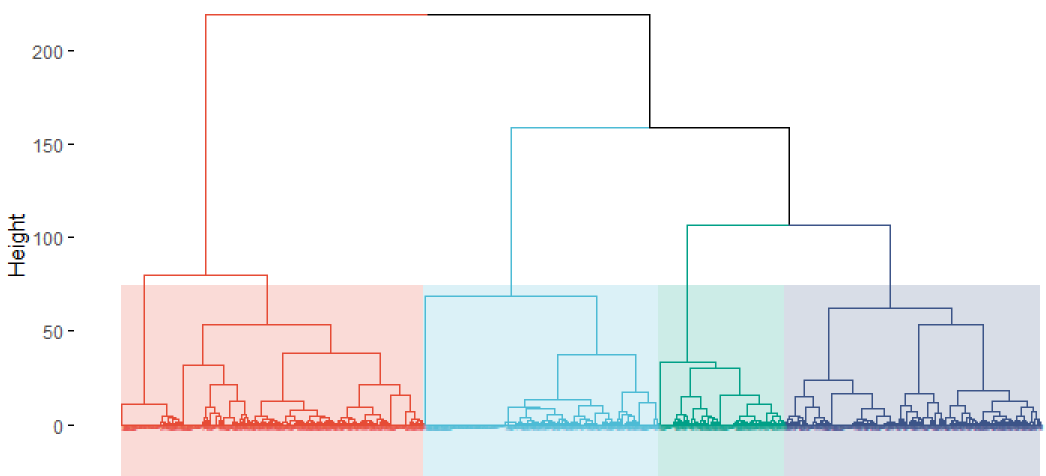
645

Figure A.3: Distribution of impacts before and after transformation, re-scaled to [0-1] interval for each grouping.

https://doi.org/10.5194/nhess-2023-228
Preprint. Discussion started: 21 December 2023



665 **Figure A.4: Correlations between the occurrences of different impact types. Correlation analysis was performed on the obtained drought impact dataset with annual aggregation before transformation for hierarchical clustering. Correlations are calculated using Spearman's Rho.**



670 **Figure A.5: Dendrogram of hierarchical clustering of DIS with the 4 clusters colored.**

https://doi.org/10.5194/nhess-2023-228
Preprint. Discussion started: 21 December 2023

675

Table A.3: Overview of evaluation metrics for obtained sequences

Item A	Item B	Support	Confidence	Lift
1	4	0.308	0.504	0.765
1	1	0.268	0.438	0.717
2	2	0.246	0.419	0.715
1	2	0.219	0.358	0.611
2	4	0.214	0.364	0.552
4	4	0.208	0.316	0.479