ADVANCING EARTH AND SPACE SCIENCES

Earth's Future

RESEARCH ARTICLE
10.1029/2024EF004590

This work has been done by Birgit M. Pfitzmann while at IBM Research.

Key Points:

- We map the global distribution of almost 300,000 abstracts from published flood, drought, and landslide research studies
- We find the distribution of published research to be biased against low-income countries and tropical regions, despite more people being affected there
- We define regions in need of targeted research and funding to reduce knowledge gaps and ultimately disaster impacts

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

L. Stein,
lina.stein@uni-potsdam.de

Citation:

Stein, L., Mukkavilli, S. K., Pfitzmann, B. M., Staar, P. W. J., Ozturk, U., Berrospi, C., et al. (2024). Wealth over Woe: Global biases in hydro-hazard research. *Earth's Future*, *12*, e2024EF004590. https://doi.org/10.1029/2024EF004590

Received 21 FEB 2024
Accepted 6 SEP 2024

Author Contributions:

Conceptualization: Lina Stein, S. Karthik Mukkavilli, Ugur Ozturk, Thomas Brunschwiler, Thorsten Wagener
Data curation: Lina Stein, S. Karthik Mukkavilli, Birgit M. Pfitzmann, Peter W. J. Staar, Cesar Berrospi
Methodology: Lina Stein, S. Karthik Mukkavilli, Birgit M. Pfitzmann, Thomas Brunschwiler, Thorsten Wagener
Software: S. Karthik Mukkavilli, Birgit M. Pfitzmann, Peter W. J. Staar, Cesar Berrospi, Thomas Brunschwiler

Wealth Over Woe: Global Biases in Hydro-Hazard Research

Lina Stein[1], S. Karthik Mukkavilli[2], Birgit M. Pfitzmann[2,3], Peter W. J. Staar[2], Ugur Ozturk[1,4], Cesar Berrospi[2], Thomas Brunschwiler[2], and Thorsten Wagener[1]

[1]Institute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany, [2]IBM Research—Europe, Zurich, Switzerland, [3]Smart City & ERZ Zurich, Zurich, Switzerland, [4]Helmholtz Centre Potsdam—GFZ German Research Centre for Geosciences, Potsdam, Germany

Abstract Floods, droughts, and rainfall-induced landslides are hydro-hazards that affect millions of people every year. Anticipation, mitigation, and adaptation to these hazards is increasingly outpaced by their changing magnitude and frequency due to climate change. A key question for society is whether the research we pursue has the potential to address knowledge gaps and to reduce potential future hazard impacts where they will be most severe. We use natural language processing, based on a new climate hazard taxonomy, to review, identify, and geolocate out of 100 million abstracts those that deal with hydro-hazards. We find that the spatial distribution of study areas is mostly defined by human activity, national wealth, data availability, and population distribution. Hydro-hazard events that impact large numbers of people lead to increased research activity, but with a strong disparity between low- and high-income countries. We find that 100 times more people need to be affected by hazards before low-income countries reach comparable research activity to high-income countries. This "Wealth over Woe" bias needs to be addressed by enabling and targeting research on hydro-hazards in highly impacted and under-researched regions, or in those sufficiently socio-hydrologically similar. We urgently need to reduce knowledge base biases to mitigate and adapt to changing hydro-hazards if we want to achieve a sustainable and equitable future for all global citizens.

Plain Language Summary Floods, droughts, and landslides are "natural hazards" responsible for the deadliest and most costly disasters globally. The scientific community studies these hazards to reduce their undesired impacts on society. To assess whether these research efforts are well-targeted, we require a global overview of where these hazards are studied and whether impacted regions are considered. Hence, we create a global map of flood, drought, and landslide research that shows whether published research is distributed equitably. We find that there is more research in regions where many people live, in wealthy regions, and in regions that have had disasters happening in the past. However, the level of research in wealthy countries is much higher despite considerably more people being affected by disasters in low-income countries. Based on our findings, we recommend regions where more research is needed for an equitable distribution of research so that all of global society is better prepared for future disasters.

1. Introduction

Hydro-hazards, such as floods, droughts, and rainfall-induced landslides, affect millions of people and cause thousands of fatalities annually. According to the Center for Research on the Epidemiology of Disasters (CRED), floods and droughts together affected more than 130 million people in 2022 alone. Critically, the risk from hydro-hazards will keep increasing due to projected climate and anthropogenic change (Arnell et al., 2019; IPCC, 2022), which already overwhelms disaster risk reduction efforts (Kreibich et al., 2022). The clear societal threats posed by hydro-hazards suggest that science should tackle knowledge gaps to better guide adaptation policies where the risk is greatest. However, existing natural hazard research likely overlooks many countries or regions which are not studied in depth despite their exposure to hydro-hazards. For example, only 6.5% of all natural hazard research studies are performed in Africa (Emmer, 2018) despite this continent having the largest predicted increase in flood exposure (Jongman et al., 2012).

Biased research distributions can be found across several disciplines including medicine (Sumathipala et al., 2004), conservation science (Di Marco et al., 2017), geoscience (North et al., 2020), and climate science (Callaghan et al., 2021). Biases have systemic causes such as differences in research funding (Overland et al., 2022; Woelbert et al., 2021), discrimination in the academic publishing system (Singh, 2006), data availability (Lindersson et al., 2020; Mwampamba et al., 2022), and language barriers (North et al., 2020).

Supervision: Thomas Brunschwiler, Thorsten Wagener
Validation: Lina Stein
Visualization: Lina Stein
Writing – original draft: Lina Stein, Thorsten Wagener
Writing – review & editing: Lina Stein, S. Karthik Mukkavilli, Ugur Ozturk, Cesar Berrospi, Thomas Brunschwiler, Thorsten Wagener

However, for hydro-hazards, there are substantial knowledge gaps regarding which environmental, anthropogenic, and socio-economic characteristics determine research foci and biases. We lack quantitative information regarding which regions are underrepresented in studies of hydro-hazards. Quantifying and mapping these biases is key to revealing and eventually addressing their underlying causes. For hydro-hazards, the large spatial variability of the components of risk (i.e., hazard, vulnerability, and exposure) complicates bias analyses. Threats from floods, droughts, and landslides are highly heterogeneous, for example, landslides are gravitational mass movements and occur predominantly in rugged not flat terrain. The exposure to any natural hazard depends on hazard magnitude and population distribution (Devitt et al., 2023). Differences in people's vulnerability, for example, due to their socio-economic situation, further determine how strongly they might be affected when a hazard occurs (Benevolenza & DeRigne, 2019). The potential for negative impacts (or risk) from hydro-hazards depends on the integration of hazard, exposure, and vulnerability. Therefore, we would not expect the global research landscape to be spatially homogeneous, given that the risk is not spread in this way. Instead, we would expect a fair research distribution to follow one or a combination of the following aspects:

1. *Socio-Hydrological Variations*: Research is conducted where scientific knowledge gaps have been identified. To advance scientific understanding, the scientific community should aim for research that is representative of the underlying socio-hydrological processes, in regard to both hazard generation and risk. Representative knowledge distribution is particularly relevant in the context of vulnerability, as it is highly spatially heterogeneous and results are difficult to transfer to other communities (King-Okumu et al., 2020; Ward et al., 2020).
2. *Impact Density*: Research is conducted where the impact or risk is largest. Impact can be measured as the number of events, fatalities, people affected, or economic loss. For our analysis, we mainly focus on the number of events and people affected. We disregard fatalities and economic losses since fatalities are underreported for drought events (UNDRR, 2021) and economic impact data disproportionately favors high-income countries (King-Okumu et al., 2020). The only exception is a supplemental analysis of landslide fatalities, as they are considered more accurate than the number of people affected (Froude & Petley, 2018).
3. *Population Density*: Finally, an equitable distribution might simply entail an equal allocation of studies according to the distribution of people.

Aiming for representative research coverage regarding hydro-climatic, landscape, and socio-economic characteristics is not only important for addressing the current hazard situation but also for predicting and projecting future risk. We investigate a corpus of 100 million scientific abstracts (Kinney et al., 2023) by extracting and geolocating those studies focused on hydro-hazards. We compare the spatial distribution of these abstracts with hydro-climatic, socio-economic, and disaster impact data to determine biases in the current knowledge base. And finally, to address these biases, we recommend high-priority regions for future research and funding. Our results integrate knowledge on hydro-hazards for disaster risk reduction and contribute toward a more sustainable and equitable global research landscape.

2. Materials and Methods

2.1. Abstract Data Mining and Annotation With Hydro-Hazards Taxonomy

Figure 1 provides an overview of the database as well as filtering and geolocation steps to identify and geolocate research related to hydro-hazards for subsequent estimation of global research distributions. Each step is described in detail below:

Abstract Database: The Semantic Scholar Academic Graph (Kinney et al., 2023) formed our basis for data mining. Currently, it contains 215 million scientific documents from all scientific fields, published and indexed by non-profit organizations like Crossref or PubMed, preprint repositories such as arXiv, and academic publishers like Springer Nature. Within the Semantic Scholar corpus, the abstracts data set provides abstract texts for around 100 million records. We utilized Deep Search (Staar et al., 2020) (https://ds4sd.github.io/), a tool that uses natural language processing to ingest and analyze unstructured data (Figure 1a). Deep Search processes text from the abstract data set and enriches the metadata (e.g., doi, title, abstract text…), for instance through language detection. As subsequent search and filtering was based on English language keywords, we used this information to filter out non-English abstracts. 95% of all abstracts were in English (Figure S1 in Supporting Information S1). The metadata associated with each abstract includes entries like unique identifiers, language, publication date, or

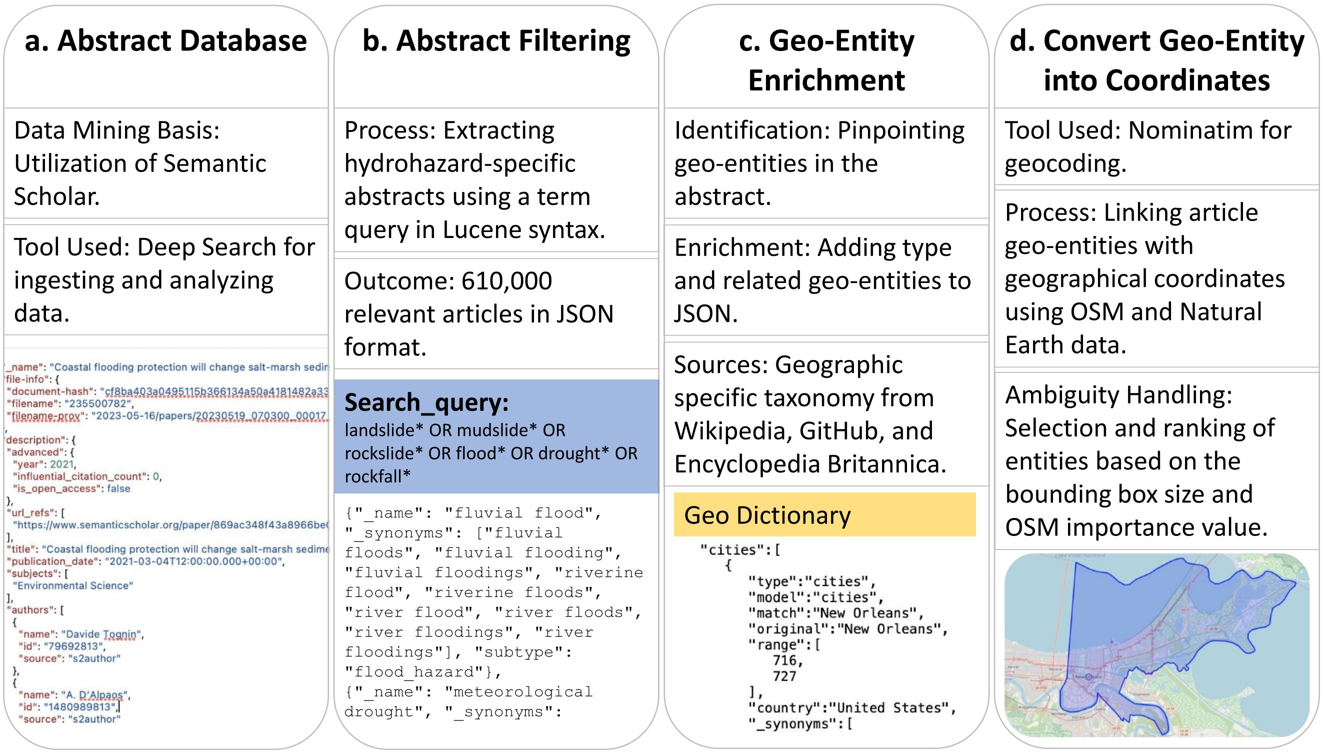

Figure 1. Overview of methodological steps for abstract search, annotation, and geolocation. The abstract database (Kinney et al., 2023) was processed using DeepSearch (Auer et al., 2022; Pyzer-Knapp et al., 2022; Staar et al., 2020).

subject (e.g., Environmental Science). We further excluded subjects related to the humanities, such as history, philosophy, and art.

Abstract filtering: We first extracted all hydro-hazard-specific abstracts from the 100 million documents using a term query (Figure 1b) in Lucene syntax (i.e., landslide OR mudslide OR rockslide OR flood OR drought OR rockfall) within Deep Search. As a result, 610,000 relevant articles remained. We then classified each abstract according to all hazards mentioned in that abstract as drought-related, flood-related, or landslide-related. Abstracts mentioning multiple hazards were counted for each category. We created a climate-specific taxonomy for hydro-hazards for the classification, which includes relevant hazard types and sub-types, along with possible synonyms. For example, "floods" are classified under "flood hazard," encompassing different forms of floods such as "flash flood", "stormwater," "outburst flood," "fluvial flood," and others. Synonyms for, for example, "fluvial flood" include "river flood," "riverine flood," etc. A full overview of hazard entities can be found in Table S1 in Supporting Information S1, while the entire taxonomy is part of the supplemental data.

Geo-entity enrichment: We employed a hybrid rule-based and gazetteer matching approach for location word identification (toponym recognition) (Hu et al., 2023). The rule-based approach identified locations based on natural feature keywords (area, basin, fold, rift, river, range,…), in combination with detecting capitalization. We build a dictionary of location names (i.e., a targeted gazetteer) to identify locations mentioned within the abstract (Figure 1c). We included location names for administrative areas, regions, lakes, rivers, and basins. Geographic taxonomy information about towns and cities with at least 100,000 inhabitants was sourced from Wikipedia's rich open knowledge base (Lehmann et al., 2015) and was further augmented with GitHub open-source collections for smaller capitals and cities by countries, as well as the Encyclopedia Britannica for lakes and rivers (Table S2 in Supporting Information S1). By limiting the gazetteer to large, administrative, and natural features we aimed to reduce possible ambiguity (Hu et al., 2023) and directly classified location entities according to type (e.g., match: "New Orleans," type: cities).

Converting geographic entities into coordinates: We used a combination of the geocoding software Nominatim (Clemens, 2015) and data from Natural Earth Data (NE, www.naturalearthdata.com) to geolocate the identified

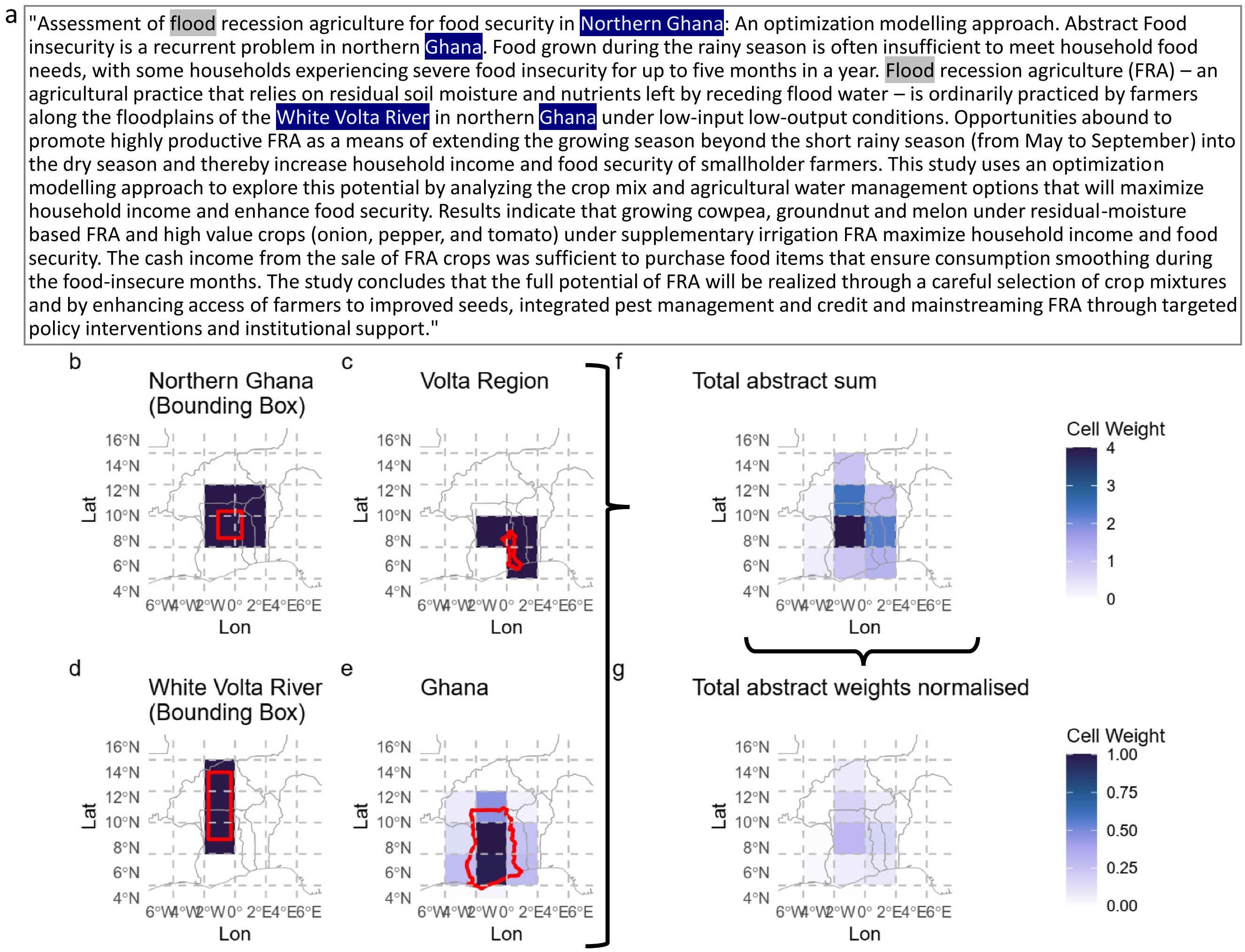

Figure 2. Schematic showing single abstract processing. (a) Abstract (Balana et al., 2019) with annotated hazards (gray) and geolocations (blue), (b–e) geo entity polygon (red) with underlying raster weights. (b) bounding box of Open Street Map entity. (c–e) polygons/bounding box extracted from Natural Earth Data. Rivers were extracted as bounding boxes for a vague estimate of catchment outline. (e) for country shapes, each cell is weighted according to the fraction covered by its shape. (f) Sum of raster (b–e). (g) Grid divided by the total sum of all cells to normalize the raster grid for each abstract to a sum of 1. This ensures comparable weights between abstract raster grids, independent of the number of geo-entities tagged.

geo-entities. Nominatim searches OpenStreetMap https://www.openstreetmap.org/copyright (OSM) (Bennett, 2010; Haklay & Weber, 2008) data. In case of ambiguity (e.g., multiple identical geo-entities), the five largest entities returned by Nominatim were selected and further ranked based on their OSM importance values, indicating search popularity (e.g., Paris, France: 0.8 vs. Paris, Texas: 0.5). We used data from NE to supplement the OSM results and to improve shape outlines of large features such as regions and continents. The matching was based on geo-entity name and identified type (e.g., "rivers," "countries"). Manual evaluation showed that this approach was more accurate in identifying regions and natural features than Nominatim alone. Final coordinates are based on feature bounding boxes for OSM and river lines, as well as exact polygon shapes for all other NE data.

2.2. Abstract to Grid Conversion

We used the geolocated entities to calculate a gridded distribution of the area each abstract covers. Figure 2 demonstrates this process. For each of the four locations identified within the abstract (Figure 2a) the grid cells that are touched by the location polygon are given the weight of 1 (unless it is a country, where cell weight is based on coverage). The sum of the four grids (Figure 2f) is then divided by the total grid sum (13.56 in this case), resulting in a weighted research distribution (Figure 2g). This process produces greater weights for cells where multiple locations overlap.

Creating a spatial grid for each abstract enabled us to calculate the density distribution of studies so that we could compare them with other data sets (e.g., population density) that were also transformed onto the same grid resolution. Similar to Callaghan et al. (2021), we chose a raster grid of 2.5°. However, unlike them, we considered not just the smallest but all locations extracted from an abstract. We commonly found that multiple equally relevant study locations are mentioned in one abstract without relevancy distinction. A country might be mentioned either as a study or modeling domain itself or just to specify the location of a smaller entity for the reader. An alternative counting method was used to calculate absolute numbers of abstracts per country. All geolocations that fell within a country (excluding continents and marine regions) were counted, and the number of unique abstracts per country was calculated.

2.3. Manual Evaluation of Annotation Quality

The combined OSM and NE tagged geo-entity data set was manually checked, and wrong results that frequently occurred were removed. For example, the frequent geo-entity "Mobile" is often misidentified as Mobile County in Alabama. A full list of these manual edits is provided in the supplement. Afterward, eight evaluators manually assessed 418 abstracts to determine geolocation annotation accuracy. The evaluation focused on three aspects: 1. Accuracy of the identified location words (Is the identified entity a location?). 2. Accuracy of the geolocation. And 3. missed locations. Of the 418 abstracts 288 (69%) had automatically annotated locations, with a total of 779 identified locations across all abstracts. Figure S2 in Supporting Information S1 gives a full overview of evaluation statistics.

Regarding aspect 1. Precision and recall are standard information retrieval metrics that are commonly used to evaluate location recognition (Hu et al., 2023). Ting (2010) defines precision as "*Total number of documents retrieved* [locations in our case] *that are relevant/Total number of documents that are retrieved*" and recall as "*Total number of documents retrieved that are relevant/Total number of relevant documents in the database*." We reach a precision value of 0.91 and a recall value of 0.78 (Figure S2a in Supporting Information S1). In comparison, Hu et al. (2023) evaluate 27 common toponym recognition methods on 26 different data sets. The 27 methods range in precision between 0.477 and 0.868 and in recall between 0.261 and 0.784. Our approach thus reaches state-of-the-art accuracy in location recognition.

Regarding aspects 1. *and* 2.: 91.1% of all annotated locations have been correctly geolocated (Figure S2b in Supporting Information S1). However, in 22% of abstracts with at least one location and in 3% of abstracts without a location entity, at least one location entity has been missed. This seems like a relatively high number. We therefore further evaluated the influence of missing and wrong locations on the research distributions. In total we identified 202 missed locations. 19% of these missed locations could not be found on OSM by the evaluators either and therefore could not be geolocated. This result reflects the limits of the OSM database. For all abstracts with missing and wrong locations that could be located (120 abstracts, Figure S2c in Supporting Information S1), we test if adding or correcting the locations influences the extent of the covered grid cells to evaluate the reliability of the final research distributions. We find that for 76% of the abstracts, the extent does not change, meaning that missed or wrong locations fall within the already identified locations (e.g., the town "Wakkanai" has been missed, but is contained within the larger entity the island of "Hokkaido," which has been identified). Additionally, the average Pearson correlation between original and corrected abstract density grids is on average 0.89, suggesting a low impact from the additional location entities. We further analyzed if the distribution of evaluated locations across country income groups differs between all evaluated locations as well as missed or wrong locations (Figure S3 in Supporting Information S1). A larger share of missed or wrong locations in low-income countries would indicate a bias in our analysis due to a bias in our location dictionary or OSM. However, Figure S3 in Supporting Information S1 reveals that this is not the case.

2.4. Bias Analysis

Biases in research distributions were determined by comparing the distributions of four data categories: 1. Impact data, 2. Hydro-meteorologic measurement stations, 3. Socio-economic data, 4. Natural and anthropogenic features of the landscape. All data sets were transformed to the same grid as the abstract data. For **impact data**, the international disaster database EM-DAT (CRED, 2023b) was combined with the Geocoded Disasters Database

10.1029/2024EF004590

(GDIS) (Rosvold & Buhaug, 2021a) to create geolocated impact data. Hazard events are only considered for EM-DAT if certain impact criteria based on severity are met, such as more than 10 dead, more than 100 affected, a state of emergency was declared, or international assistance was called. However, getting accurate impact numbers for disaster events can be a challenge (Guha-Sapir & Below, 2006), and many events are missing information in EM-DAT, for example, information on the number of deaths and the number of people affected (Jones et al., 2022). Other impact databases exist but have their own biases. A consolidated impact database from different sources is currently missing (Wyatt et al., 2023). We therefore supplement our analysis by comparing our outcomes to three additional disaster-specific, continually updated data sets commonly utilized by their respective communities: the Dartmouth Flood Observatory (Brakenridge, 2023), the NASA global landslide catalog (Kirschbaum et al., 2010), and the Global Fatal Landslide Database (Froude & Petley, 2018). Both landslide databases focus on rainfall-induced landslides and are widely used within the landslide research community.

We compared **measurement station data** to the identified research distributions to determine where a lack of data might be a factor in contributing to research gaps. We considered the distribution of stations from the WMO Integrated Global Observing System (called OSCAR) (World Meteorological Organization (WMO) & Federal Office of Meteorology and Climatology (MeteoSwiss), 2023), Global Precipitation Climatology Center (GPCC) stations (Rustemeier et al., 2022), the international soil moisture network (ISMN) (Dorigo et al., 2011), and a global streamflow stations data set (GSIM) (Do et al., 2018). We mainly refer to the World Development Indicators and Worldwide Governance Indicators (Kaufmann & Kraay, 2022; World Bank, 2023) from the World Bank Open Data Catalog for **socio-economic data** accessed via the “wbstats” R package (Piburn, 2020). Additional socio-economic indices are population (WorldPop, 2023), human development index (Kummu et al., 2018), and the adaptive capacity measure by the Notre Dame Global Adaptation Initiative (ND-GAIN) (C. Chen et al., 2015). We considered human footprint as a general measure of anthropogenic impact (Venter et al., 2016), and travel time to the nearest city above 100,000 inhabitants as a measure of closeness to urban centers (Hijmans et al., 2023; Nelson et al., 2019a). We used ESA World Cover for forest and crop coverage (Zanaga et al., 2021), and precipitation (P), potential evapotranspiration (PET), and aridity (PET/P) as measures of climate zone (Karger et al., 2017). A full list of data sets used, including details and their references, can be found in Supporting Information S1 (Table S1).

We used the Wasserstein distance (Kantorovich, 1960; Krabbenhoft et al., 2022; Schuhmacher et al., 2023) to determine differences in variable distributions between regions of high research density (>75th percentile) and the entire world as a measure of bias. The Wasserstein distance is a measure of the absolute difference between cumulative distributions and does not indicate the direction of bias. We therefore combine Wasserstein difference with a second statistic to calculate the direction of bias. For that we used the summarized difference between cumulative distribution functions (Stein et al., 2021). A positive difference between distributions indicates that an increase in variable value leads to an increase in research density. Where country-averaged values were used (e.g., for research density or impact calculation, Figure 6), we used a weighted mean average based on the fraction of cells covered by each country polygon. Country averages instead of total sums are used to compensate for different country sizes.

3. Results

3.1. Global Distribution of Hydro-Hazard Research

Out of 610,000 abstracts that include variations of the search terms “drought,” “flood,” and “landslide,” further screening (Figure S1 in Supporting Information S1) leaves us with 293,156 abstracts for analysis. We calculated research density as research per cell weighted by the size of the location entity (Callaghan et al., 2021). We define highly researched regions as all locations with a research density above the 75th quantile of all land cells. The exact regions are shown in Figure S5 in Supporting Information S1.

The global distributions of hydro-hazards research densities depicted in Figures 3a, 3d, and 3g show distinct patterns for each hazard. A noticeable hotspot for **drought** research is the west coast of the USA, while further highly researched areas can be found across much of Europe (UK, Switzerland, Italy, and Spain) and Asia (South Korea, Bangladesh). Other highly researched regions are located in Africa. Ethiopia, for example, is among the five most highly researched countries for droughts (Figure S13 in Supporting Information S1). Other African countries that are highly researched are Kenya, Nigeria, Tanzania, and Zimbabwe (Figure S5 in Supporting

10.1029/2024EF004590

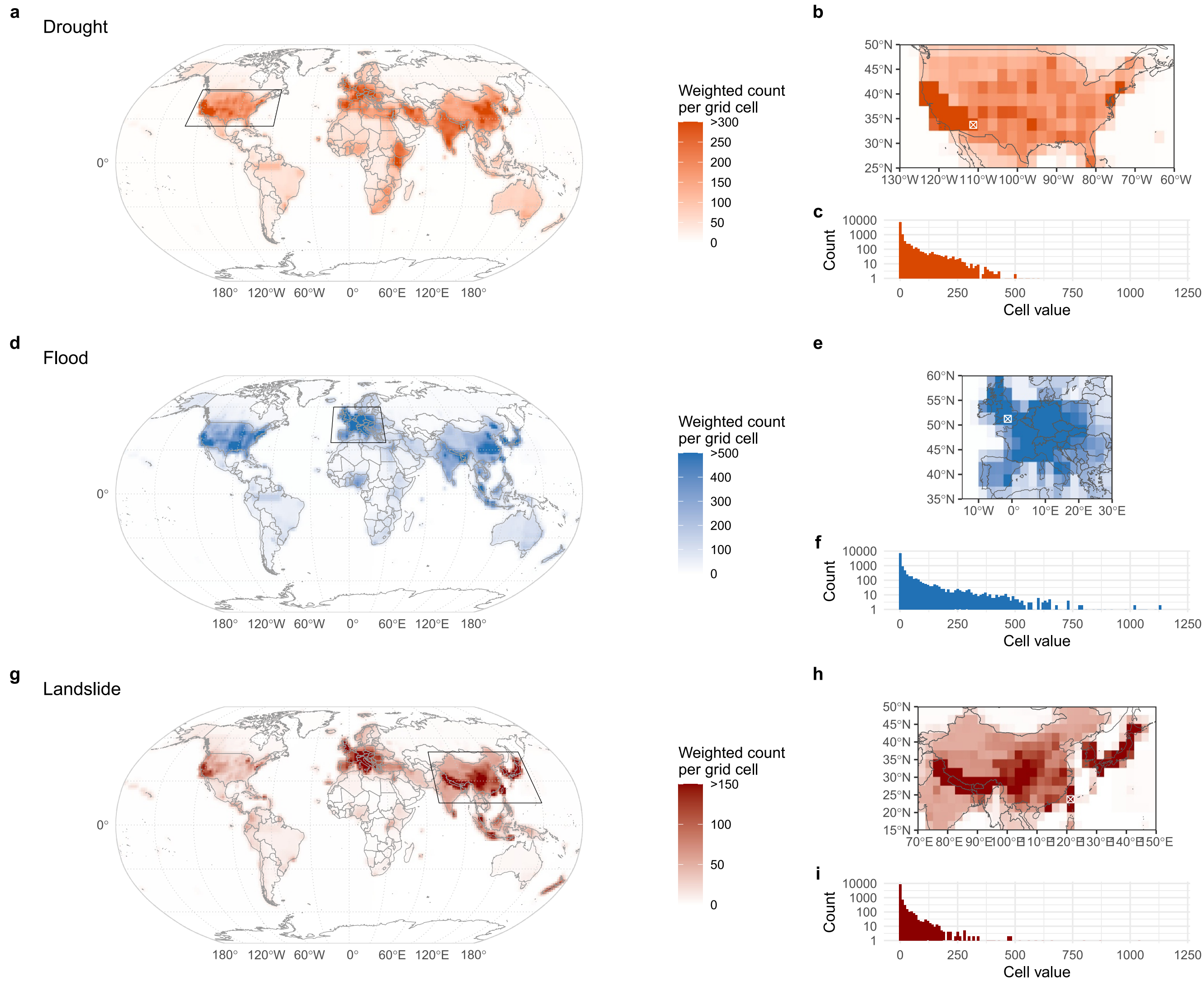

Figure 3. For each water extreme, the research distribution is displayed in three panels. A global map of weighted research count, a detailed map for the highest cell count (marked by x), and a histogram across all raster cells for droughts (a–c), floods (d–f), and landslides (g–i).

Information S1). Drought study numbers are low for Latin America, Central Africa, Russia, Kazakhstan, Mongolia, and Canada. In absolute numbers, Russia is mentioned often (Figure S6 in Supporting Information S1), but the size of the country makes individual cell weights low and we found no small-scale studies. **Flood** research density is generally higher due to larger number of articles than for the other hazards. Flood research has several clusters around Europe, the USA, and Asia, such as Bangladesh, eastern China, Japan, and South Korea. The cell with the highest flood study count is located in the south of England (a cell including London and the Thames). About 5% (8,616 in total) of all flood abstracts target the UK. For comparison, Nigeria is the country with the largest number of flood studies in Africa, with 2,595 abstracts. Flood research in South and Central America and most of Africa is low. **Landslide** research has distinct hotspots, especially in the Alps, Italy, Taiwan, Hong Kong, the Himalayas, Central China, and Japan. Taiwan is the cell with the highest research count overall. In terms of absolute numbers, China is the country with the largest number of abstracts about landslide research, with 6,571 abstracts in total.

3.2. Research Distribution Across Climate Zones

We analyze the research bias between climate zones by comparing study numbers against the number of hazard events and population numbers in each climate zone. Temperate regions have, on average, the highest research

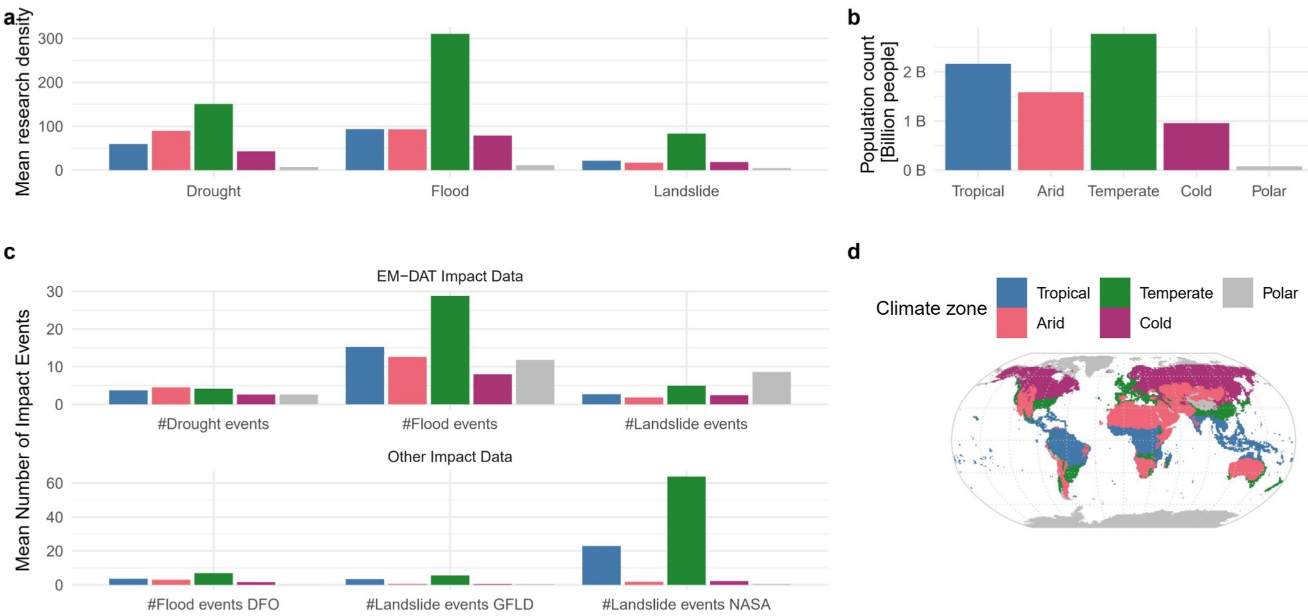

Figure 4. (a) Mean research density across broad climate zones according to Koeppen-Geiger (H. E. Beck et al., 2018), (b) population count (WorldPop, 2023) by climate zone, (c) mean number of events per cell and climate zone for EM-DAT event counts as well as one flood and two landslide data sets (Dartmouth Flood Observatory, Global Fatal Landslide Database (GFLD), NASA landslide catalog), (d) world map depicting the climate zones.

count for all three hazards (Figure 4a). In terms of hazard event counts (Emergency Management Database, EM-DAT, Figure 4c, upper panel), that distribution is mirrored by flood event occurrences, but not drought or landslide events. Most flood events (mean 28.8 per cell) also occur in temperate regions. The average flood count in tropical regions is about half that of temperate regions (mean 15.2 per cell), yet the research density is only about a third. This result suggests a flood research bias against tropical regions. A large share of flood events (mean 11.8 per cell) also occurs in polar regions, showing the lowest research density by far. Drought events are evenly distributed among climate zones. Drought research effort is much higher in temperate regions than in arid and tropical regions though, indicating a bias toward temperate and against tropical and arid regions. For landslides, the identified bias strongly depends on the choice of the event count data set (e.g., EM-DAT vs. NASA landslide catalog vs. the Global Fatal Landslide Database—GFLD, Figure 4c, lower panel). The comparison suggests biases in the event count data sets themselves. Additionally, we compare the research distribution across climate zones with the population distribution across climate zones. The dominance of research in temperate regions matches the higher share of the population in that climate zone (36%, Figure 4b). Yet, tropical regions with only 22% fewer people than temperate regions have 60% (drought), 70% (floods), and 74% (landslides) lower research densities.

3.3. Environmental and Socio-Economic Controls on Research Distributions

We further analyze how these research study distributions co-vary with different environmental and socio-economic characteristics and with the availability of hydro-meteorologic measurements. Hence, we extract the land surface with high research density (>75th quantile, Figure S5 in Supporting Information S1) and compare its characteristics with those of the whole land surface. Differences between distributions are quantified using the Wasserstein metric (Kantorovich, 1960; Krabbenhoft et al., 2022). Figure 5 shows Wasserstein distances for selected variables (all variables: Figure S8 in Supporting Information S1).

Multiple variables indicate a strong positive bias in research density towards regions that are highly influenced by human activity. Human footprint, representing aspects of human pressure on the environment (Venter et al., 2016), as well as the variables irrigated land, population count, cropland, and travel time to the nearest city as an indicator of urbanization all exhibit high Wasserstein values (>0.5). Wasserstein values are lower (on average <0.4) for climatic indices such as potential evapotranspiration, precipitation, and aridity. Average annual precipitation is the only climatic variable that has a large spread of Wasserstein values across hazards (0.14 for drought, 0.24 for flood, and 0.36 for landslide research). Furthermore, we observed opposing distribution

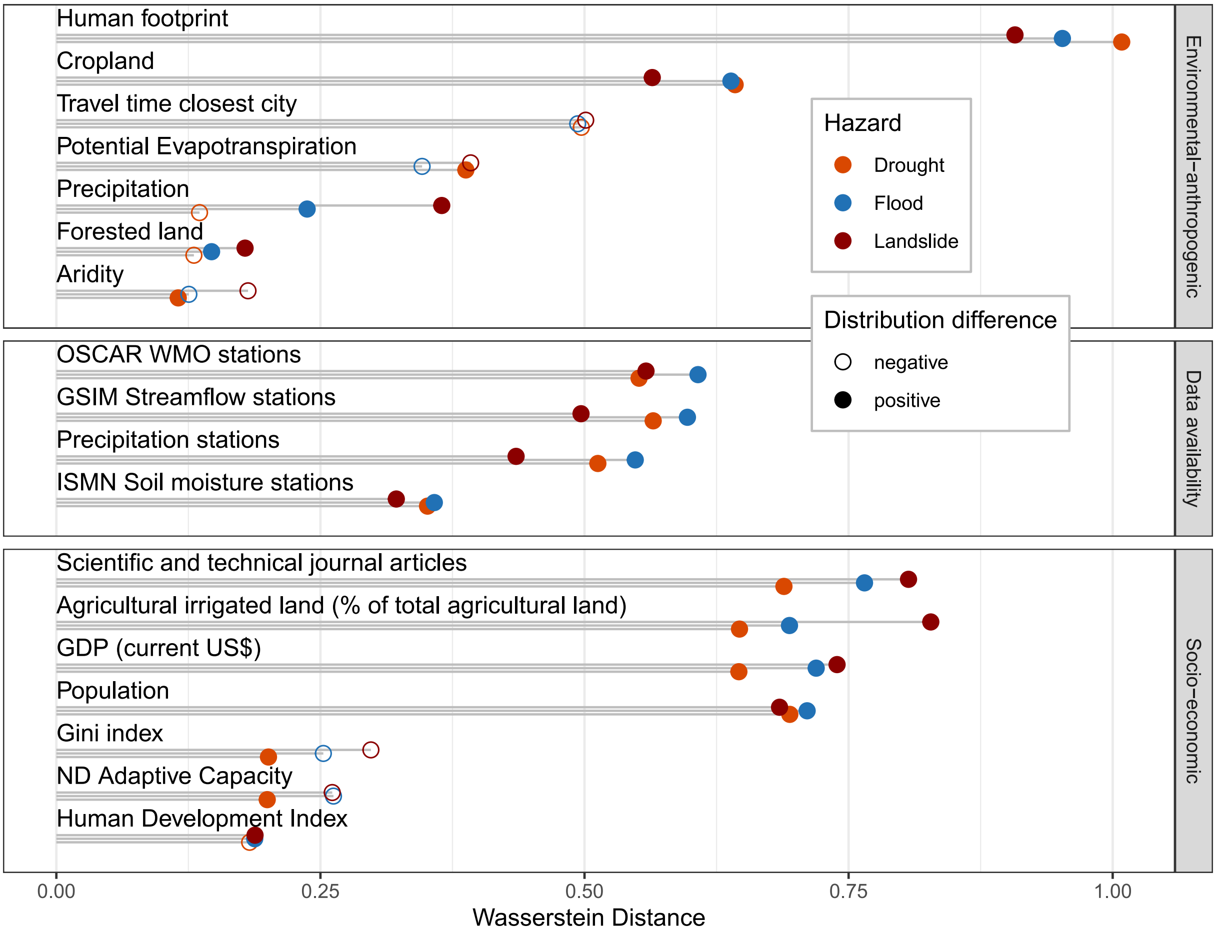

Figure 5. Comparison of climate, land, gauging data, and socio-economic characteristics between regions of high research (>75th quantile) and the entire land area. Distribution difference measured as Wasserstein distance (Krabbenhoft et al., 2022). Higher values indicate a stronger bias. Wasserstein distance only indicates the strength of bias. We infer the direction of bias from the difference between variable distributions (Stein et al., 2021). A positive (negative) distribution difference indicates more (less) research with increasing characteristics.

differences between hazards. While flood and landslide research densities increase with increasing precipitation, drought research density decreases. However, this negative relationship reflects only the average distribution. When examining detailed cumulative distributions (Figure S9 in Supporting Information S1), we observe decreasing research density with increasing precipitation from precipitation values >1250 mm. We also find biases related to data availability, that is, the research density is higher in regions with more measurement stations.

Besides human influence, further biases in hydro-hazard research activity can be found in other socio-economic dimensions. There is a positive bias in research density towards countries with a high gross domestic product (GDP) (Wasserstein distance of 0.65 for drought, 0.72 for flood, and 0.74 for landslides). The variable "Scientific and technical journal articles" from the World Bank refers to the number of articles published within the fields of science and engineering per country. Due to measuring the quantity of research similar to our study, it can be regarded as a control variable that is expected to exhibit a strongly positive value, which we confirm with an average Wasserstein distance of 0.75 across hazards. Research densities are much less biased towards other socio-economic indices than GDP and population. Income inequality (Gini Index), the ability to adapt to climate change, including hazards (adaptive capacity), and the human development index show only small biases (Wasserstein averaged across hazards: 0.25, 0.24, and 0.19, respectively).

3.4. Country Income-Level, People Affected, and Research Density

We investigate the interactions between research density and the number of affected people to analyze whether more impacted regions are also more intensely studied. In Figure 6a, we see that more research is conducted in

10.1029/2024EF004590

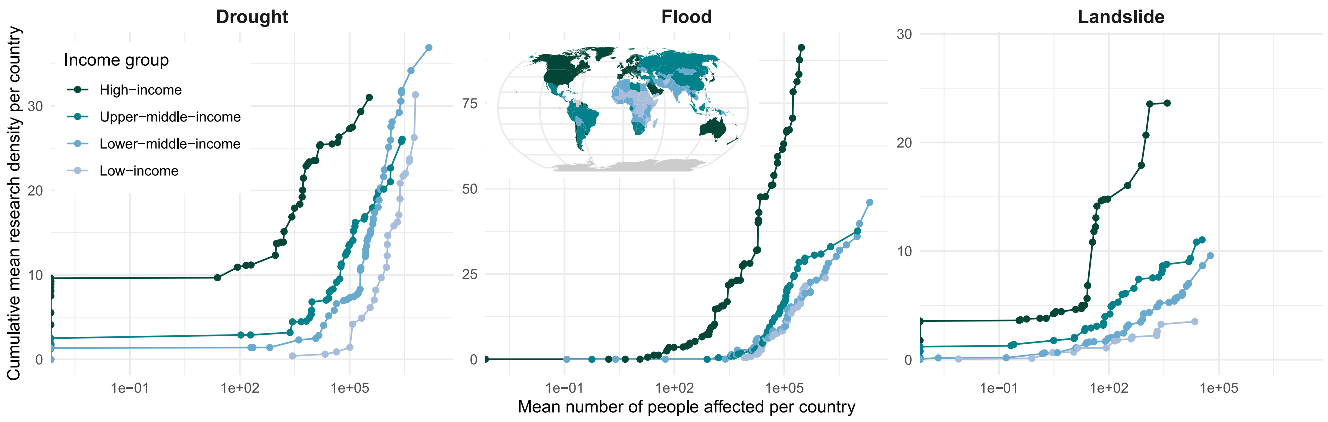

Figure 6. Country-averaged number of affected people against the cumulative distribution of the research density, averaged over all cells per country and separated by World Bank income levels (according to 2021 income classes) (World Bank, 1978). Each dot corresponds to one country.

high-income countries for all hazards, indicated by the higher baseline and earlier onset of the respective curve compared to all other income groups. For some high-income countries (e.g., for droughts in Germany, France, and Japan; or for landslides in the UK, Slovenia, and Uruguay), no people have been recorded as being affected in the EM-DAT database (CRED, 2023a), even though research has been conducted, as indicated by the distribution offset in *y*-direction. There is no visible offset for the distribution of flooding, given that Malta is the only country for which no affected people are recorded. Low, low-middle, and upper-middle-income countries all report higher numbers of people affected for the same research density than high-income countries. However, for nearly all of these countries, hazard research densities never reach the same level as for high-income countries. The only exception is drought research in lower-middle-income countries, which is largely due to the large amount of drought research in India (Figure S13 in Supporting Information S1).

There is a distinct difference in how many people need to be affected before research activity visibly increases for the different income groups. These thresholds are much lower for high-income countries across all hazards. Flood and drought research seems to be triggered when about 100 people are affected in high-income regions, for landslides it is less than 100 people. Flood and drought research activity in low-income countries only starts increasing if more than 10,000 people have been affected. Across all hazards, research density rises with the affected number of people (Figure S15 in Supporting Information S1).

4. Discussion and Conclusion

4.1. Wealth Over Woe—Poorer Countries Are Less Researched Despite Higher Hazard Impact

Low-income countries are disadvantaged across all aspects of disaster risk management. They are already strongly impacted by hydro-hazards (Hallegatte et al., 2020) and by climate change, with accelerating risk in many regions (IPCC, 2022). The need for equality across all aspects of disaster risk management has been recognized by the United Nations Office for Disaster Risk Reduction (UNDRR) and in the Sendai Framework, which aims to increase knowledge and disaster risk reduction with a particular focus on low-income countries (https://www.undrr.org/disaster-risk-reduction-least-developed-countries). Our study can contribute to achieving a more equal and sustainable research landscape, especially when local scientists and communities from target regions are involved in the research (Odeny & Bosurgi, 2022) or are being involved in sustainable research partnerships (Gill et al., 2021). Importantly, addressing these knowledge gaps will help the international community reach the Sustainable Development Goals (SDGs), many of which have synergies with current efforts in disaster risk reduction (Aitsi-Selmi et al., 2016).

Hallegatte et al. (2020) conclude that "Poor people are disproportionately affected by natural hazards and disasters." We find that low-income countries are not just disproportionately affected, but also have a disproportionately lower research density for hydro-hazards. Even though research is more prevalent in all countries where high-impact hazard events occur, the threshold for what constitutes "high" is much lower in wealthier countries (Figure 6). For flood and drought research, 100 times more people need to be affected in low-income countries

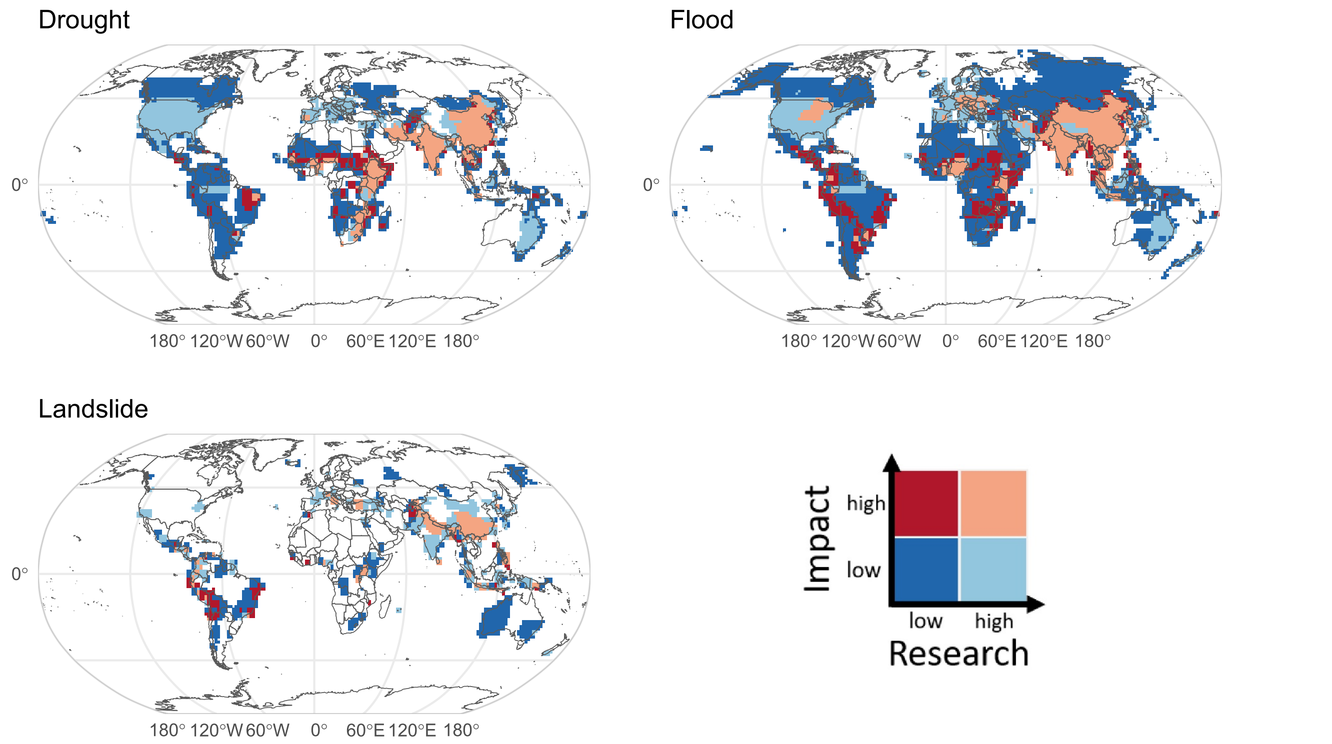

Figure 7. Research focus regions. Each cell is categorized by whether it falls into the high (>75th quantile) or low research category and high or low impact category, based on the number of people affected. Most relevant for future research are regions with low research and high impact (dark red). Classification based on 75th quantile of research and impact (number of people affected, EM-DAT).

compared to high-income countries for research densities to reach the same level. Hazard impact therefore has a relatively small influence on research activity, while country wealth is much more influential (hence Wealth over Woe). This disparity is likely due to highly unequal research funding, data availability, and research capacities between high-income and low-income countries (Skupien & Rüffin, 2020).

Our results show that low-income countries currently need to base risk assessment decisions, adaptation, and policy changes on less research than wealthier countries. Even if research findings can be transferred from hydro-climatically similar regions, socio-economic and governance conditions will most likely be very different (Figure 6). Yet, local scientific and community knowledge is highly relevant for the effectiveness of disaster risk management (Gaillard & Mercer, 2013) and can reduce disaster impact if combined with resources to implement solutions (Kreibich et al., 2022). Less research in low-income countries thus means that there is less knowledge on how the current impact imbalance might be rectified in the future. Global overviews of research distribution, such as ours, can provide valuable guidance by suggesting future research focus regions to international funding agencies including the World Bank, the UN, and the European Union. Or they can guide international research investments of individual nations, like the Global Challenges Research Fund (GCRF) of the UK Research and Innovation non-departmental body of the UK government.

4.2. How Can We Address Current and Future Hydro-Hazard Knowledge Gaps?

We assess research focus regions based on past impact and identify gaps in socio-hydrological variations covered by research activity. For an impact-based assessment, we define regions that should become research focus areas as those with combinations of a high number of people affected (>75th percentile) and low rates of research activity (<75th percentile). For droughts, regions with high research needs are predominantly the Sahel zone, the Horn of Africa, eastern Brazil, and Afghanistan (Figure 7). For floods, the areas are more scattered, but relevant regions are large areas in South and Central America as well as in eastern Africa (e.g., Somalia, Zambia, and Mozambique). In contrast to floods and droughts, which affect multiple spatial grid cells, a single landslide event will only be recorded in one cell due to its limited spatial extent. As a consequence, landslide research focus cells include major cities, for example, Freetown in Sierra Leone and Abidjan in Côte d'Ivoire (Figure 7). Under-researched landslide regions are mainly located in South America, particularly in Bolivia and Brazil. We find that all of the locations mentioned remain research focus regions even when different impact data sets are used. Though with more data, some additional regions can be added as focus regions, as shown and discussed in Supporting Information S1.

Some knowledge gained in highly researched regions may be transferable to less studied regions if similar hydro-climatic and landscape characteristics allow the assumption of process similarity (Bertola et al., 2023; Stein et al., 2021; West et al., 2022). We do find several promising hotspots of highly researched regions where flood, drought, and landslide hazards have been intensely studied. These cover mainly the US, Europe, and parts of Asia. Still, an increase in research will be particularly necessary in regions where increasing hazards and impacts are already noticeable or will likely increase in the future. For example, diminishing water availability in the Southern Hemisphere (Y. Zhang et al., 2023) indicates a need for water management and drought adaptation research, which is currently lacking. Landslide research is predominantly conducted in mountainous and temperate regions in Europe, China, and the USA (Figure 4). Yet, tropical regions, especially tropical cities, have been projected to be future hotspots of landslide risk given both population growth and climate change (Ozturk et al., 2022). While both floods and landslides are well studied in more humid regions, drought research activity is lower in very humid regions and is underrepresented in tropical regions (Figure 4). Hence, we argue that the drought risk for rainforests is likely inadequately studied, despite its importance. For example, recurrent extreme droughts in the sensitive Amazon rainforest (Lewis et al., 2011) define a potential critical tipping point for the earth system (Lenton et al., 2008). Additionally, some poorly explored regions with distinct characteristics, too dissimilar for knowledge transfer, need further exploration from a hazard process understanding viewpoint. A location-specific aspect of risk research is vulnerability since it is dependent on culture, socio-economic settings, and governance systems (King-Okumu et al., 2020). Therefore, it is paramount to ensure vulnerability to hydro-hazards is studied across different socio-hydrological settings.

Wealthier countries also collect and share more data (L. Beck et al., 2008), which further adds to the research bias towards data-rich regions (Figure 5). Some countries, such as the US, are likely highly studied simply because they collect large amounts of data through public funding and then make them freely available. In addition to increased research funding, extended data collection and data sharing are necessary. The Sendai framework and UNDRR are targeting gaps in disaster data (Aitsi-Selmi et al., 2016). However, in addition to disaster information, basic and long-term monitoring of variables such as streamflow, soil moisture, precipitation, etc. are equally necessary to improve hazard research, particularly in periods of strong climate change. Closing the data gap can be achieved by funding targeted extension of monitoring networks (Krabbenhoft et al., 2022), or by collecting and combining available data into systematic databases (e.g., Gerbens-Leenes et al., 2024). The most important point is that the data is made open-access for the most effective use (Aitsi-Selmi et al., 2016).

4.3. Limitations

We have studied the distribution of knowledge within published scientific abstracts as these are the only sources of scientific literature compiled as data sets. Therefore our approach cannot adequately recognize that at least some research might only be accessible through technical reports (i.e., gray literature) or in unpublished Master's and PhD theses. Importantly, we currently do not consider the wealth of knowledge gathered by local citizens and Indigenous people, which is often ignored or overlooked by the scientific community (Chief, 2018), but would require a different type of study to be utilized. Some research might also be overlooked due to the choice of English as the language of analysis. However, Orimoloye et al. (2021) found that 95% of disaster risk management articles are published in English. We therefore assume this limitation to be minor. Similarly, the choice of dictionaries used for geolocation might introduce a bias toward larger entities, high-income countries, and non-natural features (Acheson et al., 2017). We find that this bias did not impact the accuracy of our geolocation (Figure S2 in Supporting Information S1). Our evaluation of 418 abstracts showed, that for 26% of the abstracts, one or more locations were missed. However, the impact of missed and wrongly geolocated locations is small, as in 76% of cases the identified location extent does not change when the missing and wrong locations are added. Additionally, location extraction is biased by the limited description contained within abstracts. Although full-text analysis might have yielded more information (Westergaard et al., 2018), it would dramatically reduce the number of articles available. Fortunately, open access is rapidly growing (Björk, 2017), which means that. Hence, reviews like ours will likely become more informative in the future.

4.4. Looking Forward

In this study, we were able to map hydro-hazard literature and reveal biases related to where and how often hazards are studied in a specific location. We find that high-income countries experience much higher levels of research activity compared to lower-income countries, despite being less affected. Thresholds for numbers of people affected in relation to increased research activity appear to be significantly higher for lower-income countries compared to wealthier regions. Furthermore, the uneven distributions suggest knowledge gaps in hazard understanding since not all relevant hydro-climatic landscapes are covered equally. Where hazard events occur and where they are researched does currently not align. Tropical regions, for example, are studied less than distributions of flood, drought, and landslide events would suggest. Even more importantly, focusing research on high-income regions means that socio-economic and governance structures found in low-income countries are underrepresented. Such biases reveal where future research might be needed to cover a broad spectrum of hazard research across different environmental and socio-economic characteristics. Additionally, regions where many people have been affected by hazards in the past, but where less research has been conducted yet, offer themselves as future study regions and can thus guide research funding efforts. Specifically, Central and South America should receive more attention for flood and landslide research. In Central and Eastern Africa, more drought and flood research should be conducted.

An analysis of this scale would not have been possible without automated tools to analyze text-based data. Large language models and other text-mining tools are increasingly necessary to keep up with the vast amounts of research published (Stein et al., 2022). For comparison, based on the person-hours our manual evaluation took, an on par non-automated study would have taken about 2 years of round-the-clock work for one person to screen all the abstracts in contrast to a few hours of runtime it took us instead (not counting the time it took to develop the approach in the first place). The speed at which text analysis methods are improving will advance opportunities in research analysis. For example, we could add automatically extracted information as “hydrologic” metadata to each article, which could include location, time scale, climate regime, methods used, and more. Research could then easily be found and synthesized along these metadata (Stein et al., 2022). Authors would only need to quality-check the automatic annotations during the submission process, after which their research would immediately be mapped. Beyond search and synthesis, one could additionally generate a training data set to continuously improve and specialize automation tools. Progress in fair research distributions could thus be tracked and local research made visible.

Overall, our findings provide research funding agencies with the necessary maps to develop programs that target research inequality. Policymakers can use these maps to determine where knowledge gaps might affect their decisions. Researchers should be encouraged to develop collaborative networks with and within under-researched regions to build observational and research capacity where it is most needed. Funding agencies need to develop new funding mechanisms to support such efforts, which often fall outside current funding schemes that focus on funding researchers residing in the country of the funding agency, rather than building capacity abroad. We currently only show the state of historical research and its impact to date. However, with climate change altering hazard occurrences around the world and with rapidly changing socio-economic conditions in many places, research relevance shifts as well. If we, as a community, want to preemptively address possible future disasters (Ozturk et al., 2022), we need to map current research activities to highlight knowledge gaps in regions that are at risk in the future.

Data Availability Statement

All data sets used in this study are free and publicly available. A full detailed overview of all data sets used is provided in Supporting Information S1. The results and evaluation data on which this article is based are available in Stein et al. (2024). Due to license restrictions, the Semantic Scholar abstract data cannot be shared directly. However, the Semantic Scholar Academic Graph data set can be accessed via the Semantic Scholar API (Kinney et al., 2023). The created hazard and geo-annotations are made available and can be linked to their respective abstracts using the Semantic Scholar ID. The research density raster grids are part of the data repository.

Open Street Map data was accessed using the Nominatim API (OpenStreetMap, 2023). We use Natural Earth Data (Patterson & Kelso, 2023) accessed via the “rnaturalearth” R package (South, 2017). Impact data is sourced from the Emergency Management Database (CRED, 2023b). Geolocations for EM-Dat were taken from the

Geocoded Disasters (GDIS) Data set (Rosvold & Buhaug, 2021b). Other impact data was sourced from the Dartmouth Flood Observatory (Brakenridge, 2023), the NASA global landslide catalog (Kirschbaum et al., 2010) and the Global Fatal Landslide Database (Froude & Petley, 2018). Measurement station data was taken from the following sources: Precipitation stations—Global Precipitation Climatology Centre (GPCC) (Rustemeier et al., 2022); streamflow stations—Global Streamflow and Metadata Archive (GSIM) (Do et al., 2018); soil moisture stations—International Soil Moisture Network (ISMN) (Dorigo et al., 2011, 2013, 2021); climate stations—WMO Observing Systems Capability Analysis and Review Tool (WMO OSCAR) (World Meteorological Organization (WMO) & Federal Office of Meteorology and Climatology (MeteoSwiss), 2023). Precipitation and evapotranspiration data was taken from CHELSA (Karger et al., 2018). Human footprint data was published here (Venter et al., 2017). The fraction of cropland was taken from the ESA World Cover data set (Zanaga et al., 2021). Data on travel time from the nearest city was published here (Nelson et al., 2019b) and accessed via the "geodata" R package (Hijmans et al., 2023). Socio-economic and other indices were taken from the World Development Indicators and Worldwide Governance Indicators (Kaufmann & Kraay, 2022; World Bank, 2023) accessed via the World Bank Open Data Catalog and "wbstats" R package (Piburn, 2020). Vulnerability and adaptive capacity data were taken from the Notre Dame Global Adaptation Initiative (C. Chen et al., 2015). Population data was taken from WorldPop (2023). We additionally used Human Development Index data (Kummu et al., 2019).

Deep Search is a commercial platform and is available with limited features. The Deep Search Toolkit is a Python Software Development Kit (SDK) and Command Line Interface (CLI) allowing users to interact with the Deep Search platform (Staar et al., 2020). The Deep Search Toolkit codebase is under MIT license. For individual model usage, please refer to the model licenses found in the original packages (https://github.com/DS4SD/deepsearch-toolkit). Wasserstein distance was calculated using the "transport" R package (Schuhmacher et al., 2023). The codes to process, analyze, and plot the data and annotated abstracts are available in Stein et al. (2024).

Acknowledgments

LS and TW acknowledge support from the Alexander von Humboldt Foundation in the framework of the Alexander von Humboldt Professorship endowed by the German Federal Ministry of Education and Research (BMBF). UO acknowledges funding from the research focus point Earth and Environmental Systems of the University of Potsdam. We thank Deva Charan Jarajapu, Nirmal Kularathne, Max Serra Lasierra, and David Strahl for their help in data processing. Furthermore, we would like to thank the editors and two anonymous reviewers who helped improve the article with their suggestions. Open Access funding enabled and organized by Projekt DEAL.

References

Acheson, E., De Sabbata, S., & Purves, R. S. (2017). A quantitative analysis of global gazetteers: Patterns of coverage for common feature types. *Computers, Environment and Urban Systems*, *64*, 309–320. https://doi.org/10.1016/j.compenvurbsys.2017.03.007

Aitsi-Selmi, A., Blanchard, K., & Murray, V. (2016). Ensuring science is useful, useable and used in global disaster risk reduction and sustainable development: A view through the Sendai framework lens. *Palgrave Communications*, *2*(1), 1–9. https://doi.org/10.1057/palcomms.2016.16

Arnell, N. W., Lowe, J. A., Bernie, D., Nicholls, R. J., Brown, S., Challinor, A. J., & Osborn, T. J. (2019). The global and regional impacts of climate change under representative concentration pathway forcings and shared socioeconomic pathway socioeconomic scenarios. *Environmental Research Letters*, *14*(8), 084046. https://doi.org/10.1088/1748-9326/ab35a6

Auer, C., Dolfi, M., Carvalho, A., Berrospi, C., & Staar, P. W. J. (2022). Delivering document conversion as a cloud service with high throughput and responsiveness. In *2022 IEEE 15th International Conference on Cloud Computing (CLOUD)* (pp. 363–373). https://doi.org/10.1109/CLOUD55607.2022.00060

Balana, B. B., Sanfo, S., Barbier, B., Williams, T., & Kolavalli, S. (2019). Assessment of flood recession agriculture for food security in Northern Ghana: An optimization modelling approach. *Agricultural Systems*, *173*, 536–543. https://doi.org/10.1016/j.agsy.2019.03.021

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*(1), 180214. https://doi.org/10.1038/sdata.2018.214

Beck, L., Bernauer, T., & Kalbhenn, A. (2008). Biases in international environmental datasets: Evidence from water quality monitoring in Europe. [SSRN Scholarly Paper] Retrieved form https://papers.ssrn.com/abstract=1321036

Benevolenza, M. A., & DeRigne, L. (2019). The impact of climate change and natural disasters on vulnerable populations: A systematic review of literature. *Journal of Human Behavior in the Social Environment*, *29*(2), 266–281. https://doi.org/10.1080/10911359.2018.1527739

Bennett, J. (2010). *OpenStreetMap*. Packt Publishing Ltd. (Google-Books-ID: SZfqRcPXApoC).

Bertola, M., Blöschl, G., Bohac, M., Borga, M., Castellarin, A., Chirico, G. B., et al. (2023). Megafloods in Europe can be anticipated from observations in hydrologically similar catchments. *Nature Geoscience*, *16*(11), 982–988. https://doi.org/10.1038/s41561-023-01300-5

Björk, B.-C. (2017). Growth of hybrid open access, 2009–2016. *PeerJ*, *5*, e3878. https://doi.org/10.7717/peerj.3878

Brakenridge, G. (2023). Global active archive of large flood events. Dartmouth Flood Observatory, University of Colorado [Dataset]. *University of Colorado, USA: Dartmouth Flood Observatory*. http://floodobservatory.colorado.edu/Archives

Callaghan, M., Schleussner, C.-F., Nath, S., Lejeune, Q., Knutson, T. R., Reichstein, M., et al. (2021). Machine-learning-based evidence and attribution mapping of 100,000 climate impact studies. *Nature Climate Change*, *11*(11), 966–972. https://doi.org/10.1038/s41558-021-01168-6

Chen, C., Noble, I., Hellmann, J., Coffee, J., Murillo, M., & Chawla, N. (2015). University of Notre Dame global adaptation index country index technical report [Dataset]. *ND-GAIN: South Bend, IN, USA*. Retrieved from https://gain.nd.edu/our-work/country-index/

Chief, K. (2018). Emerging voices of tribal perspectives in water resources. *Journal of Contemporary Water Research & Education*, *163*(1), 1–5. https://doi.org/10.1111/j.1936-704X.2018.03266.x

Clemens, K. (2015). Geocoding with openstreetmap data. In *GEOProcessing 2015* (Vol. 10).

CRED. (2023a). *2022 disasters in numbers* (Tech. Rep.). CRED.

CRED. (2023b). EM-DAT [Dataset]. *CRED/UCLouvain, Brussels, Belgium*. Retrieved form www.emdat.be

Devitt, L., Neal, J., Coxon, G., Savage, J., & Wagener, T. (2023). Flood hazard potential reveals global floodplain settlement patterns. *Nature Communications*, *14*(1), 2801. https://doi.org/10.1038/s41467-023-38297-9

Di Marco, M., Chapman, S., Althor, G., Kearney, S., Besancon, C., Butt, N., et al. (2017). Changing trends and persisting biases in three decades of conservation science. *Global Ecology and Conservation*, *10*, 32–42. https://doi.org/10.1016/j.gecco.2017.01.008

Do, H. X., Gudmundsson, L., Leonard, M., & Westra, S. (2018). The Global Streamflow Indices and Metadata Archive (GSIM)—Part 1: The production of a daily streamflow archive and metadata. *Earth System Science Data*, *10*(2), 765–785. https://doi.org/10.5194/essd-10-765-2018

Dorigo, W. A., Himmelbauer, I., Aberer, D., Schremmer, L., Petrakovic, I., Zappa, L., et al. (2021). The international soil moisture network: Serving Earth system science for over a decade [Dataset]. *Hydrology and Earth System Sciences*, *25*(11), 5749–5804. (Publisher: Copernicus GmbH). https://doi.org/10.5194/hess-25-5749-2021

Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., et al. (2011). The international soil moisture network: A data hosting facility for global in situ soil moisture measurements [Dataset]. *Hydrology and Earth System Sciences*, *15*(5), 1675–1698. https://doi.org/10.5194/hess-15-1675-2011

Dorigo, W. A., Xaver, A., Vreugdenhil, M., Gruber, A., Dostálová, A., Sanchis-Dufau, A. D., et al. (2013). Global automated quality control of in situ soil moisture data from the international soil moisture network [Dataset]. *Vadose Zone Journal*, *12*(3). https://doi.org/10.2136/vzj2012.0097

Emmer, A. (2018). Geographies and scientometrics of research on natural hazards. *Geosciences*, *8*(10), 382. https://doi.org/10.3390/geosciences8100382

Froude, M. J., & Petley, D. N. (2018). Global fatal landslide occurrence from 2004 to 2016 [Dataset]. *Natural Hazards and Earth System Sciences*, *18*(8), 2161–2181. https://doi.org/10.5194/nhess-18-2161-2018

Gaillard, J., & Mercer, J. (2013). From knowledge to action: Bridging gaps in disaster risk reduction. *Progress in Human Geography*, *37*(1), 93–114. https://doi.org/10.1177/0309132512446717

Gerbens-Leenes, P. W., Vaca-Jiménez, S. D., Holmatov, B., & Vanham, D. (2024). Spatially distributed freshwater demand for electricity in Africa. *Environmental Sciences: Water Research & Technology*, *10*(8), 1795–1808. https://doi.org/10.1039/D4EW00246F

Gill, J. C., Taylor, F. E., Duncan, M. J., Mohadjer, S., Budimir, M., Mdala, H., & Bukachi, V. (2021). Invited perspectives: Building sustainable and resilient communities—Recommended actions for natural hazard scientists. *Natural Hazards and Earth System Sciences*, *21*(1), 187–202. https://doi.org/10.5194/nhess-21-187-2021

Guha-Sapir, D., & Below, R. (2006). Collecting Data on disasters: Easier said than done. *Asian Disaster Management News*, *12*(2), 9–10.

Haklay, M., & Weber, P. (2008). OpenStreetMap: User-generated street maps. *IEEE Pervasive Computing*, *7*(4), 12–18. (Conference Name: IEEE Pervasive Computing). https://doi.org/10.1109/MPRV.2008.80

Hallegatte, S., Vogt-Schilb, A., Rozenberg, J., Bangalore, M., & Beaudet, C. (2020). From poverty to disaster and back: A review of the literature. *Economics of Disasters and Climate Change*, *4*(1), 223–247. https://doi.org/10.1007/s41885-020-00060-5

Hijmans, R. J., Barbosa, M., Ghosh, A., & Mandel, A. (2023). geodata: Download geographic data [Software]. Retrieved from https://CRAN.R-project.org/package=geodata

Hu, X., Zhou, Z., Li, H., Hu, Y., Gu, F., Kersten, J., et al. (2023). Location reference recognition from texts: A survey and comparison. *ACM Computing Surveys*, *56*(5), 112:1–112:37. https://doi.org/10.1145/3625819

IPCC. (2022). In H.-O. Pörtner, D. C. Roberts, M. Tignor, E. S. Poloczanska, K. Mintenbeck, et al. (Eds.), *Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Jones, R. L., Guha-Sapir, D., & Tubeuf, S. (2022). Human and economic impacts of natural disasters: Can we trust the global data? *Scientific Data*, *9*(1), 572. https://doi.org/10.1038/s41597-022-01667-x

Jongman, B., Ward, P. J., & Aerts, J. C. J. H. (2012). Global exposure to river and coastal flooding: Long term trends and changes. *Global Environmental Change*, *22*(4), 823–835. https://doi.org/10.1016/j.gloenvcha.2012.07.004

Kantorovich, L. V. (1960). Mathematical methods of organizing and planning production. *Management Science*, *6*(4), 366–422. https://doi.org/10.1287/mnsc.6.4.366

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., et al. (2018). Data from: Climatologies at high resolution for the earth's land surface areas [Dataset]. *Dryad*. https://doi.org/10.5061/DRYAD.KD1D4

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., et al. (2017). Climatologies at high resolution for the Earth's land surface areas. *Scientific Data*, *4*(1), 170122. https://doi.org/10.1038/sdata.2017.122

Kaufmann, D., & Kraay, A. (2022). Worldwide governance indicators [Dataset]. Retrieved from www.govindicators.org

King-Okumu, C., Tsegai, D., Pandey, R. P., & Rees, G. (2020). Less to lose? Drought impact and vulnerability assessment in disadvantaged regions. *Water*, *12*(4), 1136. https://doi.org/10.3390/w12041136

Kinney, R., Anastasiades, C., Authur, R., Beltagy, I., Bragg, J., Buraczynski, A., et al. (2023). The Semantic Scholar open data platform [Software]. https://doi.org/10.48550/arXiv.2301.10140

Kirschbaum, D. B., Adler, R., Hong, Y., Hill, S., & Lerner-Lam, A. (2010). A global landslide catalog for hazard applications: Method, results, and limitations [Dataset]. *Natural Hazards*, *52*(3), 561–575. https://doi.org/10.1007/s11069-009-9401-4

Krabbenhoft, C. A., Allen, G. H., Lin, P., Godsey, S. E., Allen, D. C., Burrows, R. M., et al. (2022). Assessing placement bias of the global river gauge network. *Nature Sustainability*, *5*, 1–7. https://doi.org/10.1038/s41893-022-00873-0

Kreibich, H., Van Loon, A. F., Schröter, K., Ward, P. J., Mazzoleni, M., Sairam, N., et al. (2022). The challenge of unprecedented floods and droughts in risk management. *Nature*, *608*(7921), 80–86. https://doi.org/10.1038/s41586-022-04917-5

Kummu, M., Taka, M., & Guillaume, J. H. A. (2018). Gridded global datasets for gross domestic product and human development index over 1990–2015. *Scientific Data*, *5*(1), 180004. https://doi.org/10.1038/sdata.2018.4

Kummu, M., Taka, M., & Guillaume, J. H. A. (2019). Data from: Gridded global datasets for gross domestic product and human development index over 1990–2015 [Dataset]. *Dryad*. https://doi.org/10.5061/DRYAD.DK1J0

Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., et al. (2015). DBpedia—A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web*, *6*(2), 167–195. https://doi.org/10.3233/SW-140134

Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., & Schellnhuber, H. J. (2008). Tipping elements in the Earth's climate system. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(6), 1786–1793. https://doi.org/10.1073/pnas.0705414105

Lewis, S. L., Brando, P. M., Phillips, O. L., van der Heijden, G. M. F., & Nepstad, D. (2011). The 2010 Amazon drought. *Science*, *331*(6017), 554. https://doi.org/10.1126/science.1200807

Lindersson, S., Brandimarte, L., Mård, J., & Di Baldassarre, G. (2020). A review of freely accessible global datasets for the study of floods, droughts and their interactions with human societies. *WIREs Water*, *7*(3), e1424. https://doi.org/10.1002/wat2.1424

Mwampamba, T. H., Egoh, B. N., Borokini, I., & Njabo, K. (2022). Challenges encountered when doing research back home: Perspectives from African conservation scientists in the diaspora. *Conservation Science and Practice*, *4*(5), e564. https://doi.org/10.1111/csp2.564

Nelson, A., Weiss, D. J., van Etten, J., Cattaneo, A., McMenomy, T. S., & Koo, J. (2019a). A suite of global accessibility indicators. *Scientific Data*, *6*(1), 266. https://doi.org/10.1038/s41597-019-0265-5

Nelson, A., Weiss, D. J., van Etten, J., Cattaneo, A., McMenomy, T. S., & Koo, J. (2019b). Travel time to cities and ports in the year 2015 [Dataset]. *figshare*. https://doi.org/10.6084/m9.figshare.7638134.v3

North, M. A., Hastie, W. W., & Hoyer, L. (2020). Out of Africa: The underrepresentation of African authors in high-impact geoscience literature. *Earth-Science Reviews*, *208*, 103262. https://doi.org/10.1016/j.earscirev.2020.103262

Odeny, B., & Bosurgi, R. (2022). Time to end parachute science. *PLoS Medicine*, *19*(9), e1004099. https://doi.org/10.1371/journal.pmed.1004099

OpenStreetMap. (2023). [Software]. Retrieved from nominatim.org

Orimoloye, I. R., Ekundayo, T. C., Ololade, O. O., & Belle, J. A. (2021). Systematic mapping of disaster risk management research and the role of innovative technology. *Environmental Science and Pollution Research*, *28*(4), 4289–4306. https://doi.org/10.1007/s11356-020-10791-3

Overland, I., Fossum Sagbakken, H., Isataeva, A., Kolodzinskaia, G., Simpson, N. P., Trisos, C., & Vakulchuk, R. (2022). Funding flows for climate change research on Africa: Where do they come from and where do they go? *Climate & Development*, *14*(8), 705–724. https://doi.org/10.1080/17565529.2021.1976609

Ozturk, U., Bozzolan, E., Holcombe, E. A., Shukla, R., Pianosi, F., & Wagener, T. (2022). How climate change and unplanned urban sprawl bring more landslides. *Nature*, *608*(7922), 262–265. https://doi.org/10.1038/d41586-022-02141-9

Patterson, T., & Kelso, N. V. (2023). Natural Earth data [Dataset]. Retrieved from https://www.naturalearthdata.com/downloads/110m-physical-vectors/

Piburn, J. (2020). wbstats: Programmatic access to the World Bank API [Software]. *Oak Ridge, Tennessee: Oak Ridge National Laboratory*. https://doi.org/10.11578/dc.20171025.1827

Pyzer-Knapp, E. O., Pitera, J. W., Staar, P. W. J., Takeda, S., Laino, T., Sanders, D. P., et al. (2022). Accelerating materials discovery using artificial intelligence, high performance computing and robotics. *Npj Computational Materials*, *8*(1), 1–9. https://doi.org/10.1038/s41524-022-00765-z

Rosvold, E., & Buhaug, H. (2021a). GDIS, a global dataset of geocoded disaster locations. *Scientific Data*, *8*(1), 61. https://doi.org/10.1038/s41597-021-00846-6

Rosvold, E., & Buhaug, H. (2021b). Geocoded disasters (GDIS) dataset [Dataset]. *NASA Socioeconomic Data and Applications Center (SEDAC)*. https://doi.org/10.7927/ZZ3B-8Y61

Rustemeier, E., Hänsel, S., Finger, P., Schneider, U., & Ziese, M. (2022). GPCC Climatology version 2022 at 2.5°: Monthly land-surface precipitation climatology for every month and the total year from Rain-Gauges built on GTS-based and historical data [Dataset]. https://doi.org/10.5676/DWD_GPCC/CLIM_M_V2022_250

Schuhmacher, D., Bähre, B., Gottschlich, C., Hartmann, V., Heinemann, F., & Schmitzer, B. (2023). transport: Computation of optimal transport plans and Wasserstein distances [Software]. Retrieved from https://cran.r-project.org/package=transport

Singh, D. (2006). Publication bias—A reason for the decreased research output in developing countries. *African Journal of Psychiatry*, *9*(3), 153–155. https://doi.org/10.4314/ajpsy.v9i3.30216

Skupien, S., & Rüffin, N. (2020). The geography of research funding: Semantics and beyond. *Journal of Studies in International Education*, *24*(1), 24–38. https://doi.org/10.1177/1028315319889896

South, A. (2017). rnaturalearth: World map data from natural [Software]. *Earth*. Retrieved from https://CRAN.R-project.org/package=rnaturalearth

Staar, P. W. J., Dolfi, M., & Auer, C. (2020). Corpus processing service: A knowledge graph platform to perform deep data exploration on corpora [Software]. *Applied AI Letters*, *1*(2), e20. https://doi.org/10.1002/ail2.20

Stein, L., Clark, M. P., Knoben, W. J. M., Pianosi, F., & Woods, R. A. (2021). How do climate and catchment attributes influence flood generating processes? A large-Sample study for 671 catchments across the contiguous USA. *Water Resources Research*, *57*(4), e2020WR028300. https://doi.org/10.1029/2020WR028300

Stein, L., Mukkavilli, S. K., Pfitzmann, B. M., Staar, P. W. J., Ozturk, U., Berrospi, C., et al. (2024). Supplementary data and code "Wealth over Woe: Global biases in hydro-hazard research" [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.10490256

Stein, L., Mukkavilli, S. K., & Wagener, T. (2022). Lifelines for a drowning science—Improving findability and synthesis of hydrologic publications. *Hydrological Processes*, *36*(11), e14742. https://doi.org/10.1002/hyp.14742

Sumathipala, A., Siribaddana, S., & Patel, V. (2004). Under-representation of developing countries in the research literature: Ethical issues arising from a survey of five leading medical journals. *BMC Medical Ethics*, *5*(1), 5. https://doi.org/10.1186/1472-6939-5-5

Ting, K. M. (2010). Precision and recall. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning* (pp. 781). Springer US. https://doi.org/10.1007/978-0-387-30164-8_652

UNDRR. (2021). GAR special report on drought 2021. Retrieved from http://www.undrr.org/publication/gar-special-report-drought-2021

Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., et al. (2016). Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nature Communications*, *7*(1), 12558. https://doi.org/10.1038/ncomms12558

Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., et al. (2017). Data from: Global terrestrial human footprint maps for 1993 and 2009 [Dataset]. *Dryad*. https://doi.org/10.5061/DRYAD.052Q5

Ward, P. J., Blauhut, V., Bloemendaal, N., Daniell, J. E., de Ruiter, M. C., Duncan, M. J., et al. (2020). Review article: Natural hazard risk assessments at the global scale. *Natural Hazards and Earth System Sciences*, *20*(4), 1069–1096. https://doi.org/10.5194/nhess-20-1069-2020

West, C., Rosolem, R., MacDonald, A. M., Cuthbert, M. O., & Wagener, T. (2022). Understanding process controls on groundwater recharge variability across Africa through recharge landscapes. *Journal of Hydrology*, *612*, 127967. https://doi.org/10.1016/j.jhydrol.2022.127967

Westergaard, D., Stærfeldt, H.-H., Tønsberg, C., Jensen, L. J., & Brunak, S. (2018). A comprehensive and quantitative comparison of text-mining in 15 million full-text articles versus their corresponding abstracts. *PLoS Computational Biology*, *14*(2), e1005962. https://doi.org/10.1371/journal.pcbi.1005962

Woelbert, E., Lundell-Smith, K., White, R., & Kemmer, D. (2021). Accounting for mental health research funding: Developing a quantitative baseline of global investments. *The Lancet Psychiatry*, *8*(3), 250–258. https://doi.org/10.1016/S2215-0366(20)30469-7

World Bank. (1978). *World development report 1978 (World Development Report No. 1)*. World Bank.

World Bank. (2023). World development indicators [Dataset]. Retrieved from https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators

World Meteorological Organization (WMO), & Federal Office of Meteorology and Climatology (MeteoSwiss). (2023). WMO integrated global observing system (OSCAR) [Dataset]. Retrieved from https://oscar.wmo.int/surface//index.html#/

WorldPop. (2023). WorldPop (School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur) and Center for International Earth Science

Information Network (CIESIN), Columbia University (2018). Global High Resolution Population Denominators Project—Funded by The Bill and Melinda Gates Foundation (OPP1134076) [Dataset]. https://doi.org/10.5258/SOTON/WP00647

Wyatt, F., Robbins, J., & Beckett, R. (2023). Investigating bias in impact observation sources and implications for impact-based forecast evaluation. *International Journal of Disaster Risk Reduction*, *90*, 103639. https://doi.org/10.1016/j.ijdrr.2023.103639

Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., et al. (2021). ESA WorldCover 10 m 2020 v100 [Dataset]. *Zenodo*. https://doi.org/10.5281/ZENODO.5571936

Zhang, Y., Li, C., Chiew, F. H. S., Post, D. A., Zhang, X., Ma, N., et al. (2023). Southern Hemisphere dominates recent decline in global water availability. *Science*, *382*(6670), 579–584. https://doi.org/10.1126/science.adh0716

References From the Supporting Information

Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J.-C., Fritz, N., Froissard, F., et al. (2008). From near-surface to root-zone soil moisture using an exponential filter: An assessment of the method based on in-situ observations and model simulations. *Hydrology and Earth System Sciences*, *12*(6), 1323–1337. https://doi.org/10.5194/hess-12-1323-2008

Alday, J. G., Camarero, J. J., Revilla, J., & Resco de Dios, V. (2020). Similar diurnal, seasonal and annual rhythms in radial root expansion across two coexisting Mediterranean oak species. *Tree Physiology*, *40*(7), 956–968. https://doi.org/10.1093/treephys/tpaa041

Al-Yaari, A., Dayau, S., Chipeaux, C., Aluome, C., Kruszewski, A., Loustau, D., & Wigneron, J.-P. (2018). The AQUI soil moisture network for satellite microwave remote sensing validation in South-western France. *Remote Sensing*, *10*(11), 1839. https://doi.org/10.3390/rs10111839

Ardö, J. (2013). A 10-year dataset of basic meteorology and soil properties in Central Sudan [Dataset]. *Dataset Papers in Geosciences*, *2013*, 1–6. https://doi.org/10.7167/2013/297973/dataset

Bell, J., Palecki, M., Baker, B., Collins, W., Lawrimore, J., Leeper, R., et al. (2013). U.S. Climate reference network soil moisture and temperature observations. *Journal of Hydrometeorology*, *14*(3), 977–988. https://doi.org/10.1175/JHM-D-12-0146.1

Beyrich, F., & Adam, W. (2007). Site and Data Report for the Lindenberg Reference Site in CEOP—Phase 1, Berichte des Deutschen Wetterdienstes, 230, Offenbach am Main, 2007 (Tech. Rep.).

Biddoccu, M., Ferraris, S., Opsi, F., & Cavallo, E. (2016). Long-term monitoring of soil management effects on runoff and soil erosion in sloping vineyards in Alto Monferrato (North West Italy). *Soil and Tillage Research*, *155*, 176–189. https://doi.org/10.1016/j.still.2015.07.005

Bircher, S., Skou, N., Jensen, K., Walker, J., & Rasmussen, L. (2012). A soil moisture and temperature network for SMOS validation in Western Denmark. *Hydrology and Earth System Sciences*, *16*(5), 1445–1463. https://doi.org/10.5194/hess-16-1445-2012

Blöschl, G., Blaschke, A. P., Broer, M., Bucher, C., Carr, G., Chen, X., et al. (2016). The hydrological open air laboratory (HOAL) in Petzenkirchen: A hypothesis-driven observatory. *Hydrology and Earth System Sciences*, *20*(1), 227–255. https://doi.org/10.5194/hess-20-227-2016

Bogena, H., Kunkel, R., Pütz, T., Vereecken, H., Kruger, E., Zacharias, S., et al. (2012). TERENO—long-term monitoring network for terrestrial environmental research. *Hydrologie und Wasserbewirtschaftung*, *56*, 138–143.

Bogena, H., Montzka, C., Huisman, J., Graf, A., Schmidt, M., Stockinger, M., et al. (2018). The TERENO-Rur hydrological observatory: A Multiscale multi-Compartment research platform for the advancement of hydrological science. *Vadose Zone Journal*, *17*(1), 1–22. https://doi.org/10.2136/vzj2018.03.0055

Bogena, H. R. (2016). TERENO: German network of terrestrial environmental observatories. *Journal of large-scale research facilities JLSRF*, *2*, A52. https://doi.org/10.17815/jlsrf-2-98

Brocca, L., Melone, F., & Moramarco, T. (2008). On the estimation of antecedent wetness condition in rainfall-runoff modeling. *Hydrological Processes*, *22*(5), 629–642. https://doi.org/10.1002/hyp.6629

Brocca, L., Melone, F., Moramarco, T., & Morbidelli, R. (2009). Antecedent wetness conditions based on ERS scatterometer data. *Journal of Hydrology*, *364*(1–2), 73–87. https://doi.org/10.1016/j.jhydrol.2008.10.007

Brocca, L., Hasenauer, S., Lacava, T., Melone, F., Moramarco, T., Wagner, W., et al. (2011). Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe. *Remote Sensing of Environment*, *115*(12), 3390–3408. https://doi.org/10.1016/j.rse.2011.08.003

Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., & Piguet, B. (2007). In situ soil moisture observations for the CAL/VAL of SMOS: The SMOSMANIA network. In *2007 IEEE International Geoscience and Remote Sensing Symposium* (pp. 1196–1199). https://doi.org/10.1109/IGARSS.2007.4423019

Calvet, J.-C., Fritz, N., Berne, C., Piguet, B., Maurel, W., & Meurey, C. (2016). Deriving pedotransfer functions for soil quartz fraction in southern France from reverse modeling. *Soil*, *2*(4), 615–629. https://doi.org/10.5194/soil-2-615-2016

Canisius, F. (2011). *Calibration of Casselman, Ontario soil moisture monitoring network* (Tech. Rep.) (p. 37). Agriculture and Agri-Food Canada.

Capello, G., Biddoccu, M., Ferraris, S., & Cavallo, E. (2019). Effects of tractor passes on hydrological and soil erosion processes in tilled and grassed vineyards. *Water*, *11*(10), 2118. https://doi.org/10.3390/w11102118

Cappelaere, B., Descroix, L., Lebel, T., Boulain, N., Ramier, D., Laurent, J.-P., et al. (2009). The AMMA-CATCH experiment in the cultivated Sahelian area of south-west Niger, Investigating water cycle response to a fluctuating climate and changing environment. *Journal of Hydrology*, *375*(1–2), 34–51. https://doi.org/10.1016/j.jhydrol.2009.06.021

Chen, N., Xiao, C., Pu, F., Wang, X., Wang, C., Wang, Z., & Gong, J. (2015). Cyber-physical geographical information service-enabled control of diverse in-situ sensors. *Sensors*, *15*(2), 2565–2592. https://doi.org/10.3390/s150202565

Chen, N., Zhang, X., & Wang, C. (2015). Integrated open geospatial web service enabled cyber-physical information infrastructure for precision agriculture monitoring. *Computers and Electronics in Agriculture*, *111*, 78–91. https://doi.org/10.1016/j.compag.2014.12.009

Cook, D., & Sullivan, R. (2018). Surface Energy Balance system (SEBS) Instrument Handbook. In *U.S. Department of Energy, atmospheric radiation measurement user facility, Richland, Washington, DOE/SC- ARM-TR-092*. https://doi.org/10.2172/1004944

Cook, D. R. (2016). Soil temperature and moisture profile (STAMP) system Handbook. In *U.S. Department of Energy, atmospheric radiation measurement user facility, Richland, Washington, DOE/SC-ARM-TR-186*. https://doi.org/10.2172/1332724

Darouich, H., Ramos, T. B., Pereira, L. S., Rabino, D., Bagagiolo, G., Capello, G., et al. (2022). Water use and soil water balance of Mediterranean vineyards under rainfed and drip irrigation management: Evapotranspiration partition and soil management modelling for resource conservation. *Water*, *14*(4), 554. (Publisher: MDPI). https://doi.org/10.3390/w14040554

Dente, L., Su, Z., & Wen, J. (2012). Validation of SMOS soil moisture products over the Maqu and Twente regions. *Sensors*, *12*(8), 9965–9986. https://doi.org/10.3390/s120809965

Flammini, A., Corradini, C., Morbidelli, R., Saltalippi, C., Picciafuoco, T., & Giráldez, J. V. (2018). Experimental analyses of the evaporation dynamics in bare soils under natural conditions. *Water Resources Management*, *32*(3), 1153–1166. (Publisher: Springer). https://doi.org/10.1007/s11269-017-1860-x

Flammini, A., Morbidelli, R., Saltalippi, C., Picciafuoco, T., Corradini, C., & Govindaraju, R. S. (2018). Reassessment of a semi-analytical field-scale infiltration model through experiments under natural rainfall events. *Journal of Hydrology*, *565*, 835–845. (Publisher: Elsevier). https://doi.org/10.1016/j.jhydrol.2018.08.073

Fuchsberger, J., Kirchengast, G., & Kabas, T. (2021). WegenerNet high-resolution weather and climate data from 2007 to 2020. *Earth System Science Data*, *13*(3), 1307–1334. https://doi.org/10.5194/essd-13-1307-2021

Galle, S., Grippa, M., Peugeot, C., Bouzou Moussa, I., Cappelaere, B., Demarty, J., et al. (2015). AMMA-CATCH a hydrological, meteorological and ecological long term observatory on West Africa: Some recent results. In *AGU Fall Meeting Abstracts* (Vol. 2015, p. GC42A–01).

González-Zamora, Á., Sánchez, N., Pablos, M., & Martínez-Fernández, J. (2019). CCI soil moisture assessment with SMOS soil moisture and in situ data under different environmental conditions and spatial scales in Spain. *Remote Sensing of Environment*, *225*, 469–482. https://doi.org/10.1016/j.rse.2018.02.010

Hajdu, I., Yule, I., Bretherton, M., Singh, R., & Hedley, C. (2019). Field performance assessment and calibration of multi-depth AquaCheck capacitance-based soil moisture probes under permanent pasture for hill country soils. *Agricultural Water Management*, *217*, 332–345. https://doi.org/10.1016/j.agwat.2019.03.002

Hollinger, S., & Isard, S. (1994). A soil moisture climatology of Illinois. *Journal of Climate*, *7*(5), 822–833. https://doi.org/10.1175/1520-0442(1994)007⟨0822:ASMCOI⟩2.0.CO;2

Ikonen, J., Vehviläinen, J., Rautiainen, K., Smolander, T., Lemmetyinen, J., Bircher, S., & Pulliainen, J. (2016). The Sodankylä in-situ soil moisture observation network: An example application to Earth observation data product evaluation. *Geoscientific Instrumentation, Methods and Data Systems*, *5*(1), 95–108. https://doi.org/10.5194/gi-5-95-2016

Ikonen, J., Smolander, T., Rautiainen, K., Cohen, J., Lemmetyinen, J., Salminen, M., & Pulliainen, J. (2018). Spatially distributed evaluation of ESA CCI Soil Moisture products in a northern boreal forest environment. *Geosciences*, *8*(2), 51. https://doi.org/10.3390/geosciences8020051

Jackson, T., Cosh, M., Bindlish, R., Starks, P., Bosch, D., Seyfried, M., et al. (2011). Validation of advanced microwave scanning radiometer soil moisture products. In *Geoscience and Remote Sensing, IEEE Transactions on* (Vol. 48(12), pp. 4256–4272). https://doi.org/10.1109/TGRS.2010.2051035

Jensen, K. H., & Refsgaard, J. C. (2018). HOBE: The Danish hydrological observatory. *Vadose Zone Journal*, *17*(1), 1–24. https://doi.org/10.2136/vzj2018.03.0059

Jin, R., Li, X., Yan, B., Li, X., Luo, W., Ma, M., et al. (2014). A nested ecohydrological wireless sensor network for capturing the surface heterogeneity in the Midstream areas of the Heihe River Basin, China. *IEEE Geoscience and Remote Sensing Letters*, *11*(11), 2015–2019. https://doi.org/10.1109/LGRS.2014.2319085

Kang, J., Li, X., Jin, R., Ge, Y., Wang, J., & Wang, J. (2014). Hybrid optimal design of the eco-hydrological wireless sensor network in the middle reach of the Heihe River Basin, China. *Sensors*, *14*(10), 19095–19114. https://doi.org/10.3390/s141019095

Kang, C. S., Kanniah, K. D., & Kerr, Y. H. (2019). Calibration of SMOS soil moisture retrieval algorithm: A case of tropical site in Malaysia. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(6), 3827–3839. (Publisher: IEEE). https://doi.org/10.1109/tgrs.2018.2888535

Kirchengast, G., Kabas, T., Leuprecht, A., Bichler, C., & Truhetz, H. (2014). WegenerNet: A pioneering high-resolution network for monitoring weather and climate. *Bulletin of the American Meteorological Society*, *95*(2), 227–242. https://doi.org/10.1175/BAMS-D-11-00161.1

Larson, K., Small, E., Gutmann, E., Bilich, A., Braun, J., Zavorotny, V., & Larson, C. (2008). Use of GPS receivers as a soil moisture network for water cycle studies. *Geophysical Research Letters*, *35*(24). https://doi.org/10.1029/2008GL036013

Leavesley, G., David, O., Garen, D., Lea, J., Marron, J., Perkins, T., & Strobel, M. (2010). A modelling framework for improved agricultural water-supply forecasting (Tech. Rep.).

Lebel, T., Cappelaere, B., Galle, S., Hanan, N., Kergoat, L., Levis, S., et al. (2009). AMMA-CATCH studies in the Sahelian region of west-Africa: An overview. *Journal of Hydrology*, *375*(1–2), 3–13. https://doi.org/10.1016/j.jhydrol.2009.03.020

Liu, S., Mo, X., Li, H., Peng, G., & Robock, A. (2001). Spatial variation of soil moisture in China: Geostatistical characterization. *Journal of The Meteorological Society of Japan*, *79*(1B), 555–574. https://doi.org/10.2151/jmsj.79.555

Loew, A., Dall'Amico, J. T., Schlenz, F., & Mauser, W. (2009). The upper danube soil moisture validation site: Measurements and activities. In H. Lacoste (Ed.), *Earth observation and water cycle science* (Vol. 674, p. 56).

Marczewski, W., Slominski, J., Slominska, E., Usowicz, B., Usowicz, J., S, R., et al. (2010). Strategies for validating and directions for employing SMOS data, in the Cal-Val project SWEX (3275) for wetlands. *Hydrology and Earth System Sciences Discussions*, *7*. https://doi.org/10.5194/hessd-7-7007-2010

Mattar, C., Santamaría-Artigas, A., Durán-Alarcón, C., Olivera-Guerra, L., & Fuster, R. (2014). LAB-net the first Chilean soil moisture network for remote sensing applications. In *Quantitative Remote Sensing Symposium (RAQRS)* (pp. 22–26).

Mattar, C., Santamaría-Artigas, A., Durán-Alarcón, C., Olivera-Guerra, L., Fuster, R., & Borvarán, D. (2016). The LAB-net soil moisture network: Application to Thermal remote sensing and surface energy balance. *Data*, *1*(1), 6. https://doi.org/10.3390/data1010006

Moghaddam, M., Entekhabi, D., Goykhman, Y., Li, K., Liu, M., Mahajan, A., et al. (2011). A wireless soil moisture smart sensor web using physics-based optimal control: Concept and initial demonstrations. Selected topics in applied Earth observations and remote sensing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *3*(4), 522–535. https://doi.org/10.1109/JSTARS.2010.2052918

Moghaddam, M., Silva, A., Clewely, D., Akbar, R., Hussaini, S., Whitcomb, J., et al. (2016). Soil moisture profiles and temperature data from SoilSCAPE sites, USA [Dataset]. https://doi.org/10.3334/ORNLDAAC/1339

Morbidelli, R., Corradini, C., Saltalippi, C., Flammini, A., & Rossi, E. (2011). Infiltration-soil moisture redistribution under natural conditions: Experimental evidence as a guideline for realizing simulation models. *Hydrology and Earth System Sciences*, *15*(9), 2937–2945. (Publisher: Copernicus GmbH). https://doi.org/10.5194/hess-15-2937-2011

Morbidelli, R., Saltalippi, C., Flammini, A., Rossi, E., & Corradini, C. (2014). Soil water content vertical profiles under natural conditions: Matching of experiments and simulations by a conceptual model. *Hydrological Processes*, *28*(17), 4732–4742. (Publisher: Wiley Online Library). https://doi.org/10.1002/hyp.9973

Morbidelli, R., Saltalippi, C., Flammini, A., Cifrodelli, M., Picciafuoco, T., Corradini, C., & Govindaraju, R. S. (2017). In situ measurements of soil saturated hydraulic conductivity: Assessment of reliability through rainfall–runoff experiments. *Hydrological Processes*, *31*(17), 3084–3094. https://doi.org/10.1002/hyp.11247

Mougin, E., Hiernaux, P., Kergoat, L., Manuela, G., Rosnay, P., Timouk, F., et al. (2009). The AMMA-CATCH Gourma observatory site in Mali: Relating climatic variations to changes in vegetation, surface hydrology, fluxes and natural resources. *Journal of Hydrology*, *375*(1–2), 14–33. https://doi.org/10.1016/j.jhydrol.2009.06.045

Musial, J. P., Dabrowska-Zielinska, K., Kiryla, W., Oleszczuk, R., Gnatowski, T., & Jaszczynski, J. (2016). Derivation and validation of the high resolution satellite soil moisture products: A case study of the Biebrza Sentinel-1 validation sites. *Geoinformation Issues*, *8*(1), 37–53.

Nguyen, H. H., Kim, H., & Choi, M. (2017). Evaluation of the soil water content using cosmic-ray neutron probe in a heterogeneous monsoon climate-dominated region. *Advances in Water Resources*, *108*, 125–138. https://doi.org/10.1016/j.advwatres.2017.07.020

Ojo, E. R., Bullock, P., L. Heureux, J., Powers, J., McNairn, H., & Pacheco, A. (2015). Calibration and evaluation of a frequency domain reflectometry sensor for real-time soil moisture monitoring. *Vadose Zone Journal*, *14*(3). https://doi.org/10.2136/vzj2014.08.0114

Osenga, E. C., Arnott, J. C., Endsley, K. A., & Katzenberger, J. W. (2019). Bioclimatic and soil moisture monitoring across elevation in a mountain watershed: Opportunities for research and resource management. *Water Resources Research*, *55*(3), 2493–2503. https://doi.org/10.1029/2018WR023653

Osenga, E. C., Vano, J. A., & Arnott, J. C. (2021). A community-supported weather and soil moisture monitoring database of the Roaring Fork catchment of the Colorado River Headwaters. *Hydrological Processes*, *35*(3), e14081. (Publisher: Wiley Online Library). https://doi.org/10.1002/hyp.14081

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, *10*(1), 89. https://doi.org/10.1186/s13643-021-01626-4

Peischl, S., Walker, J., Rüdiger, C., Ye, N., Kerr, Y., Kim, E., et al. (2012). The AACES field experiments: SMOS calibration and validation across the Murrumbidgee River catchment. *Hydrology and Earth System Sciences Discussions*, *16*(6), 1697–1708. https://doi.org/10.5194/hess-16-1697-2012

Pellarin, T., Laurent, J.-P., Cappelaere, B., Decharme, B., Descroix, L., & Ramier, D. (2009). Hydrological modelling and associated microwave emission of a semi-arid region in South-western Niger. *Journal of Hydrology*, *375*(1–2), 262–272. https://doi.org/10.1016/j.jhydrol.2008.12.003

Petropoulos, G. P., & McCalmont, J. P. (2017). An operational in situ soil moisture & soil temperature monitoring network for West Wales, UK: The WSMN network. *Sensors*, *17*(7), 1481. (Publisher: Multidisciplinary Digital Publishing Institute). https://doi.org/10.3390/s17071481

Raffelli, G., Previati, M., Canone, D., Gisolo, D., Bevilacqua, I., Capello, G., et al. (2017). Local- and plot-scale measurements of soil moisture: Time and spatially resolved field techniques in plain, hill and mountain sites. *Water*, *9*(9), 706. https://doi.org/10.3390/w9090706

Robock, A., Vinnikov, K., Srinivasan, G., Entin, J., Hollinger, S., Speranskaya, N., et al. (2000). The global soil moisture data bank. *Bulletin of the American Meteorological Society*, *81*(6), 1281–1300. https://doi.org/10.1175/1520-0477(2000)081⟨1281:TGSMDB⟩2.3.CO;2

Rosnay, P., Gruhier, C., Timouk, F., Baup, F., Mougin, E., Hiernaux, P., et al. (2009). Multi-scale soil moisture measurements at the Gourma meso-scale site in Mali. *Journal of Hydrology*, *375*(1–2), 241–252. https://doi.org/10.1016/j.jhydrol.2009.01.015

Rüdiger, C., Hancock, G., Hemakumara, H., Jacobs, B., Kalma, J., Martinez, C., et al. (2007). Goulburn River experimental catchment data set. *Water Resources Research*, *43*(10). https://doi.org/10.1029/2006WR005837

Schaefer, G., Cosh, M., & Jackson, T. (2007). The USDA natural resources conservation service soil climate analysis network (SCAN). *Journal of Atmospheric and Oceanic Technology*, *24*(12), 2073–2077. https://doi.org/10.1175/2007JTECHA930.1

Schlenz, F., dall'Amico, J. T., Loew, A., & Mauser, W. (2012). Uncertainty assessment of the SMOS validation in the upper Danube catchment. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(5), 1517–1529. https://doi.org/10.1109/tgrs.2011.2171694

Shuman, D. I., Nayyar, A., Mahajan, A., Goykhman, Y., Li, K., Liu, M., et al. (2010). Measurement scheduling for soil moisture sensing: From physical models to Optimal control. *Proceedings of the IEEE*, *98*(11), 1918–1933. https://doi.org/10.1109/JPROC.2010.2052532

Smith, A., Walker, J., Western, A., Young, R., Ellett, K., Pipunic, R., et al. (2012). The Murrumbidgee soil moisture monitoring network data set. *Water Resources Research*, *48*(7). https://doi.org/10.1029/2012WR011976

Su, Z., Wen, J., Dente, L., Velde, R., Wang, L., Ma, Y., et al. (2011). The Tibetan Plateau observatory of plateau scale soil moisture and soil temperature (Tibet-Obs) for quantifying uncertainties in coarse resolution satellite and model products. *Hydrology and Earth System Sciences*, *15*(7), 2303–2316. https://doi.org/10.5194/hess-15-2303-2011

Tagesson, T., Fensholt, R., Guiro, I., Rasmussen, M., Huber, S., Mbow, C., et al. (2014). Ecosystem properties of semi-arid savanna grassland in West Africa and its relationship to environmental variability. *Global Change Biology*, *21*(1), 250–264. https://doi.org/10.1111/gcb.12734

Van Cleve, K., Chapin, F. S. S., & Ruess, R. W. (2015). Bonanza creek long term ecological research project climate database [Dataset]. *University of Alaska Fairbanks*. Retrieved form http://www.lter.uaf.edu/

Vreugdenhil, M., Dorigo, W., Broer, M., Haas, P., Eder, A., Hogan, P., et al. (2013). Towards a high-density soil moisture network for the validation of SMAP in Petzenkirchen, Austria. In *2013 IEEE International Geoscience and Remote Sensing Symposium—IGARSS* (pp. 1865–1868). https://doi.org/10.1109/IGARSS.2013.6723166

Wigneron, J.-P., Dayan, S., Kruszewski, A., Aluome, C., Al-Yaari, A., Fan, L., et al. (2018). The Aqui network: Soil moisture sites in the "Les Landes" forest and graves vineyards (Bordeaux Aquitaine Region, France). In *IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 3739–3742). IEEE. https://doi.org/10.1109/IGARSS.2018.8517392

Xaver, A., Zappa, L., Rab, G., Pfeil, I., Vreugdenhil, M., Hemment, D., & Dorigo, W. A. (2020). Evaluating the suitability of the consumer low-cost Parrot Flower Power soil moisture sensor for scientific environmental applications. *Geoscientific Instrumentation, Methods and Data Systems*, *9*(1), 117–139. https://doi.org/10.5194/gi-9-117-2020

Yang, K., Qin, J., Zhao, L., Chen, Y., Tang, W., Han, M., et al. (2013). A multi-scale soil moisture and Freeze-Thaw monitoring network on the third pole. *Bulletin of the American Meteorological Society*, *94*(12), 1907–1916. https://doi.org/10.1175/BAMS-D-12-00203.1

Young, R., Walker, J., Yeoh, N., Smith, A., Ellett, K., Merlin, O., & Western, A. (2008). *Soil moisture and meteorological observations from the Murrumbidgee catchment*. Department of Civil and Environmental Engineering, The University of Melbourne.

Zacharias, S., Bogena, H., Samaniego, L., Mauder, M., Fuß, R., Pütz, T., et al. (2011). A network of terrestrial environmental observatories in Germany. *Vadose Zone Journal*, *10*(3), 955–973. https://doi.org/10.2136/vzj2010.0139

Zappa, L., Forkel, M., Xaver, A., & Dorigo, W. (2019). Deriving field scale soil moisture from satellite observations and ground measurements in a Hilly Agricultural region. *Remote Sensing*, *11*(22), 2596. https://doi.org/10.3390/rs11222596

Zappa, L., Woods, M., Hemment, D., Xaver, A., & Dorigo, W. (2020). *Evaluation of remotely sensed soil moisture products using crowdsourced measurements*. Cyprus: SPIE. (Backup Publisher: Eighth International Conference on Remote Sensing and Geoinformation of Environment).

Zhang, X., Chen, N., Chen, Z., Wu, L., Li, X., Zhang, L., et al. (2018). Geospatial sensor web: A cyber-physical infrastructure for geoscience research and application. *Earth-Science Reviews*, *185*, 684–703. https://doi.org/10.1016/j.earscirev.2018.07.006

Zhao, T., Shi, J., Lv, L., Xu, H., Chen, D., Cui, Q., et al. (2020). Soil moisture experiment in the Luan River supporting new satellite mission opportunities. *Remote Sensing of Environment*, *240*, 111680. https://doi.org/10.1016/j.rse.2020.111680

Zheng, J., Zhao, T., Lü, H., Shi, J., Cosh, M. H., Ji, D., et al. (2022). Assessment of 24 soil moisture datasets using a new in situ network in the Shandian River Basin of China [Dataset]. *Remote Sensing of Environment*, *271*, 112891. https://doi.org/10.1016/j.rse.2022.112891

Zreda, M., Desilets, D., Ferré, T., & Scott, R. (2008). Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons. *Geophysical Research Letters*, *35*(21). https://doi.org/10.1029/2008GL035655

Zreda, M., Shuttleworth, W. J., Zeng, X., Zweck, C., Desilets, D., Franz, T., & Rosolem, R. (2012). COSMOS: The COsmic-ray soil moisture observing system. *Hydrology and Earth System Sciences*, *16*(11), 4079–4099. https://doi.org/10.5194/hess-16-4079-2012