Global and free datasets for the study of floods, droughts and human societies

Sara Lindersson[1,2], Luigia Brandimarte[3], Johanna Mård[1,2], Giuliano Di Baldassarre[1,2]

[1] Uppsala University, Sweden. [2] CNDS Centre of Natural Hazards and Disaster Science. [3] KTH Royal Institute of Technology, Sweden.

1 BACKGROUND

Global water challenges and the corresponding need for large-scale datasets are coinciding with an ongoing big data revolution.

Water-related disaster risk remain high in many parts of the world, from which there is an urgency to reduce humanitarian and economic losses.

Large-scale comparative research has the benefits of detecting patterns, reaching generalizable results, and building hypotheses about water risk propagation. Large-scale studies, however, raise data needs that are completely different from those of a single case study.

As it happens, the availability of planetary-scale geospatial datasets is rapidly growing. This rapid development has been triggered by new monitoring technologies, new processing capabilities, and a growing willingness to share data.

These available large-scale datasets span from being remotely sensed and gridded - to being measured on the ground and given in point format. In between, a wide number of datasets are springing up, building on and combining with other datasets.

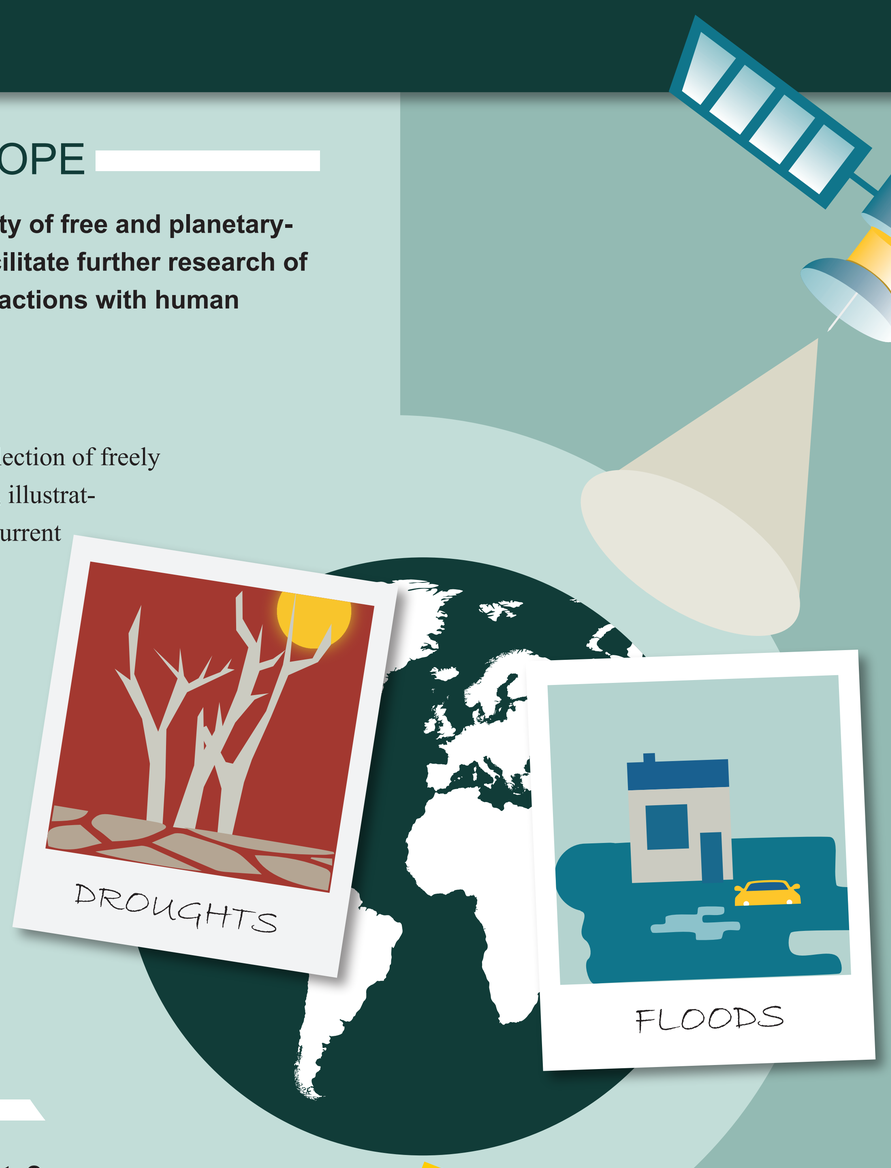
2 RESEARCH SCOPE

We review the current availability of free and planetary-scale geospatial datasets to facilitate further research of floods, droughts and their interactions with human societies.

More specifically, we:

- create and structure an extensive collection of freely accessible global geospatial datasets, illustrating research data opportunities and current data gaps.
- discuss general challenges for using these global datasets in comparative studies.

The purpose of presenting one list of datasets for a wide range of variables is to highlight research opportunities across scientific disciplines.



3 DATA REVIEW METHODS

We compiled a list of variables to include in the review, based on published conceptualizations of human-water disaster systems. In the search for datasets, we used a wide range of data sources, including peer reviewed literature (from data providers and data users), large geodata providers (e.g. NASA and ESA), geodata blogs, etc.

Strict inclusion criteria for datasets

- Accessibility: Online and free to use for research
- Spatial extent: Global or near-global
- Spatial resolution for grids: Finer than 50 km

Resolutions given priority (varies with variable)

- Spatial resolution: 30 m - 10 km
- Temporal resolution: Days - Months

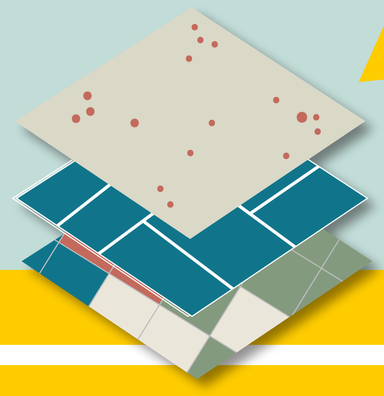
How did we chose among datasets?

We have given priority to...

... geographical data formats over tabular data formats.

... dynamic datasets capturing temporal changes, over static datasets.

... datasets showing improvement in accuracy and coverage by combining several parent datasets, rather than presenting each parent dataset individually.



4 COLLECTION OF DATASETS

The resulting data collection includes 124 datasets, divided into 7 categories and 36 sub-categories, shown below. The numbers indicate the number of datasets included in each category.



53% of datasets are offered in **raster** data format. The remaining datasets are either **vector** or **tabular**.

40% of datasets are **static**, lacking temporal depth. The **temporal coverage** of the remaining datasets span from 10 to 100 years.

24% of the time-varying datasets have unfixed **temporal resolutions** (e.g. event based). **Annual** and **monthly** time steps are most common among the fixed temporal resolutions.

Psst!

The collection of datasets is available at http://doi.org/10.5281/zenodo.3368882

5 DISCUSSION POINTS

Despite the global spatial extent, these datasets do not exhibit a fair geographic representation across the globe. The reasons for this differ accross the wide range of datasets, e.g. technical constraints in remotely sensed datasets favouring certain environmental conditions, or reporting biases in databases favouring certain socioeconomical conditions.

New technologies contribute to unprecedented fine spatial resolutions, but also generate limitations on usability. Challenges for downscaled datasets, for instance, involve: time varying datasets advised not to be used in time series analysis, fine spatial resolution datasets advised not to be used in local studies, datasets built on a wide variety of data sources prohibiting systematic uncertainty assessments, and data dependency issues preventing interaction studies.

Relating to disaster risk: Our review confirms that the geographic representation of disaster hazards and exposure is far more mature than the geographic representation of vulnerability aspects (still being mostly presented in tabular datasets, at national or regional scales). We are however optimistic that the following years will see further progress in capturing the vulnerabilities of humanity as well, not least due to current progress in very high-resolution satellite imagery in combination with machine learning techniques.

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

sara.lindersson@geo.uu.se
www.hydrosocialextremes.org

CNDS
Centre of Natural Hazards & Disaster Science

CC BY