

# Reviewer 1

**LINK:** <https://doi.org/10.5194/essd-2025-482-RC1>

**REVIEW:**

The manuscript presents the Unified Global Landslide Catalogue (UGLC), a global dataset obtained through the integration and harmonization of numerous pre-existing landslide inventories, catalogues, and archives, with the aim of providing a standardized basis for global-scale analyses, modelling, and risk-management applications.

The initiative to collect and make globally distributed landslide information accessible under an open license undoubtedly represents a significant and commendable effort, consistent with the mission of journals oriented toward data publication. Moreover, the definition of a common ontological structure and standardized attributes constitutes a methodological contribution that could potentially foster interoperability and data sharing at the international level.

However, despite these valuable elements, the work presents substantial critical issues that limit both the scientific robustness of the dataset and its effective reusability for advanced quantitative applications. In particular, key concerns relate to the epistemological heterogeneity of the source data, the absence of independent quantitative validation, spatial, temporal, and informational inconsistencies, and—most importantly—ambiguity regarding the appropriate domains of scientific use.

The data are not new in an observational sense, but have been reorganized and harmonized by the authors to homogenize their attributes. The novelty is therefore structural and ontological rather than empirical. Indeed, the work compiles landslide data from different global sources, making it potentially useful for reconnaissance of available sources, global descriptive analyses, and data-discovery activities; however, their usefulness for quantitative modelling, susceptibility assessment, or machine learning appears limited due to the strong heterogeneity of the original datasets. For example, the Itatica dataset cannot be used for certain landslide susceptibility assessments, just as inventories produced for earthquake-induced landslides cannot be used for modelling different from seismic contexts. Some datasets are discontinuous, such as Itatica, and therefore cannot be used, for instance, in machine-learning modelling.

The major issue with the UGLC is that unifying attributes is not sufficient to make data homogeneous; it is essential to trace the purpose for which each dataset was generated, because data collection is functional only to its original objective. Mixing datasets collected for different purposes may lead to careless use and the risk of obtaining results that are not truly supported.

“This study offers a conceptual framework and workflow that can serve as a template for the international standardisation of landslide data management and can even be applied in the development of smaller-scale landslide datasets.” As stated above, having an ontologically correct and uniform dataset is not enough to claim that it can be used indiscriminately. The heterogeneity of the data composing the UGLC is so large that, personally, I would find it difficult to use it as a unified dataset.

The description of materials and data sources is overall adequate, and the original datasets are accessible. However, the criteria for semantic normalization, the definition and assignment of reliability, and the procedures for managing uncertainties are less clear. In the methods section, more methodological details are required, including examples of how attributes from the original datasets/catalogues were transferred into the new global catalogue.

Regarding Italian datasets, other published inventories exist that were not considered in the catalogue. How do the authors manage the geographical overlap of data derived from different catalogues/inventories?

Furthermore, it is essential that the authors better clarify the limits of scientific use. Many datasets were created for specific purposes and are not automatically transferable to other analytical contexts. Attribute

unification does not make the data scientifically homogeneous, and mixing datasets with different purposes may produce unsupported results.

Is the dataset only partially accessible? All tiles from 8 to 15, from Italy to Japan, are missing for download. Therefore, the dataset described in the article is not fully downloadable. In addition, some links are inaccessible, making the section with links to newspaper reports, for example, useless. This occurs, for instance, for data derived from the Cooperative Open Online Landslide Repository – NASA and from the Global Fatal Landslide Catalog.

No error estimates or sources of uncertainty are provided. Independent quantitative validation is lacking, which unfortunately represents a critical issue for a data paper.

No common quality standards emerge among the integrated datasets.

Each dataset carries the problem of data completeness. Beyond spatial-temporal accuracy, data should also be complete. Do the original datasets describe this aspect? If not, the authors should include a disclaimer. This does not refer to areas where data do not exist, but to situations where data exist but are incomplete. Unfortunately, the declaration of completeness is a limitation of many inventories and catalogues, and this aspect must be clearly stated.

Yes, geographic data standards are followed. Although the global open-access collection effort is appreciable, scientific quality remains limited by heterogeneity, incompleteness, and lack of validation. Currently, the dataset is technically usable only in partial form (only half of the globe). Metadata are not fully adequate. The structure is clear, but technical database validation is missing. The language is consistent, though layout issues occur in Tables 4, 5, and 6.

In practice, reusability for quantitative analyses is limited by intrinsic heterogeneity. Attribute standardization creates an apparent uniformity that does not correspond to real scientific homogeneity. Therefore, the work represents a significant effort toward global integration and standardization of landslide data and constitutes a potentially useful open information infrastructure for synoptic consultation and large-scale descriptive analyses. However, relevant critical issues limit its scientific maturity: epistemological heterogeneity of the data, current absence of quantitative validation, and incompleteness and uncertainty of the information, resulting in limited reusability for modelling. Consequently, the manuscript appears potentially publishable only after substantial revisions, authors should clarify the epistemological limits of the catalogue and strengthen data-quality assessment. I also invite the authors to scale down claims regarding global modelling use. My recommendation is major revision.

## **RESPONSE TO Reviewer 1:**

We sincerely thank the reviewer for the rigorous and conceptually deep evaluation of our manuscript. We particularly appreciate the emphasis on epistemological heterogeneity, which we recognise as a fundamental aspect of global landslide data integration.

We clarify that the UGLC does not produce a scientifically homogeneous dataset in an observational sense. Indeed it is based on the sampling strategy endorsed in each single original database.

The novelty of the work is structural and ontological, not empirical.

In the revised manuscript, we have explicitly clarified that UGLC is an organisational and semantic integration framework, rather than a homogenised dataset suitable for indiscriminate quantitative use. Importantly, we emphasise that the catalogue preserves the epistemological context of each source dataset through explicit provenance attributes (e.g., OLD\_DATASET, OLD\_ID, TYPE, PHYSICAL\_FACTORS). Landslide records remain valid observations within their original acquisition context, and UGLC does not attempt to reinterpret or generalise them beyond their native meaning.

To address the reviewer's concern regarding potential misuse, we have introduced a dedicated section clarifying intended and non-intended uses. In particular, we explicitly state that:

- the catalogue is suitable for data discovery, exploratory analyses, and large-scale descriptive studies;
- modelling applications require careful, user-driven filtering based on trigger type, landslide classification, spatial/temporal accuracy, and dataset provenance;
- the dataset should not be used as a statistically homogeneous input without preprocessing.

These additions are reflected in the revised manuscript ( line 44-51 section: “1. Introduction”; section: 1.2 “Emerging need for harmonised landslide data”; section: 1.3 Improving geospatial aspects of landslide catalogues”; line 384-390 section: “5.1 Analysis and statistical insights on UGLC”; section: “5.2 Epistemological limitations and intended use”; section: “7.4 Landslide models and predictive applications”; section: “8. Conclusion” )

About the semantic normalisation and methodological transparency, we have substantially expanded the methodological description and the normalisation process is now clarified as follows:

- Attribute harmonisation is driven by dataset-specific JSON lookup tables, available in the public GitHub repository;
- These lookup tables were manually curated through a systematic, dataset-by-dataset analysis, ensuring semantic consistency while minimising interpretative bias;
- The process is intentionally conservative, aiming to standardise existing information (e.g., resolving naming inconsistencies, typos, case sensitivity issues) rather than infer new attributes;
- Interpretative steps were limited and applied only where sufficient contextual metadata supported them;
- In the absence of reliable information, attributes are explicitly marked as ND or -99999, preserving the integrity of the original data.

This clarification reinforces that the harmonisation process does not artificially homogenise the data, but instead regularises their representation.

These additions are reflected in the revised manuscript ( section: “4.3 Semantic standardisation and attribute harmonisation”; section: “4.5 Data normalisation”)

On duplicate detection and dataset overlap we have clarified that duplicate records are identified using a strict matching criterion based on: identical spatial coordinates and identical temporal attributes.

When duplicates are detected, they are interpreted as informational redundancy across datasets. In such cases the most informative record is retained, to have fully identical records reduced to a single entry.

This approach is conservative and may not detect near-duplicates, prioritising traceability and reproducibility over aggressive deduplication.

These additions are reflected in the revised manuscript ( section: “4.4 Data merge”)

About the reliability, uncertainty, and completeness, we agree that these aspects require clearer clarification.

RELIABILITY attribute is now explicitly defined as a relative confidence ranking, based on spatial and temporal precision not a quantitative uncertainty measure, and has not been independently validated as a statistical accuracy metric

We also explicitly state that:

- no independent error estimation or uncertainty propagation is performed;
- completeness is inherited from source datasets and is rarely explicitly quantified in the source datasets;
- UGLC does not guarantee completeness and should not be interpreted as such.

A formal disclaimer on data completeness has been added.

These additions are reflected in the revised manuscript ( line: 185-197 section: “3.2 Attributes structure and specifics”)

Regarding the missing tiles we thank the reviewer for identifying the issue. This was caused by a technical inconsistency in the tiling workflow and did not affect the master dataset.

The issue has now been resolved, all tiles have been regenerated and verified and all polygon data are distributed as a global CSV to ensure full accessibility (Broken external links have been updated or removed where necessary).

On modelling claims we fully acknowledge reviewer's concern and have revised the manuscript to moderate claims related to modelling applications. The text now clearly states that UGLC is designed to support large-scale analyses, but requires context-aware filtering and domain expertise for modelling purposes.

These additions are reflected in the revised manuscript ( line 44-51 section: "1. Introduction; section: "5.2 Epistemological limitations and intended use"; section: "7.4 "Landslide models and predictive applications; section: "8. Conclusion")

We thank the reviewer again for highlighting critical conceptual aspects. The manuscript has been substantially revised to improve clarity, transparency, and correct scientific positioning of the dataset.

---

# Reviewer 2

**LINK:** <https://doi.org/10.5194/essd-2025-482-RC2>

**REVIEW:**

Dear Editors,

I have carefully reviewed the manuscript entitled “Unified Global Landslide Catalogue (UGLC): A single, standardised global-scale landslide dataset”. The authors present an ambitious and valuable contribution aimed at harmonising 29 heterogeneous landslide inventories into a unified global catalogue comprising more than one million events. The work addresses a well-recognised gap in the geomorphology and disaster-risk communities: the absence of a standardised, ontologically coherent global landslide dataset.

Overall, the manuscript is clearly written, well organised, and supported by a strong scientific motivation. The authors convincingly highlight the need for harmonised global data, noting that landslide information is often fragmented, inconsistent, and collected across diverse spatial and temporal scales—conditions that hinder large-scale analyses and modelling efforts. The proposed UGLC has the potential to become a reference dataset for global susceptibility modelling, climate–landslide interaction studies, and risk assessment.

## Major Comments

### 1. Insufficient methodological transparency in the normalisation process

The manuscript emphasises the novelty of the proposed ontological framework, described as “the first published multilevel ontological methodology specifically developed for landslide phenomena”. Despite this, the description of the normalisation workflow remains too limited. The authors should clarify in greater detail how semantic conflicts between datasets were resolved; how duplicate events were identified and merged; how missing or inconsistent attributes were handled; how spatial accuracy was inferred when not explicitly provided; how point and polygon datasets were harmonised under a common schema.

Given that the scientific value of UGLC depends critically on these steps, a more explicit methodological description is essential

### 1. Reliability classification requires stronger justification

Table 4 introduces a reliability score based on spatial and temporal precision. While the concept is intuitive, the manuscript does not explain:

- the rationale behind the chosen spatial thresholds (<100 m, <250 m, <500 m);
- how reliability is computed for polygon datasets;
- how temporal uncertainty is treated when only partial dates are available.

A conceptual justification—or ideally a sensitivity analysis—would significantly strengthen the credibility of the reliability metric.

### 1. Dataset imbalance and geographic bias

The catalogue exhibits substantial imbalance among contributing datasets (e.g., IFFI alone contributes more than 600,000 points). Although the authors acknowledge that data availability is higher at subnational levels and in regions with high susceptibility, they do not discuss how this imbalance affects:

- global statistical analyses,
- susceptibility modelling,
- interpretation of temporal trends.

A dedicated subsection addressing dataset bias and its implications is needed.

1. Lack of validation or demonstration of scientific utility

The manuscript would benefit greatly from including at least one of the following:

- a comparison with existing global datasets (e.g., GLC, NASA COOLR);
- an example application (e.g., density mapping, susceptibility modelling);
- a validation exercise assessing spatial accuracy for a subset of events.

At present, the manuscript describes the dataset but does not demonstrate its practical or scientific utility.

The manuscript represents a highly valuable contribution to the field. However, additional methodological detail, clearer justification of assumptions, and a more explicit discussion of limitations and biases are required.

## **RESPONSE TO Reviewer 2 :**

We thank the reviewer for the constructive and insightful feedback, and for recognising the potential value of the UGLC dataset.

- On methodological transparency:

We have significantly expanded the description of the normalisation workflow. In particular, the semantic harmonisation is performed through dataset-specific JSON lookup tables, publicly available in the GitHub repository. These tables were manually curated through a systematic review of each dataset to ensure scientific consistency and minimal interpretative bias. The process focuses on standardising existing information, rather than inferring new attributes; while missing or unreliable information is explicitly preserved using standardised no-data values.

Moreover, we also clarified:

- the duplicate detection criteria (exact match on coordinates and date);
- conservative handling of spatial accuracy inference;
- the preservation of every original geometries for both point and polygon datasets.

These additions are reflected in the revised manuscript (line: 180-197 section: "3.2 Attributes structure and specifics"; section: "4.2 Data conversion"; section: "4.3 Semantic standardisation and attribute harmonisation"; line: 265-269 section: "4.4 Data merge"; line: 323-333 section: "4.5 Data normalisation")

- On reliability classification:

We clarified that the RELIABILITY attribute is a ranking system, not a probabilistic metric. It is designed to support data filtering based on practical thresholds reflecting typical geospatial precision levels. We also clarified how this metric is applied mainly to point data. While polygon data are considered spatially accurate by definition, but may or may not be temporally accurate. Consequently, polygons can have a RELIABILITY value of 1 or 2.

Regarding the treatment of temporal uncertainty when only partial dates are available every case was explained in (line: 310-321 section: "4.5 Data normalisation").

We also clarify that a global validation (e.g., via multi-temporal satellite imagery) would require a whole dedicated effort beyond the scope of this study, and this limitation is now explicitly stated in the manuscript. These additions are reflected in the revised manuscript ( line: 185-197 section: "3.2 Attributes structure and specifics"; section: "5.2 Epistemological limitations and intended use")

- On dataset imbalance and bias:

We have added a dedicated subsection explicitly addressing:

- geographic imbalance;
- bias related to trigger types (rainfall and seismic);
- implications for statistical analysis and modelling.

We clarify that these biases reflect data availability and reporting practices, rather than global landslide occurrence patterns.

These additions are reflected in the revised manuscript (line: 384-390 and 405-410 section: "5.1 Analysis and statistical insights on UGLC"; section: "5.2 Epistemological limitations and intended use")

- On validation and scientific utility:

We agree that demonstrating utility is important.

A direct comparison with existing global datasets (e.g., GLC, NASA COOLR) would be largely redundant, as these datasets are already included within the UGLC (e.g., GLC is incorporated into NASA COOLR, which is part of UGLC).

Moreover, independent validation falls beyond the scope of this data paper and would require a dedicated effort. The original datasets were curated and published by their respective providers, often through peer-reviewed or institutional frameworks, and are therefore considered to have undergone prior quality control.

To address the reviewer's concern, we have added an exploratory analysis comparing UGLC point data with global rainfall intensity (NASA GPM-IMERG, 2000–2025) and seismic hazard (peak ground acceleration from the Global Earthquake Model – GEM, 10% probability of exceedance in 50 years,  $V_{s30} = 760\text{--}800$  m/s).

This comparison highlights the expected spatial clustering of landslides in regions characterised by intense rainfall and seismic activity, and helps illustrate inherent biases in the dataset; as well as its practical and scientific utility.

In addition, we have:

- clarified possible applications and their limitations;
- explicitly framed UGLC as a data infrastructure rather than a validated modelling dataset.

These additions are reflected in the revised manuscript ( line 44-51 section: "1. Introduction"; section: 1.2 "Emerging need for harmonised landslide data"; section: 1.3 Improving geospatial aspects of landslide catalogues"; line: 384-390 and 405-410 section: "5.1 Analysis and statistical insights on UGLC"; section: 5.2 "Epistemological limitations and intended use"; section: 7.4 "Landslide models and predictive applications"; section: "8. Conclusion").

Final remark

Thanks to your suggestions the manuscript has been strengthened through improved methodological transparency, clearer justification of assumptions, and a more explicit discussion of limitations and bias.