Unified Global Landslide Catalogue (UGLC): A single, standardised global-scale landslide dataset

Saverio Mancino[1,5,*], Anna Sblano[2,*], Francesco Paolo Lovergine[3], Vincenzo Massimi[2], Tushar Sethi[4,5], Domenico Capolongo[1], and Giuseppe Amatulli[4,5,6]

[1]Department of Earth and Geoenvironmental Sciences, University of Bari Aldo Moro, 70125 Bari, Italy
[2]Planetek Italia, 70132 Bari, Italy
[3]Institute for Electromagnetic Sensing of the Environment (IREA), National Research Council of Italy (CNR), 70126 Bari, Italy
[4]Margosa Environmental Solutions Ltd., Brandon House, 1st Floor, 90 The Broadway, Chesham, HP5 1EG, UK
[5]Spatial Ecology, Brandon House, 1st Floor, 90 The Broadway, Chesham, HP5 1EG, UK
[6]School of the Environment, Yale University, New Haven, CT 06511, USA
[*]These authors contributed equally to this work.

Correspondence: Saverio Mancino (saverio.mancino@uniba.it) and Giuseppe Amatulli (giuseppe.amatulli@yale.edu)

Abstract. Landslides are a serious threat to all communities due to their potential for property damage and loss of life. Triggered by different natural, climatic and anthropogenic factors, landslides are complex phenomena and difficult to identify, monitor, and manage (Kirschbaum et al., 2015) https://doi.org/10.1016/j.geomorph.2015.03.016. Accurate and comprehensive data are essential in the mitigation of landslide risk, where both the likelihood and impact of landslides on communities must be quantified. Robust datasets allow for the development of dependable prevention strategies such as land use planning and early warning systems. These proactive measures play a crucial role in landslide risk mitigation (Gomez et al., 2020) https://doi.org/10.1007/s11069-023-05848-8.

This study presents a single global scale standardised landslide catalogue, the Unified Global Landslide Catalogue (UGLC), which is intended as an harmonised landslide information framework designed to support data discovery, comparative analyses, and large-scale descriptive investigations relevant to risk-related studies. UGLC integrates multiple open data landslide datasets and reports spatiotemporal data with trigger factors for landslides. Landslide occurrence data are collected from extensive field surveys, GPS data, GIS techniques, satellite imagery, and historical records sourced from government agencies, universities, and researchers.

UGLC contains more than 1 million landslide events as point and polygonal data, from the period spanning circa 1700 to 2023. The catalogue is standardised across 18 field attributes, and systematically grouped into seven main categories: (1) UGLC Reference – a unique event identifier; (2) Source Reference that enables back-tracing to the original data source; (3 and 4) Spatial Accuracy and Temporal Accuracy – precisely describe the geographic location and temporal resolution of recorded events, respectively; (5) Geological Information, including triggering factors; (6) Reliability, which assigns a trustworthiness value to the data; and (7) Notes and Information containing supplementary details such as source links, authorship, scientific publications, and other relevant metadata.

UGLC is intended as a robust catalogue of standardised landslide information worldwide. The aim is to provide a reliable and user-friendly source for the characterisation of landslide occurrence. Uniquely, it presents a comprehensive range of data for global analysis and thus compensates for the shortcomings of small-scale heterogeneous datasets. UGLC will facilitate a deeper understanding of landslide phenomena in relation to the surrounding landscape, climate, and impact on human populations and the built environment (Kirschbaum et al., 2015) https://doi.org/10.1016/j.geomorph.2015.03.016. UGLC is publicly available at Zenodo (DOI: https://doi.org/10.5281/zenodo.18643456).

1 Introduction

Landslides represent a significant global geological hazard, as they annually inflict thousands of casualties and economic losses estimated at USD 20 billion (Sim et al., 2022). Their effects are largely felt by vulnerable communities in mountainous regions (Hovius and Stark, 2006) and developing countries (Liu et al., 2024), and are exacerbated by anthropogenic factors (Jaboyedoff et al., 2018), climate change-driven extreme weather (Gariano and Guzzetti, 2016) and increasing urban encroachment into susceptible areas. Hence, these phenomena disrupt infrastructure (Mitsugi, 2018), ecosystems (Hearn and Hart, 2011), and essential services such as food and water security (Liu et al., 2024). Addressing the escalating threat of landslides requires newer and more effective risk management strategies that draw from reliable data and advanced technologies. The compilation of large landslide datasets requires a deep, multidisciplinary understanding of landslide dynamics and may combine a variety of methods. Traditional ones such as field surveys are reliable but limited in scale and accessibility. On the other hand, advanced emerging approaches comprising machine learning to process remote sensed data like high-resolution satellite imagery and Digital Elevation Models (DEMs) allow for deep geospatial analysis and automated landslide detection. (Wang et al., 2020; Koneru et al., 2024).

Landslide information is often characterised by fragmented datasets collected across diverse spatial-temporal scales. Its heterogeneity and acute lack of standardisation greatly hinder large-scale analysis and modelling efforts (Tehrani et al., 2022). Consequently, there is an urgent and unmet need for a consolidated and standardised global catalogue of landslides that supports data discovery, exploratory analyses, and structured integration of heterogeneous landslide information useful in decision-making for risk mitigation (Jiang and Wang, 2024).

This paper presents the Unified Global Landslide Catalogue (UGLC) as a solution to the gaps in landslide data: its meticulous construction and domain-specific ontology normalise, harmonise, and integrate information from 29 fragmented global sources that encompass over a million recent and historical landslide events. However, the harmonisation process of several datasets developed under heterogeneous conditions and for different purposes does not eliminate differences in mapping strategies, data acquisition objectives, spatial completeness, temporal coverage, or classification criteria. UGLC preserves data completeness and incompleteness from the original datasets, which may result in the absence of records from a given region. However, unavailable data does not impute an absence of landslide occurrence. Context-aware filtering for analytical use is thus essential, as UGLC does not constitute a scientifically homogeneous dataset.

1.1 State of the art of landslide datasets and catalogues

Landslide data repositories can be broadly categorised into two main types: large-scale landslide datasets and detailed landslide catalogues. Large-scale datasets are systematic spatial and temporal compilations often assembled at national to global scales using relatively standardised methodologies to promote consistency and reproducibility (Gomez et al., 2020). However, these are frequently characterised by sparse ancillary information, inconsistent thematic content, and heterogeneous spatial coverage, which limits their utility for detailed localised analysis. Causes of such irregularities range from restrictions on regional data access to limited incentives for compiling transnational datasets into open-access global inventories. Furthermore, datasets that document older landslides and those from remote regions are often few in both number and record contained, resulting in a natural under-representation of certain geographic areas and historical periods (Petley, 2012). Nevertheless, data availability is better at subnational levels, particularly in regions with high landslide susceptibility, as inventories are often more accessible, and the increased attention to the phenomenon drives systematic data acquisition.

In contrast to datasets, landslide catalogues typically represent subnational to local inventories developed from diverse sources and often heterogeneous data classes, including historical archives, scientific literature, and field survey reports. Catalogues tend to provide richer descriptive metadata, such as event triggers, geomorphological characteristics, and impact data but are usually compiled without conformity to any standardised data collection protocols, resulting in considerably variable data structure and quality.

High-quality landslide data organised according to a structured ontology is key to improving the state-of-the-art (Tehrani et al., 2022). With this aim, UGLC assimilates landslide records under a framework of 18 standardised attributes, which promotes semantic consistency and standardised classifications while inheriting the spatial and temporal characteristics of the original datasets. The global catalogue design facilitates integration into GIS-based workflows while preserving the intrinsic heterogeneity of the original inventories, thereby supporting dataset-specific analyses and filtering strategies. In response to increasing exposure to disaster risk and the destructive potential of landslides, UGLC can therefore serve as locus of disparately catalogued global landslide data to aid urgent revisions in spatial planning and disaster management strategies (on Climate Change, 2022; UNDRR, 2021).

1.2 Emerging need for harmonised landslide data

The greater focus on landslide-related catastrophes coincides with an increase in remote sensed data availability on such phenomena. Scientists have thus been able to direct research efforts towards advanced data-driven methods and emergent landslide-related models are increasingly calling for spatially explicit but large datasets for accurate simulations, predictions, and mitigation planning. However, large-scale data integration is impeded by substantial differences in how data are collected, structured, and interpreted. In addition to technical differences such as file formats or attribute naming, there are deeper epistemological divergences related to the purpose, methodology, and context of data acquisition. Bringing consistency to the syntax and semantics of cataloguing methods is essential to creating the continental-scale landslide datasets that are now needed for advancing research.

From a semantic perspective, similar landslide phenomena may be classified under different schemes or terminologies across datasets. For example, the same landslide process may be recorded using different naming conventions (e.g., “rock_fall”, “Rockfall”, or “rock fall”), while incomplete or inconsistent attributes (e.g., “unknown”, “null”, or “other”) require explicit handling. In some cases, differences also occur at the interpretative level, such as the classification of “topple” versus “rock fall”. At the same time, syntactic and structural inconsistencies, such as variations in attribute fields, geometry types (points versus polygons), and temporal representation, further complicate direct comparison and integration of data. These differences prevent direct interchangeability across datasets, since each inventory is shaped by its original scientific or operational objectives. As a result, combining multiple sources without a consistent framework can lead to ambiguity, loss of information, or misinterpretation.

A harmonisation strategy that preserves the original meaning of the data while enabling consistent and interoperable data representation should resolve such cataloguing challenges in large scale landslide inventories. These needs and challenges have prompted the methodology developed in this study. We expect that the resulting UGLC will meet the demand for enhanced analysis, including automatic landslide detection and segmentation methods, which are are often based on machine learning algorithms applied to high-resolution satellite multispectral imagery or Synthetic Aperture Radar (SAR) interferometry (Morales et al., 2022; Di Napoli et al., 2020; Bhuyan et al., 2023).

1.3 Improving geospatial aspects of landslide catalogues

Mapping a landslide event typically entails generating points and/or polygon vectors. Points represent key features such as the landslide’s initiation zone, centroid, or depositional terminus. Point-based datasets are usually employed in regional to global inventories where precise delineation is not feasible. Although they are efficient for spatial referencing and statistical aggregation, point representations provide only limited information on the spatial extent and geomorphic footprint of the event. Meanwhile, polygon-based datasets delineate the entire landslide boundary, encompassing both the source and the depositional zones. This format enables more comprehensive spatial analyses, such as volume estimation, terrain deformation modelling, and spatial susceptibility mapping, particularly when high-resolution remote sensing data or detailed field observations are available.

The prevalence of point datasets in global landslide inventories is often due to their ease of generation from widely available imagery and anecdotal event reports. However, poor-resolution imagery and the lack of precise spatial boundaries can inhibit analysis. Moreover, point datasets from different sources are fundamentally inconsistent in terms of data structure, semantic definitions, collection methodologies, and spatial geometries. These inconsistencies pose a significant barrier to transboundary analyses and undermine efforts to construct globally coherent landslide assessments.

A major challenge for the geomorphology research community is to bridge the gap between wide-ranging point datasets and localised and richly detailed catalogues. To this end, the development of a consolidated and standardised global landslide catalogue is essential. Such an effort must involve robust big-data harmonisation strategies that are grounded in a domain-expert-driven ontological framework. This would ensure semantic consistency, structural integrity, and multi-source integration of heterogeneous datasets into a cohesive, analysable whole. In the following sessions, we describe the development of UGLC, starting with a summary of the input data sources and followed by the methods to integrate them within a prescribed ontology.

2 Data sources

The proposed UGLC is created from a collection of several open access landslide datasets containing landslide events until 2024. Only datasets published under a CC BY 4.0 licence, which allows the sharing and editing of data, were chosen for the catalogue. A comprehensive and standardised analysis was performed on each incorporated landslide dataset to examine and reorganise all embedded information. The main differences in the datasets are related to the area of interest, the number of records contained, the method of creation, the search for any reference publications, and its licence for use.

UGLC comprises separate point and polygon collections, but both of these catalogues adhere to the same attribute standard. The current version of UGLC contains 17 datasets in the point catalogue and 11 in polygon one. Future versions may increase the number of datasets incorporated in each catalogue. We provide below a brief summary of the datasets compiled in the point and polygon catalogues, while a detailed description for each can be found in the appendix A and B.

2.1 Point datasets

Table 1 provides an overview of the 17 point-based landslide catalogues compiled at global, national, and regional levels. Additional technical details of the dataset, including data access links and bibliographic references, are provided in the appendix A: section 9. For each constituent dataset, the table reports the acronym, version (where applicable), and total number of recorded events. These datasets differ significantly in terms of data acquisition methodology, spatial resolution, temporal coverage, and thematic scope.

Overall, these point-based datasets reflect a wide range of acquisition techniques, from manual archival interpretation (e.g., IFFI, NTMI) to advanced remote sensing workflows (e.g., NZK, PCLD, RBR).

From a temporal perspective, the datasets cover very different periods: some catalogues collect historical events going back several centuries, such as SLIDO (even before 0) and NTMI (since 1488), while others focus on more recent periods, such as NZK (2016 only) or 1N (2015-2027). Although some offer broad national coverage, e.g. BGS, UAP, others are thematically or regionally focused, capturing specific processes or localised hazard contexts (e.g., CAFLAG, 1N, NZK).

2.2 Polygonal datasets

Table 2 summarises several datasets of landslides mapped in polygonal format, providing essential information on their spatial and temporal extent. In total, it contains 11 global to local landslide datasets distinguishable by their geographical scale. The table reports the dataset acronyms, the geographical coverage, and the number of polygons contained.

These polygonal datasets vary widely in temporal resolution and acquisition strategy, as they span from manual photo-interpretation and field validation (e.g., IFFI, GEUS, PH) to advanced remote sensing and AI-driven classification approaches (e.g., PALI, JLD). Some are tailored to specific triggers and geographies (e.g. MAL, HLD, ASM), while others provide generalised coverage suitable for broader hazard assessments (e.g., ETFGI, COOLR). The appendix B has further technical details and access information for each dataset: section 9.

Dataset	ID Name	Nº Points	Geographical Coverage	Time Period	Reference
Cooperative Open Online Landslide Repository (version 2023)	COOLR	49 713	Global	1956-2023	(Kirschbaum et al., 2010, 2015)
Global Fatal Landslide Database (Version 2017)	GFLD	5490	Global	2004-2017	(Froude and Petley, 2018)
Italian Rainfall Induced Landslides Catalog (V2 - 2023)	ITALICA	6312	Italy	1996-2021	(Peruccacci et al., 2023)
Landslide Inventories across the United States (V2 - 2022)	UAP	176 427	USA, Alaska, Puerto Rico	1996-2021	(Mirus et al., 2020)
Australia Landslide Catalogue (Version 2018)	ALC	1653	Australia	1900-2016	(Australia, 2016)
Preliminary Canadian Landslide Database (V8.0)	PCLD	10 134	Canada	1771-2023	(Brideau et al., 2024)
Shallow Landslide Inventory for 2000-2019 (V1.0)	RBR	7944	Eastern DRC, Rwanda, Burundi	2001-2019	(Depicker et al., 2020, 2021)
Map of co-seismic Landslides for the 7.8 Kaikoura earthquake (V2.0)	NZK	29 519	New Zealand	2016	(Massey et al., 2021)
Mass Movements Information System (SIMMA) of the Colombian Geological Service (Version 2023)	CA	2407	Colombia	1929-2023	(Herrera-Coy et al., 2023)
National Landslide Database - Index data (BGS) (Version 2020)	BGS	18 217	UK	1664-2023	(Pennington et al., 2015)
Landslide Events Data (GSI) (Version 2016)	NTMI	2811	Ireland	1488-2020	(Ireland, 2012)
Vermont Geological Survey's preliminary landslide inventory (Version 2024)	VLS	3049	USA	2002-2019	(Survey, 2021)
Statewide Landslide Information Database for Oregon (V4.5)	SLIDO	15 377	Canada	a.C-2023	(DOGAMI, 2024)
1N (2015-2027): French Landslide Observatory – OMIV (Version 2024)	1N	25	France	2015-202	(Malet et al., 2015)
The CAmpi Flegrei LAndslide Geodatabase (V2.0)	CAFLAG	2302	Italy	1828-2017	(Esposito and Matano, 2023)
ETGFI - Earthquake-Triggered Ground-Failure Inventories (Version 2022)	ETGFI	115 402	Global	1908-2021	(Schmitt et al., 2017)
Inventory of landslide phenomena in Italy (Version 2021)	IFFI	622 447	Italy	1920-2021	(Trigila et al., 2021)

Table 1. General information about point datasets involved into UGLC creation.

Dataset	ID Name	Nº Polygons	Geographical Coverage	Time Period	Reference
Patagonian Andes Landslides Inventory (Version 2022)	PALI	10 026	Argentina, Chile	2020-2021	(Morales et al., 2022)
Inventory of landslides triggered by the 2015 Mw 6.0 Sabah earthquake (V1.0)	MAL	5198	Malaysia	2008-2015	(Ferrario, 2022)
Inventories of landslides triggered by the 2019 Cotabato - Davao del Sur (Philippines) seismic sequence (V1.0)	PH	10 593	Philippines	2019	(Ferrario et al., 2023)
Haiti Landslide Dataset (V1.0)	HLD	4178	Haiti	2021	(Martinez et al., 2021)
Polygon inventory of 12.920 Asia Summer Monsoon (ASM) Triggered landslides in Nepal (V1.0)	ASM	12 920	Nepal	2018-2020	(University, 2024)
Danish landslide inventory (Version 2022)	GEUS	3202	Denmark	2014-2020	(Luetzenburg et al., 2022)
Utah Landslide Inventory Polygons (Version 2018)	UTH	2381	Utah	1937-2012	(Survey, 2018)
Japan landslide dataset for semantic segmentation (V1.0)	JLD	330	Japan	nd	(Bragagnolo et al., 2020)
Cooperative Open Online Landslide Repository Polygons (Version 2023)	COOLR	20 060	Global	1897-2023	(Kirschbaum et al., 2010, 2015)
Earthquake-Triggered Ground-Failure Inventories Polygons (Version 2022)	ETFGI	491 839	Global	1876-2020	(Schmitt et al., 2017)
Inventory of landslide phenomena in Italy (Version 2021)	IFFI	423 399	Italy	1920-2021	(Trigila et al., 2021)

Table 2. General information about polygonal datasets involved into UGLC creation

3 UGLC design

The development of UGLC was guided by the need to reorganise yet retain information from original data sources. Its design objective was to offer a centralised, accessible resource to serve both casual GIS users and expert modellers. To achieve this, the UGLC data were structured to support easy integration into geospatial workflows while maintaining the detail and consistency required for structured and reproducible geospatial analyses, and preserving the heterogeneity of source data.

3.1 Ontology in data normalisation

According to our research, this is the first published multilevel ontological methodology specifically developed for landslide phenomena. We are not aware of any similarly structured domain-specific global framework established prior to this work. (Wen et al., 2023). This unique systematisation was essential for the integration of a highly heterogeneous set of local, national, and global landslide datasets, each characterised by distinct attributes, structures, formats, and original purpose. Expert interpretation was used to harmonise the various data sources, resolve inconsistencies, and standardise attributes by ensuring scientific clarity and usability of the catalogue. Section 4.4 “Data Normalisation” provides a detailed explanation of the normalisation ontology.

3.2 Attributes structure and specifics

Spatio-temporal accuracy is fundamental to developing a reliable landslide catalogue, as it directly affects the quality and reproducibility of derived analyses. To ensure consistency across diverse data sources, UGLC embeds detailed normalised information for each landslide event into 18 well-defined attributes. These attributes are systematically organised into seven principal categories that form the backbone of UGLC’s ontological structure:

1. **UGLC Reference:** uniquely identifies the event record within the new catalogue;
2. **Source Reference:** uniquely identifies the original event record from the source catalogue, allowing record tracing;
3. **Spatial Accuracy:** describes the spatial accuracy and geographic location of the record;
4. **Temporal Accuracy:** describes the temporal accuracy of the record;
5. **Geological Information:** describes the kinematic characteristics and physical factors of the recorded phenomenon;
6. **Reliability:** describes a relative confidence ranking based on spatial and temporal precision;
7. **Notes and information:** additional information about the record type, source, authors, related scientific publication and other linked information.

A complete list of attributes, along with detailed descriptions and classification logic, is provided in Table 3.

Attribute name	Type	Description
WKT_GEOM	WKT	Stores georeferencing as vector geometries (POINT or POLYGON/MULTIPOLYGON) in WGS84 (EPSG:4326) for GIS compatibility.
NEW DATASET	String	UGLC ID name.
ID	Int	Unique ID generated for each landslide record.
OLD DATASET	String	Name of the original dataset.
OLD ID	Int	ID used in the original dataset (if available).
VERSION	String	Latest version of the original dataset.
COUNTRY	String	Country name derived from data source or coordinates.
ACCURACY	String	Estimated coordinate precision in metres (if possible).
START DATE	Date	Record date (YYYY/MM/DD). If missing, the earliest reliable date is used. Defaults to 1678/01/01 if absent due to Pandas limitations.
END DATE	Date	Latest known date of the record.
TYPE	String	Geological/kinematic landslide type, standardised using Varnes classification (Hungr et al., 2014) (Table 5).
PHYSICAL FACTORS	String	Physical factors affecting the landslide (Preparatory - P, Triggering - T) (Table 6).
RELIABILITY	Int	Reliability class based on spatial (ACCURACY) and temporal (START DATE, END DATE) accuracy (Table 4).
RECORD TYPE	String	Report or event. Reports provide more technical details.
FATALITIES	Int	Number of fatalities.
INJURIES	Int	Number of injuries.
NOTES	String	Additional record-related notes.
LINK	String	URLs to sources, studies, or extra details.

Table 3. Detailed description of all attributes in UGLC catalogues.

As for Spatial Accuracy, the COUNTRY field identifies the sovereign state where the landslide occurred. To avoid ambiguity, administrative subdivisions such as “states” or “provinces” are encoded in a separate field. The country name is taken directly from the source data when available and is otherwise derived from coordinates. The ACCURACY field quantifies the estimated spatial error in metres, where such information can be retrieved or inferred. In records where it could not be estimated, it was set as -99999.

Temporal information is managed through the START DATE and END DATE attributes, which store the occurrence date or estimated time window of the event.

Additionally, each record receives a RELIABILITY score (0-10) that weights spatial and temporal accuracy to evaluate overall trustworthiness (see Table 4).

The RELIABILITY attribute is defined as a categorical ranking ranging from Class 1 (highest reliability) to Class 10 (lowest reliability), based on the spatial and temporal precision of each record. Classes are assigned according to predefined thresholds of spatial accuracy and temporal detail, with higher classes reflecting more precise and complete information. This attribute is intended as a practical indicator to support data filtering and selection, rather than a quantitative measure of uncertainty. It does not represent a statistically validated estimate of positional or temporal error. The classification scheme is based on commonly used ranges of spatial precision in geospatial data and reflects typical levels of accuracy associated with landslide mapping practices.

Point-based records can span the full RELIABILITY range (1–10), reflecting variability in both spatial and temporal precision, while polygon geometries are inherently spatially accurate and are therefore limited to classes 1 or 2, where RELIABILITY is primarily determined by temporal precision. This parameter can serve as a filter for users modelling with specific data quality requirements. Classes 1 to 4 offer the most reliable data for detailed analysis and modelling.

The attribute "TYPE" follows an ontology based on the extended Varnes classification (Hungr et al., 2014), incorporating additional ground instability phenomena, already present in the native datasets.

As for PHYSICAL FACTORS, these are classified according to a standardised framework, which distinguishes between preparatory and triggering factors (Cogan and Gratchev, 2019; Steger et al., 2023).

4 Methodology

The UGLC development workflow comprised three main steps:

- **Download of open-access landslide datasets**
  Landslide open data (landslide datasets, catalogues and inventories) from credible sources were selected for download only if licenses permitted data modification for analysis and reuse;

- **Data normalisation procedures**
  A detailed analysis provided a uniform and appropriate ontology for standardising all collected data. It led to the development of dataset-specific scripting procedures and their integration into UGLC framework;

Spatial Reliability	Temporal Precision	Reliability Description	Class
<100 m	TRUE	Exact point	1
<100 m	FALSE	Almost exact point	2
>100 m, <250 m	TRUE	Very high reliability point	3
>100 m, <250 m	FALSE	High reliability point	4
>250 m, <500 m	TRUE	Medium reliability point	5
>250 m, <500 m	FALSE	Low reliability point	6
>500 m, <1000 m	TRUE	Very low reliability point	7
>500 m, <1000 m	FALSE	Poor reliability point	8
>1000 m	TRUE/FALSE	Uncertain reliability point	9
-99999	TRUE/FALSE	Unreliable point	10

Table 4. Decision table for the reliability record ranking based on spatial accuracy attribute ("ACCURACY") and temporal accuracy attributes (where "START DATE" = "END DATE").

Landslide Categories		
Complex	Earth slide	Mudslide
Soil creep	Riverbank collapse	Rock slide
Debris flow	Rock fall	Rotational sliding
Earth flow	Translational sliding	Earth spreading
Lahar	Rock spreading	Mud flow
Sinkhole	Not defined	

Table 5. The kinematic type of each landslide record, standardised using the extended classification of Varnes, together with other surface instability phenomena (Hungr et al., 2014).

- **Data merging and formatting**

  In the final phase, the standardised data were integrated to generate both catalogues in their respective file formats and tile-based partitions to facilitate access to geographically segmented information.

Physical Factors	**UGLC id**
Rainfall activity	rainfall (T)
Seismic activity	seismic (T)
Volcanic activity	volcanic (T)
Human-induced factors	human (T,P)
Climatic factors	climate (T,P)
Post-fire conditions	postfire (P)
Post-deforestation processes	deforestation (P)
Erosional and biological factors	natural (T,P)
Not Defined	ND

Table 6. List of physical factors contributing to landslide activation, categorised as triggering (T) or preparatory (P) factors.

4.1 Data collection

The collection of landslide data was particularly challenging due to the extreme variability of the information within. Native landslide dataset availability, quality, structure, and detail level vary depending on the purpose for which they were collected, often linked to the geographical asset and social context. The data implementation workflow was designed to be highly adaptable, ensuring full compatibility with all native dataset formats (CSV, Shapefile, or Geopackage) and attribute structures. All datasets integrated in UGLC are openly accessible and can be directly downloaded from their respective repositories (or requested from the data provider, as in the case of the BGS National Landslide Database (Pennington et al., 2015)).

4.2 Data conversion

A preliminary analysis of the datasets revealed heterogeneity and ambiguity, such as inconsistencies in nomenclature, units of measurement, and data formats. For example, a single variable may have had different names, structures, or coding across several datasets, which then required a normalisation process to ensure data consistency and interoperability.

To address this issue, dataset-specific JSON lookup tables were constructed after a deep data exploration to convert non-standard values into a uniform format, and transformation scripts developed with Python libraries to facilitate data normalisation, where needed. The standardised parameters were:

– Country: Country-of-origin information was introduced through the “assign country to points” function, and countries were linked to the respective event coordinates;

– Accuracy: A qualitative estimate was added based on the information present in each event record;

- Date: All dates were converted to the YYYY/MM/DD format. Events with null dates were assigned the date that coincided with the year of the latest dataset version, and dates before 1678 were assigned the default of 1 January 1678. The latter overcame the timestamp limits of the Python Pandas library, because timestamps are typically stored as 64-bit integers and represent a maximum time span of approximately 584 years, specifically from the year 1677 to 2262. Beyond this range, the values exceed the storage capacity of a 64-bit integer (development team, 2020);

- Type: Landslide categories were standardised using a reference table;

- Physical Factors: A distinction between preparatory and triggering factors were standardised using a reference table;

For the management of missing data, i.e., no data or null data, two forms were chosen based on the type of variable.

- "ND" for all null string ("String") variables

- "-99999" for all null integer ("Int") values.

4.3 Semantic standardisation and attribute harmonisation

Semantic standardisation was performed through a harmonisation workflow aimed at ensuring structural and semantic consistency while preserving the informational content of the original datasets. Due to the significant heterogeneity of source inventories, the process focused on semantic alignment rather than observational homogenisation. This approach aims to minimise interpretative bias and reduce the risk of semantic misclassification across heterogeneous inventories.

The attribute conversion was governed by lookup tables encoded in JSON format, which define reproducible attribute re-mapping rules between native dataset fields and the UGLC schema. These lookup tables act as preservative conversion dictionaries, prioritising semantic equivalence and information conservation, while avoiding inference when metadata were incomplete or ambiguous.

The lookup tables were designed following an extensive manual data exploration phase. For each inventory, native attributes were systematically analysed to identify the full set of unique classification elements, including inconsistencies, alternative terminologies, and typographical variations. This semantic attribute re-mapping showed how the information was organised in each dataset, rather than forcing it into a predefined classification scheme.

Any interpretative approach was applied only when sufficiently supported by metadata.

As an illustrative example, heterogeneous landslide classification schemes were mapped onto the UGLC TYPE attribute through semantic correspondences, while triggering factors were harmonised within the PHYSICAL FACTOR field when explicitly provided by the source datasets.

Each dataset-specific JSON lookup table is located within the corresponding repository folder and is publicly available, ensuring full methodological transparency and reproducibility.

- Points catalogue datasets folders: https://github.com/UnibaGEO/UGLC_point/tree/master/csv

- Polygons catalogue datasets folders: https://github.com/UnibaGEO/UGLC_poly/tree/master/csv

4.4 Data merge

Before any merging operation, all individual records from the various source datasets were consolidated by standardising their formats and harmonising attribute definitions by creating a uniform data structure in which each record was assigned a unique ID to maintain traceability.

Thereafter, duplicate records were identified using a strict matching criterion based on identical spatial coordinates and identical temporal attributes. These cases were interpreted as informational duplicates, where the same landslide event is reported across multiple datasets. When duplicates were detected, the most informative record was retained, based on attribute completeness. In cases of identical records, only one entry was preserved. This conservative approach prioritises traceability and reproducibility over aggressive duplicate removal and does not account for near-duplicate events.

Moreover, to ensure scientific consistency in representing the phenomena of landslide and surface instability, all events classified as “avalanche", “liquefaction”, and “glacial lake outburst floods” were also excluded, as these processes differ kinematically and geologically from other gravitational phenomena such as landslides. Four output files in two different formats were produced as a result of the merge:

1. Point dataset called *UGLC_point* in CSV format;

2. Point dataset in GeoPackage format, split into 105 tiles that cover the entire Earth’s surface. Each tile is uniquely identified within the global grid by a pair of variables i and j, placed at the end of the filename *UGLC_point_i_j*. Empty tiles are automatically excluded from storage, thus ensuring optimised file management and performance;

3. Polygonal dataset called *UGLC_poly* in CSV format;

4. Polygonal dataset in GeoPackage format, split into 105 tiles that cover the entire Earth’s surface. As above, each tile is uniquely identified within the global grid by a pair of variables i and j, placed at the end of the filename *UGLC_poly_i_j*. Empty tiles are automatically excluded from storage, thus ensuring optimised file management and performance.

Tables 1 and 2 list the datasets used and the total number of validated landslide events included in the catalogue. UGLC polygonal dataset contains 984126 landslide records, a number that remained unchanged as no duplicate or nonconforming entries were found for this dataset. In contrast, the UGLC point dataset was refined during the validation process, resulting in a total of $1\,061\,450$ landslide point records (Fig. 2) .

4.5 Data normalisation

The lack of a uniform framework poses significant challenges to the integration and reuse of heterogeneous datasets, leading to potential inconsistencies in the interpretation and application of data. Therefore, in this study, the multi-source data were normalised within a common standard for landslide studies to ensure compatibility for wide-ranging applications. The harmonisation process involved detailed analysis and expert judgement to classify key attributes. This critical step produced a coherent and comprehensive representation of landslide records worldwide. The following categories were standardised during the process:

- Landslide Type classification: Based on the Varnes extended version (Hungr et al., 2014), this distinguishes six main types of movement according to the nature of the material involved (rock, debris, soil) and the mode of movement (falling, toppling, sliding, flowing, lateral expansion, subsidence). However, in order to better represent the variety and complexity of the phenomena observed in the source catalogues, it was necessary to assimilate similar categories (For example, multiple source labels such as 'Rock Fall', 'rockfalling', "Rock Topple" and 'RF' were mapped into a unified 'rock fall' class) and include other gravitational ground instabilities phenomena (like sinkholes) in this classification.

  Additionally, expert interpretation and corrected re-interpretation of certain types of classification resolved the issue of fragmentary information found in some records. For example, specific phenomena such as riverbank collapse, often omitted or incorrectly categorised from traditional classifications, were included to represent these commonplace events in riverine contexts (Figure5).

- Physical Factors: These represent phenomena that either create preparatory conditions for landslides or actually trigger them. The factors are classified into two functional categories: preparatory (P) and triggering (T) (See Table 6). In most of the source catalogues, information on physical factors was either incomplete or absent and thus called for a careful evaluation and normalisation of data. Where possible, fragmentary information was traced back to reconstruct the missing information in the correct context. For example, the associated physical factor for data on seismic-induced landslides was set as "seismic (T)". However, where original records did not provide reliable information or references on the relevant physical factors, the category was set as "Not Defined" (ND).

- Record date: Record date normalisation was done to ensure data consistency and completeness, including when native date formats were either varied or incomplete. Date information was converted to "YYYY/MM/DD" format, and the following date attribute standardisation criteria were applied:
  - If the date attribute in the native dataset contained a full date (year, month, day), the format was assigned for both "START DATE" and "END DATE" attributes;
  - If the date attribute in the native dataset contained only year and month, "START DATE" attribute was setted as the first day of that month, while "END DATE" as the last day of that same month;
  - If the date attribute in the native dataset contained only the year, "START DATE" attribute was set as the first day of that year, while "END DATE" as the last day of that same year;
  - If no reliable date attribute in the native dataset was available, "START DATE" attribute was set as the start date of acquisitions for that dataset or to "1678/01/01" for no-data(due to Pandas library limitation (development team, 2020)), while "END DATE" attribute was set as the the end date of native dataset data acquisitions.

- Accuracy: The Accuracy attribute was normalised to standardise the spatial coordinate accuracy of each landslide record to estimate in metres the relative deviation between the landslide coordinate position in the record and the actual ground-truthed position of the landslide. To assess this a standardisation criterion was followed to convert or infer accuracy information:
  - In cases where source records specified a numerical value for spatial accuracy, the value was reported or converted in

metres, without any further operations; - Where native records did not provide any numerical value a expert interpretation of metadata was employed to estimate probable accuracy in instances of incomplete or ambiguous location data. The process involved cross-referencing auxiliary information, such as nearby landmarks or descriptive metadata, to determine coordinates that closely approximated the event's actual location (for example, using a text description of acquisition processes like "detected by helicopter" or "digitalised from historical maps" etc.); - Where records did not provide any kind of reference on spatial accuracy, the attribute was not inferred or interpreted, and instead filled with "-99999" as no-data.

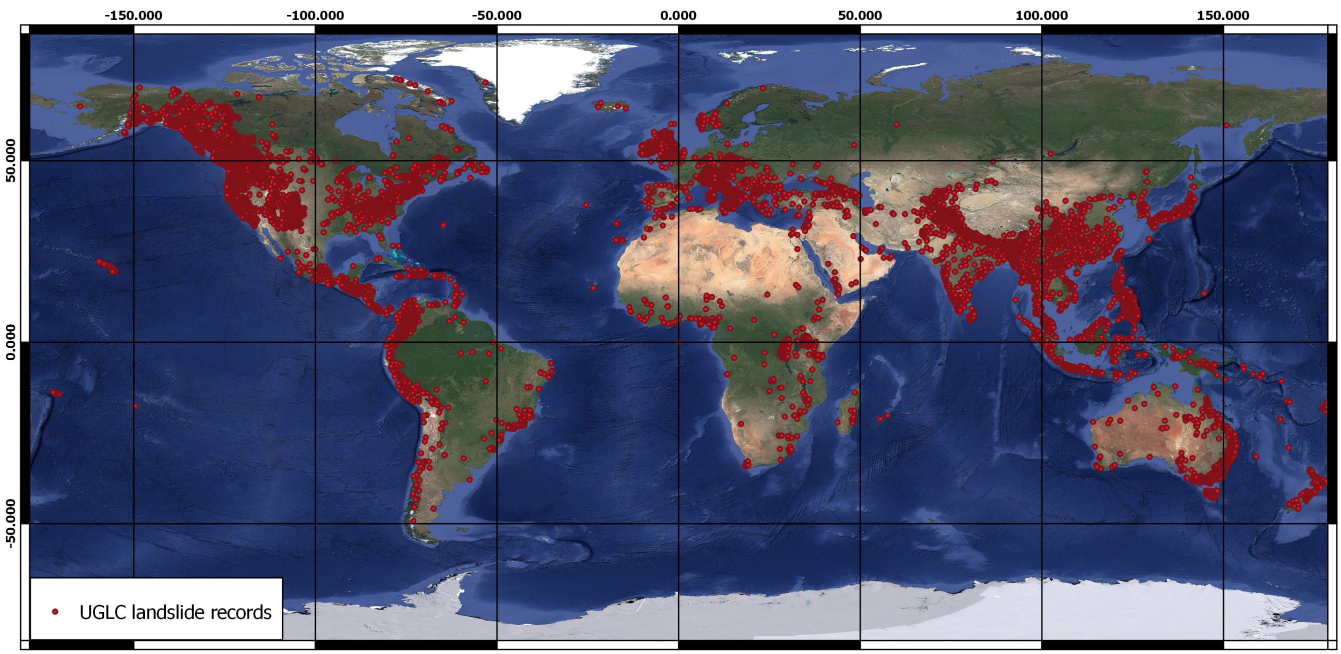

Figure 1. Global distribution of landslide record points contained in the UGLC point catalogue (Mancino et al., 2025). on a global basemap (Google Satellite imagery ©2025 CNES / Airbus, Maxar Technologies, Google).

5 Results

5.1 Analysis and statistical insights on UGLC

A major challenge in creating such a large standardised catalogue was achieving information semantic and syntax consistency while condensing significantly heterogeneous data. Extensive inconsistencies in nomenclature are often a barrier to data analysis, especially in fields such as geology. The adopted point and polygon data schema ensured a consistent integration of all

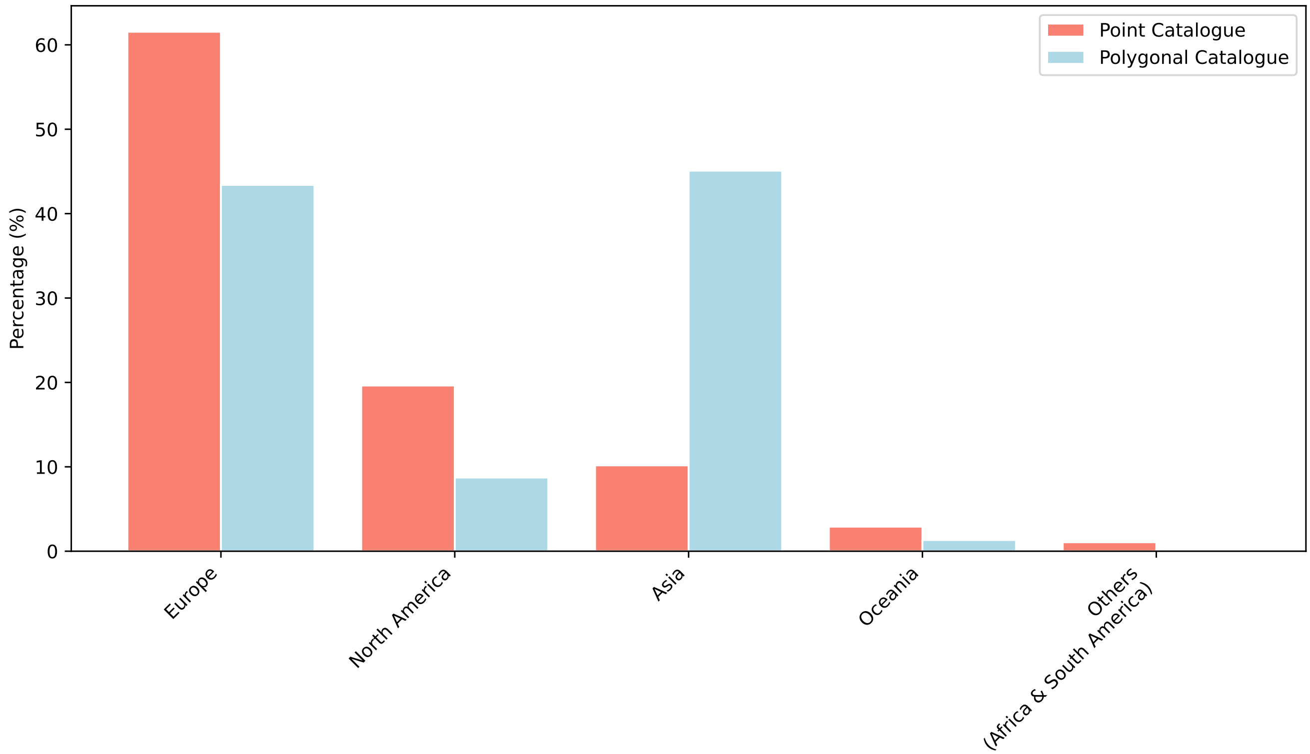

Figure 2. UGLC Data Distribution per Continent for both Point and Polygonal Catalogue

datasets while preserving their original geometry types. Thereafter, several statistical analyses were conducted to better understand the content of both catalogues (point and polygonal) and to explore key aspects of the constituent data. This assessment was essential for setting out the framework for appropriate and targeted use of the catalogues and highlighting the potential for future scientific development. In particular, the analysis identified a pronounced disparity in the geographical distribution of landslide records across continents (Figures 1, 2).

Within the point catalogue, Europe exhibits the highest proportion of globally recorded events (61.55%) followed by North America (19.63%) and Asia (10.17%). Africa, South America and Oceania collectively constitute a very low share (below 3.97%). Meanwhile, the polygonal catalogue presents a different distribution pattern, with Asia leading (45.09%) and followed by Europe (43.40%) and North America (8.73%). Here too, Africa, South America, and Oceania collectively constitute a miniscule proportion of recorded events (1.43%). This imbalance becomes more apparent by going into more detail with a state-by-state analysis, as the landslide records appear in an unbalanced distribution in density and across geography (Figure 3).

In the state-wise density data, it is particularly evident that some relatively small countries like Italy, UK, and New Zealand lead landslide data collection alongside large countries such as the USA and China (and sometimes markedly surpass them, as is the case of Italy, which alone contributes to more than 57% of the entire catalogue's data). This data distribution evidences greater attention to landslide phenomena in frequently afflicted countries. (Stanley and Kirschbaum, 2017).

Despite the high density of studies in certain landslide-prone regions, data accessibility remains a significant limitation, particularly due to restricted or non-public datasets in various areas, including parts of Europe and Asia. This limited access restricted the scope of the analysis, regardless of the spatial distribution or availability of published research.

Significant variations in time coverage between datasets are also evident. Addressing these discrepancies in formatting and granularity was a significant effort (Figure4). Addressing temporal inaccuracies involved finding exact event dates and correcting broader temporal ranges (e.g., decades or centuries). For records with incomplete or poorly formatted temporal data, representative time ranges were assigned based on the available context to ensure logical alignment with the recorded phenomena. The approach not only improved temporal consistency, but also enhanced the utility of the catalogue by preserving valuable, albeit imprecise, historical data. Providing temporal consistency in the catalogue also mitigated the risk of data misinterpretation from inconsistent native data formats.

Spatial accuracy is a similarly critical factor in cataloguing landslide phenomena, as it determines the geospatial reliability of each record (Figure 5). Native datasets often present difficulties, including poorly formatted coordinates, varying levels of spatial precision, and inconsistencies in georeferencing methods. The UGLC accuracy distribution ranges from highly precise values (<10 metres) to broader approximations (>10 kilometres), and reflects the inherent variability in the quality and reporting practices of the source (Figure 5). As a corollary measure to the standardisation processes, the catalogue's reliability attribute aids in reflecting the general robustness of individual records.

Both the point and polygon catalogues exhibit notably high record reliability, with the majority of entries falling within classes 1 and 2 (see Table 4). In the point catalogue specifically, the records with lower reliability classes collectively represent less than 15% of the total (Figure 6). Nevertheless, the whole spatial and temporal standardisation process forms a reliable framework, which is depicted by the reliability class parameter.

The data analysis of UGLC also presented the spatio-temporal density and landslide type distributions (Figure 7). These data asymmetries are due to the influence of the extremely heterogeneous datasets that were consolidated into standardised landslide types. The integrated product reflects the harmonisation of various terminologies, classification schemes, and granularity, as well as the recovery of valuable data from erroneous event logs.

Landslide types having a significantly higher variance, such as "complex" or "earth flow", indicate the presence of a higher rate of expert-interpreted data compared to those with natively more unambiguous and consistent types, which are therefore easily interpreted, such as "rockfall" or "sinkhole". It was also possible to analyse the distribution of various physical factors associated with each recorded landslide (Figure 8). The graph reveals a higher prevalence of missing information about physical factors, followed by Triggering factors (T) and Preparatory factors (P).

The dominance of common trigger factors, such as rainfall and seismic activity, highlights the statistical prevalence of these phenomena in native datasets. These patterns reflect both natural processes and reporting biases associated with data availability and observation practices. However, this distribution of factors is also influenced by the uneven geographical coverage of the data. Areas of frequently documented landslides also tend to be those where certain triggering factors are more prominent. This underscores the need to address spatial heterogeneity in future data collection to enhance global representativeness.

Data completeness is inherited from the original datasets and is generally not explicitly documented. Therefore, the absence of records in a given area should not be interpreted as the absence of landslide occurrence.

Analysis of the overall distribution of the various types of landslide standardised in UGLC shows how the frequencies of each type of landslide vary widely throughout the world (Figure 9). The majority of landslides in the graph fall under undefined categories ("ND"), meaning that native datasets lack information on the kinematics for each landslide record. Nevertheless, for non-null categories, the types "complex", "earth slide", "rock fall" and "soil creep" are the most prevalent, while types such as "lahar" and "earth spreading" are minimally represented.

The dataset's spatial variability is apparent through dense clusters found in extensively studied regions, such as South Asia and Central America, which are known for their climatic and geological activity. Conversely, areas such as Africa and Russia are under-represented because of data gaps, likely resulting from mapping challenges or limitations in data access. These analyses underscore the challenge of worldwide data availability and uneven data resolution, which UGLC strives to overcome.

To further explore the spatial distribution of landslide occurrences, UGLC point data were qualitatively compared with global rainfall patterns and seismic hazard. Rainfall intensity was derived from monthly precipitation data based on the GPM IMERG product (Huffman et al. (2019)), aggregated over the period 2000–2025 (Figure 10). Seismic hazard was represented using peak ground acceleration (PGA) from the Global Earthquake Model (GEM; Pagani et al. (2020)), corresponding to a 10% probability of exceed in 50 years under reference site conditions (Vs30 = 760–800 m/s) (Figure 11).

This comparison is intended as an exploratory assessment rather than a quantitative validation. The observed spatial distribution of landslides shows clear clustering in regions characterised by high rainfall intensity and elevated seismic hazard, consistent with the dominant triggering mechanisms represented in the dataset. At the same time, these patterns reflect not only geomorphological controls but also systematic reporting and data availability biases. Regions subject to frequent rainfall or earthquake-induced landslides are more extensively monitored and documented, leading to higher data density in the catalogue. Conversely, areas with limited monitoring infrastructure or lower reporting capacity remain under-represented.

5.2 Epistemological limitations and intended use

Given the scale and heterogeneity of the data integrated within the UGLC, it is essential to explicitly acknowledge its epistemological limitations and define appropriate domains of use.

The compilation of landslide datasets from multiple sources presents inherent challenges, primarily due to incomplete or inconsistent metadata, variability in parameter nomenclature, and the absence of universally adopted classification standards for landslide types and associated physical triggers. These issues reflect the diversity of scientific, operational, and institutional contexts in which the original datasets were produced.

The UGLC integrates datasets developed under heterogeneous conditions and for different purposes. As such, it does not constitute a scientifically homogeneous dataset in an observational sense. The harmonisation process standardises the representation of landslide information, but does not eliminate differences in mapping strategies, data acquisition objectives, spatial completeness, temporal coverage, or classification criteria.

Consequently, the UGLC should be interpreted as an organisational and semantic integration layer rather than a uniform analytical dataset. Its primary function is to enable structured access to globally distributed landslide information while preserving the provenance and contextual meaning of the original records.

The catalogue is well suited for data discovery, exploratory analyses, and large-scale descriptive assessments. However, its use in quantitative modelling or predictive applications requires careful, context-aware filtering based on dataset provenance, landslide typology, triggering mechanisms, and data reliability. The dataset should therefore not be treated as a statistically homogeneous input without appropriate preprocessing and domain-specific considerations. It is also important to note that data completeness is inherited from the original datasets and is rarely explicitly documented. As a result, the absence of records in a given region should not be interpreted as the absence of landslide occurrence.

As also highlighted by the comparison with global rainfall and seismic hazard datasets (Section 5.1), the spatial distribution of UGLC events reflects both triggering mechanisms and reporting biases, and should therefore be interpreted with caution in quantitative analyses because they may can affect frequency statistics, temporal interpretations, and predictive models training.

Despite the integration of multiple global and regional inventories, significant spatial imbalances remain. Large areas, like parts of Africa, South America and North Asia areas are under-represented due to a lack of monitoring infrastructure, logistical constraints, and disparities in data availability and reporting practices. Furthermore, variations in national data-sharing policies and institutional transparency continue to restrict access to landslide information at the global scale.

The RELIABILITY attribute represents a practical classification of spatial and temporal precision but is not a validated accuracy metric. A rigorous validation would require comparison with independent reference data, such as multi-temporal satellite imagery, but this is challenging at the global scale due to dataset heterogeneity, broad temporal coverage, and uneven data availability. A comprehensive validation would therefore require a dedicated effort beyond the scope of this study. Consequently, RELIABILITY should be interpreted as a pragmatic filtering indicator to support appropriate data use.

These limitations highlight the broader challenge of developing globally consistent geohazard datasets. Many existing landslide inventories are designed for regional or national applications and are often not interoperable due to differences in metadata standards and data management practices.

UGLC does not aim to resolve these inconsistencies but provides a structured framework that facilitates interoperability while explicitly preserving data heterogeneity. The workflow and data model presented in this study can therefore serve as a reference for future efforts toward the standardisation of landslide data, both at global and regional scales.

6 Future development

The ongoing evolution of UGLC requires a plan of continuous improvement and expansion. To ensure its long-term relevance, consistency, and reliability, future developments will focus on two key aspects:

- a yearly update check.

– future implementation of a WebGIS portal.

6.1 Yearly update check

A systematic yearly update will be performed to maintain the integrity and comprehensiveness of UGLC. This process will include:

– Existing data verification:
  All native landslide datasets integrated into UGLC will be reviewed for updates. Any modifications, corrections, or additions made by the original data providers will be incorporated into the latest version.

– New Sources Identification:
  An annual check will be performed to identify any newly available landslide datasets that were not included in the previous version. These may come from government agencies, research institutions, or independent projects.

– Crowd-sourced and Scientific Data Incorporation:
  UGLC will leverage data submissions from diverse sources, such as scientific journals, universities, government agencies, and citizen scientists. Reports submitted through the dedicated Google Form (UGLC new point data integration) will be evaluated for accuracy, reliability, and relevance before integration into the catalogue. We encourage organisations and agencies that manage landslide occurrences to complete the form for each landslide event. For multiple event entries, we suggest contacting the corresponding authors to facilitate the incorporation of new datasets into the UGLC.

6.2 Future WebGIS portal

Future developments will also focus on creating a dedicated WebGIS portal to improve accessibility and usability for visualisation and new data entry. This platform will facilitate:

– Optimised Data Submission:
  A structured and user-friendly interface will allow for easier submission of new landslide records, including both point-based and polygonal data, in accordance with UGLC standards ensuring consistency across all submitted records;

– Interactive Data Visualisation:
  Users will be able to explore UGLC through an interactive map, providing intuitive filtering and querying functionalities.

6.3 OpenStreetMap crowdsourcing

To further improve landslide data update efforts, UGLC will leverage the OpenStreetMap (OSM) framework and its community for assisted photo-interpreted mapping of landslide polygons. The initiative aims to:

– Encourage community participation:
  Engage the OSM community in identifying and mapping landslide-affected areas using satellite imagery and other geospatial datasets.

– Facilitate annual data integration:
  Landslide polygons mapped through OSM will be collected each year and incorporated into the annual update, similar to UGLC point data collected via the Google Form.

– Enhance Mapping Accuracy:
  The initiative will improve landslide detection and delineation by incorporating a large community of contributors who will enrich the quality of the catalogue.

6.4 Landslide models and predictive applications

UGLC provides a structured and interoperable framework for integrating heterogeneous landslide information at the global scale. While the dataset supports exploratory analyses and large-scale investigations, its application in predictive modelling requires careful context-specific filtering strategies and domain-specific interpretation. The catalogue structure, together with attributes describing landslide typology and physical triggering factors where available, enables users to derive analytically consistent subsets suited to specific modelling objectives.

Increasingly available high-resolution geo-environmental datasets, including terrain, hydrological, and soil variables can supplement such modelling applications, and UGLC is designed to be used with these. Datasets such as SoilGrids250m (Poggio et al., 2021), Geomorpho90m (Amatulli et al., 2020), Hydrography90m (Amatulli et al., 2022), among others, provide geomorphological and environmental descriptors commonly associated with landslide occurrence.

UGLC was indeed conceived as a harmonised landslide information framework designed to facilitate large-scale analyses, comparative investigations, and reproducible modelling experiments. While the catalogue does not eliminate the intrinsic heterogeneity of the original inventories, it provides a consistent semantic and structural basis that can support global-scale exploratory and predictive studies when used with appropriate methodological caution.

7 Conclusion

To date, open access global landslide data has inhibited probabilistic continental to global model performance in terms of both detail and geospatial precision (Loche et al., 2022). Data heterogeneity introduces limitations in such models, as they are unlike traditional local-scale ones derived from more robust and in situ data (Reichenbach et al., 2018). To address this fundamental challenge associated with data heterogeneity, UGLC development focused on two principal goals. First, the creation of a single, standardised, and accessible landslide catalogue sought to simplify the global organisation and interpretation of these phenomena by reducing structural and semantic inconsistencies characteristic of most open landslide datasets. This includes enhancing the interpretability of landslide data and enabling systematic analyses of the relationships between environmental features, climatic conditions, and societal impacts. The second goal was to provide a harmonised foundational catalogue designed to support landslide analyses and modelling-oriented investigations, including those employing Machine Learning and emerging AI methodologies (Pradhan et al., 2010; Kirschbaum et al., 2016; Reichenbach et al., 2018; Tehrani et al., 2022; Shang et al., 2023).

In pursuit of these objectives, the design and content of UGLC provide a structured and semantically consistent framework suited for large-scale descriptive analyses, comparative investigations, and dataset-specific modelling experiments. UGLC facilitates the development of global landslide monitoring strategies and wide-ranging susceptibility and risk studies, while preserving the intrinsic heterogeneity of the original inventories. The methodology employed for UGLC also represents a basis for improving structural consistency and interoperability among landslide records. It is inherently a scalable framework that will support the evolution of new landslide datasets and that of UGLC itself.

Presently, UGLC offers a global foundation for landslide data sharing and analysis. In this initial release, only the point catalogue contains integrated data and invites new contributions via the following Google form: UGLC new point data integration. UGLC's ongoing expansion is intended to foster international collaboration and promote the open exchange of geospatial landslide information through the implementation of standardised formats, seeking the progressive improvement of the consistency, documentation, and level of detail in recorded data, while minimising redundancy. To ensure its continuous development, UGLC encourages contributions from a broad network of collaborators, including scientific journals, research institutions, government agencies, and citizen crowd-sourcing to gather daily event reports, using an easily accessible user-friendly mode.

8 Data availability

The complete UGLC dataset is openly available through Zenodo at https://doi.org/10.5281/zenodo.18643456 (Mancino et al., 2025), where it is distributed as tiled GeoPackage files together with a global CSV version. The archived release is assigned a persistent DOI and represents the version described in this paper.

The UGLC point and polygon datasets are distributed under the Creative Commons Attribution–NonCommercial 4.0 International (CC BY-NC 4.0) licence. The licence information is included in both the Zenodo archive and the GitHub repositories. Users are requested to cite the Zenodo DOI when using the dataset.

9 Code availability

The Python source code used to generate the UGLC dataset is openly available on GitHub:

- Points catalogue:
  https://doi.org/10.5281/zenodo.20925165

- Polygons catalogue:
  https://doi.org/10.5281/zenodo.20925213

The repositories contain all processing scripts required to reproduce the UGLC generation workflow. Raw input datasets are not redistributed because they originate from third-party sources with different licensing conditions. Users should obtain these datasets directly from their original providers before running the workflow.

Author contributions. SM and GA designed the study. SM and AS developed, implemented, and benchmarked the Python workflow and processing chain for the datasets harmonisations. AS managed all dataset permit request gathering from the owners. FL supervised and reviewed the Python code and contributed to defining the overall architecture and quality standards of the repository. FL, SM and DC were responsible for the ontology behind data standardisation. SM and GA were responsible for data analysis and graphics production. TS was responsible for editing and finalising the manuscript. SM and AS wrote the first draft of the manuscript, and all authors contributed to the writing of the manuscript and discussed the results.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. The authors accept no responsibility for any liability arising from the use of this research paper and its associated dataset. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. The authors thank all those involved in the development and dissemination of the datasets used in this work, as cited in the appendix A and B. The authors thank Dr. Ivan Marchesini from IRPI-CNR Perugia-Italy for his valuable discussions and feedback during the implementation of the catalogue, and Dr. Rosa Colacicco for her valuable support in the development of the criteria for the conversion ontology of physical factors. Our thanks also go to Margosa Environmental Solutions for its in-kind contribution through TS on user networks and application expertise.

Financial support. This research was supported by multiple institutions. GA acknowledges the financial support provided by Yale University, School of the Environment. SM's Ph.D. scholarship (in the framework of D.M. 117/2023 PNRR Investment 3.3) and AS' research fellowship were co-funded by Planetek Italia and the Italian National Recovery and Resilience Plan (PNRR) – both institutions are focused on advancing studies in landslide risk management.

References

Amatulli, G., McInerney, D., Sethi, T., Strobl, P., and Domisch, S.: Geomorpho90m, empirical evaluation and accuracy assessment of global high-resolution geomorphometric layers, Scientific Data, 7, 162, 2020.

Amatulli, G., Domisch, S., Tuanmu, M., Parmentier, B., Ranipeta, A., Malczyk, J., and Jetz, W.: Hydrography90m: A new high-resolution global hydrographic dataset, Earth System Science Data, 14, 4525–4550, 2022.

Australia, G.: Australia Landslide Catalogue, https://oasishub.co/dataset/australia-landslide-catalogue, 2016.

Bhuyan, K., Tanyaş, H., Nava, L., Puliero, S., Meena, S., Floris, M., van Westen, C., and Catani, F.: Generating multi-temporal landslide inventories through a general deep transfer learning strategy using HR EO data, Scientific reports, 13, https://doi.org/10.1038/s41598-022-27352-y, 2023.

Bragagnolo, L., Rezende, L., da Silva, R., and Grzybowski, J.: Japan landslide dataset for semantic segmentation, https://doi.org/10.5281/zenodo.3775870, 2020.

Brideau, M.-A., Lau, C.-A., Brayshaw, D., Lipovsky, P., Cronmiller, D., and Friele, P.: Preliminary Canadian Landslide Database, https://doi.org/10.5281/zenodo.10799126, 2024.

Cogan, J. and Gratchev, I.: A study on the effect of rainfall and slope characteristics on landslide initiation by means of flume tests, Landslides, 16, 2369–2379, https://doi.org/10.1007/s10346-019-01261-0, 2019.

Depicker, A., Jacobs, L., Delvaux, D., Havenith, H.-B., Maki Mateso, J.-C., Govers, G., and Dewitte, O.: The added value of a regional landslide susceptibility assessment: The western branch of the East African Rift, Geomorphology, 353, 106 886, https://doi.org/10.1016/j.geomorph.2019.106886, 2020.

Depicker, A., Govers, G., Jacobs, L., Campforts, B., Uwihirwe, J., and Dewitte, O.: Interactions between deforestation, landscape rejuvenation, and shallow landslides in the North Tanganyika–Kivu rift region, Africa, Earth Surface Dynamics, 9, 445–462, https://doi.org/10.5194/esurf-9-445-2021, 2021.

development team, P.: pandas-dev/pandas: Pandas, https://doi.org/10.5281/zenodo.3509134, 2020.

Di Napoli, M., Carotenuto, F., Cevasco, A., Confuorto, P., Di Martire, D., Firpo, M., Pepe, G., Raso, E., and Calcaterra, D.: Machine learning ensemble modelling as a tool to improve landslide susceptibility mapping reliability, Landslides, 17, 1897–1914, https://doi.org/10.1007/s10346-020-01392-9, 2020.

DOGAMI: Statewide Landslide Information Database for Oregon (SLIDO), Release 4.5, https://www.oregon.gov/dogami/slido/pages/data.aspx, 2024.

Esposito, G. and Matano, F.: A geodatabase of historical landslide events occurring in the highly urbanized volcanic area of Campi Flegrei, Italy, Earth System Science Data, 15, 1133–1149, https://doi.org/10.5194/essd-15-1133-2023, 2023.

Ferrario, M.: Landslides triggered by the 2015 Mw 6.0 Sabah (Malaysia) earthquake: inventory and ESI-07 intensity assignment, Natural Hazards and Earth System Sciences, 22, 3527–3542, https://doi.org/10.5194/nhess-22-3527-2022, 2022.

Ferrario, M., Perez, J., Dizon, M., Livio, F., Rimando, J., and Michetti, A.: Environmental effects following a seismic sequence: the 2019 Cotabato -Davao del Sur (Philippines) earthquakes, https://doi.org/10.1007/s11069-024-06467-7, 2023.

Froude, M. and Petley, D.: Global fatal landslide occurrence from 2004 to 2016, Natural Hazards and Earth System Sciences, 18, 2161–2181, https://doi.org/10.5194/nhess-18-2161-2018, 2018.

Gariano, S. and Guzzetti, F.: Landslides in a changing climate, Earth-Science Reviews, 162, 227–252, https://doi.org/10.1016/j.earscirev.2016.08.011, 2016.

Gomez, D., Garcia, E., and Aristizábal, E.: Spatial and temporal landslide distributions using global and open landslide databases, Natural Hazards, 117, 22–55, https://doi.org/10.1007/s11069-023-05848-8, 2020.

Hearn, G. and Hart, A.: Chapter Five - Geomorphological Contributions to Landslide Risk Assessment: Theory and Practice, vol. 15, pp. 107–148, Elsevier, https://doi.org/10.1016/B978-0-444-53446-0.00005-7, 2011.

Herrera-Coy, M., Calderón, L., Herrera-Pérez, I., Bravo-López, P., Conoscenti, C., Delgado, J., Sánchez-Gómez, M., and Fernández, T.: Landslide Susceptibility Analysis on the Vicinity of Bogotá-Villavicencio Road (Eastern Cordillera of the Colombian Andes), Remote Sensing, 15, 3870, https://doi.org/10.3390/rs15153870, 2023.

Hovius, N. and Stark, C.: Landslide-driven erosion and topographic evolution of active mountain belts, vol. 49, pp. 573–590, Springer, https://doi.org/10.1007/978-1-4020-4037-5_30, 2006.

Huffman, G., Stocker, E., Bolvin, D., Nelkin, E., and Tan, J.: GPM IMERG Final Precipitation L3 Half Hourly 0.1° x 0.1° V06, NASA Goddard Earth Sciences Data and Information Services Center (GES DISC), https://doi.org/10.5067/GPM/IMERG/3B-HH/06, 2019.

Hungr, O., Leroueil, S., and Picarelli, L.: The Varnes classification of landslide types, an update, Landslides, 11, 167–194, https://doi.org/10.1007/s10346-013-0436-y, 2014.

Ireland, G. S.: Landslide Locations and Extents Ireland (ROI/NI) ITM, https://data.gov.ie/dataset/landslide-locations-and-extents-ireland-roini-itm, 2012.

Jaboyedoff, M., Michoud, C., Derron, M.-H., Voumard, J., Leibundgut, G., Sudmeier-Rieux, K., Nadim, F., and Leroi, E.: Human-Induced Landslides: Toward the Analysis of Anthropogenic Changes of the Slope Environment, pp. 217–232, CRC Press, ISBN 9781315375007, 2018.

Jiang, L. and Wang, X.: Dataset Constrution through Ontology-Based Data Requirements Analysis, Applied Sciences, 14, 2237, https://doi.org/10.3390/app14062237, 2024.

Juang, C., Stanley, T., and Kirschbaum, D.: Using citizen science to expand the global map of landslides: Introducing the Cooperative Open Online Landslide Repository (COOLR), PLoS ONE, 14, e0218 657, https://doi.org/10.1371/journal.pone.0218657, 2019.

Kirschbaum, D., Adler, R., Hong, Y., Hill, S., and Lerner-Lam, A.: A global landslide catalog for hazard applications: method, results, and limitations, Natural Hazards, 52, 561–575, https://doi.org/10.1007/s11069-009-9401-4, 2010.

Kirschbaum, D., Stanley, T., and Zhou, Y.: Spatial and temporal analysis of a global landslide catalog, Geomorphology, 249, 4–15, https://doi.org/10.1016/j.geomorph.2015.03.016, 2015.

Kirschbaum, D., Stanley, T., and Yatheendradas, S.: Modeling landslide susceptibility over large regions with fuzzy overlay, Landslides, 13, 485–496, https://doi.org/10.1007/s10346-015-0577-2, 2016.

Koneru, S., Badavathula, H., Vadttitya, P., and Kosaraju, S.: Landslide identification using convolutional neural network, pp. 416–423, CRC Press, ISBN 9781003471059, https://doi.org/10.1201/9781003471059-54, 2024.

Liu, S., Li, Y.-E., Wang, B., Cai, A.-D., Feng, C., Lan, H., and Zhao, R.-C.: Challenges and countermeasures for developing countries in addressing loss and damage caused by climate change, Advances in Climate Change Research, 15, 353–363, https://doi.org/10.1016/j.accre.2024.02.003, 2024.

Loche, M., Alvioli, M., Marchesini, I., Bakka, H., and Lombardo, L.: Landslide susceptibility maps of Italy: Lesson learnt from dealing with multiple landslide types and the uneven spatial distribution of the national inventory, Earth-Science Reviews, 232, 104 125, https://doi.org/10.1016/j.earscirev.2022.104125, 2022.

Luetzenburg, G., Svennevig, K., Bjørk, A., Keiding, M., and Kroon, A.: A national landslide inventory for Denmark, Earth System Science Data, 14, 3157–3165, https://doi.org/10.5194/essd-14-3157-2022, 2022.

Malet, J.-P., Hibert, C., Radiguet, M., Gautier, S., Larose, E., Amitrano, D., Jongmans, D., Bièvre, G., and RESIF: French Landslide Observatory – OMIV (Temporary data) (MT-campagne) (RESIF - SISMOB), https://doi.org/10.15778/RESIF.1N2015, 2015.

Mancino, S., Sblano, A., Lovergine, F., Sethi, T., Capolongo, D., and Amatulli, G.: Unified Global Landslide Catalogue (UGLC), https://doi.org/10.5281/zenodo.16755044, 2025.

Martinez, S., Allstadt, K., Slaughter, S., Schmitt, R., Collins, E., Schaefer, L., and Ellison, S.: Landslides triggered by the August 14, 2021, magnitude 7.2 Nippes, Haiti, earthquake, Tech. rep., U.S. Geological Survey, https://doi.org/10.3133/ofr20211112, 2021.

Massey, C., Townsend, D., Rosser, B., Morgenstern, R., Jones, K., Lukovic, B., and Davidson, J.: Version 2.0 of the landslide inventory for the Mw 7.8 14 November 2016, https://doi.org/10.17603/ds2-1ftv-hm22, 2021.

Mirus, B., Jones, E., Baum, R., Godt, J., Slaughter, S., Crawford, M., Lancaster, J., Stanley, T., Kirschbaum, D., Burns, W., Schmitt, R., Lindsey, K., and McCoy, K.: Landslides across the USA: occurrence, susceptibility, and data limitations, Landslides, 17, 2271–2285, https://doi.org/https://doi.org/10.1007/s10346-020-01424-4, 2020.

Mitsugi, H.: Foreword by Hiroto Mitsugi for the Journal of the International Consortium on Landslides, Landslides, 15, 2323–2324, https://doi.org/10.1007/s10346-018-1076-z, 2018.

Morales, B., Garcia-Pedrero, A., Lizama, E., Lillo-Saavedra, M., Gonzalo-Martín, C., Chen, N., and Somos-Valenzuela, M.: Patagonian Andes Landslides Inventory: The Deep Learning's Way to Their Automatic Detection, Remote Sensing, 14, 4622, https://doi.org/10.3390/rs14184622, 2022.

on Climate Change, I. P.: Climate Change 2022: Impacts, Adaptation, and Vulnerability, Cambridge University Press, Cambridge, UK and New York, NY, USA, https://doi.org/10.1017/9781009325844, 2022.

Pagani, M., Weatherill, G., Garcia-Pelaez, J., Crowley, H., Silva, V., Henshaw, P., Butler, L., Simionato, M., Vigano, D., Danciu, L., and Monelli, D.: Global Earthquake Model (GEM) Seismic Hazard Map (PGA, 10% probability of exceedance in 50 years), https://www.globalquakemodel.org, 2020.

Pennington, C., Freeborough, K., Dashwood, C., Dijkstra, T., and Lawrie, K.: The National Landslide Database of Great Britain: Acquisition, communication and the role of social media, Geomorphology, 249, 44–51, https://doi.org/10.1016/j.geomorph.2015.03.013, 2015.

Peruccacci, S., Gariano, S., Melillo, M., Solimano, M., Guzzetti, F., and Brunetti, M.: The ITAlian rainfall-induced LandslIdes CAtalogue, an extensive and accurate spatio-temporal catalogue of rainfall-induced landslides in Italy, Earth System Science Data, 15, 2863–2877, https://doi.org/10.5194/essd-15-2863-2023, 2023.

Petley, D.: Global patterns of loss of life from landslides, Geology, 40, 927–930, https://doi.org/10.1130/G33217.1, 2012.

Poggio, L., de Sousa, L., Batjes, N., Heuvelink, G., Kempen, B., Ribeiro, E., Rossiter, D., and Gardi, C.: SoilGrids 2.0, https://doi.org/10.5281/zenodo.5099749, 2021.

Pradhan, B., Lee, S., and Buchroithner, M.: A GIS-based back-propagation neural network model and its cross-application and validation for landslide susceptibility analyses, Computers, Environment and Urban Systems, 34, 216–235, https://doi.org/10.1016/j.compenvurbsys.2009.12.004, 2010.

Reichenbach, P., Rossi, M., Malamud, B., Mihir, M., and Guzzetti, F.: A review of statistically-based landslide susceptibility models, Earth-Science Reviews, 180, 60–91, https://doi.org/10.1016/j.earscirev.2018.03.001, 2018.

Schmitt, R., Tanyas, H., Jessee, M., Zhu, J., Biegel, K., Allstadt, K., Jibson, R., Thompson, E., van Westen, C., Sato, H., Wald, D., Godt, J., Gorum, T., Xu, C., Rathje, E., and Knudsen, K.: An open repository of earthquake-triggered ground-failure inventories, https://doi.org/10.5066/F7H70DB4, 2017.

Shang, H., Su, L., Chen, W., Tsangaratos, P., Ilia, I., Liu, S., Cui, S., and Duan, Z.: Spatial Prediction of Landslide Susceptibility Using Logistic Regression (LR), Functional Trees (FTs), and Random Subspace Functional Trees (RSFTs) for Pengyang County, China, Remote Sensing, 15, https://doi.org/10.3390/rs15204952, 2023.

Sim, K., Lee, M., Remenyte-Prescott, R., and Wong, S.: An Overview of Causes of Landslides and Their Impact on Transport Networks, in: Advances in Modelling to Improve Network Resilience, pp. 114–124, 2022.

Stanley, T. and Kirschbaum, D.: A heuristic approach to global landslide susceptibility mapping, Natural Hazards, 87, 145–164, https://doi.org/10.1007/s11069-017-2757-y, 2017.

Steger, S., Moreno, M., Crespi, A., Zellner, P., Gariano, S., Brunetti, M., Melillo, M., Peruccacci, S., Marra, F., Kohrs, R., Goetz, J., Mair, V., and Pittore, M.: Deciphering seasonal effects of triggering and preparatory precipitation for improved shallow landslide prediction using generalized additive mixed models, Natural Hazards and Earth System Sciences, 23, 1483–1506, https://doi.org/10.5194/nhess-23-1483-2023, 2023.

Survey, U. G.: Utah Landslide Inventory Polygons, https://gis.utah.gov/products/sgid/geoscience/landslide-inventory-polygons/, 2018.

Survey, V. G.: Vermont Landslide Inventory, https://anrgeodata.vermont.gov/datasets/landslides/explore?location=43.896903%2C-72.678150%2C8.82, 2021.

Tehrani, F., Calvello, M., Liu, Z., Zhang, L., and Lacasse, S.: Machine learning and landslide studies: recent advances and applications, Natural Hazards, 114, 1197–1245, https://doi.org/10.1007/s11069-022-05423-7, 2022.

Trigila, A., Iadanza, C., Lastoria, B., Bussettini, M., and Barbano, A.: Dissesto idrogeologico in Italia: pericolosità e indicatori di rischio, vol. 356/2021, ISPRA, rapporto 2021 edn., ISBN 978-88-448-1085-6, 2021.

UNDRR: International Cooperation in Disaster Risk Reduction: Target F, https://reliefweb.int/report/world/international-cooperation-disaster-risk-reduction-target-f, 2021.

University, P.: Polygon inventory of 12,920 Asia Summer Monsoon (ASM) Triggered landslides in Nepal (NERC Grant NE/L002582/1), https://www.data.gov.uk/dataset/d614bc9b-2696-4bd6-be01-b461cee575d1/polygon-inventory-of-12920-asia-summer-monsoon-asm/-triggered-landslides-in-nepal-nerc-grant-ne-, 2024.

Wang, H., Zhang, L., Yin, K., Luo, H., and Li, J.: Landslide identification using machine learning, Geoscience Frontiers, 12, https://doi.org/10.1016/j.gsf.2020.02.012, 2020.

Wen, M., Qiu, Q., Zheng, S., Ma, K., Zheng, S., Xie, Z., and Tao, L.: Construction and application of a multilevel geohazard domain ontology: A case study of landslide geohazards, Applied Computing and Geosciences, 20, 100 134, https://doi.org/10.1016/j.acags.2023.100134, 2023.

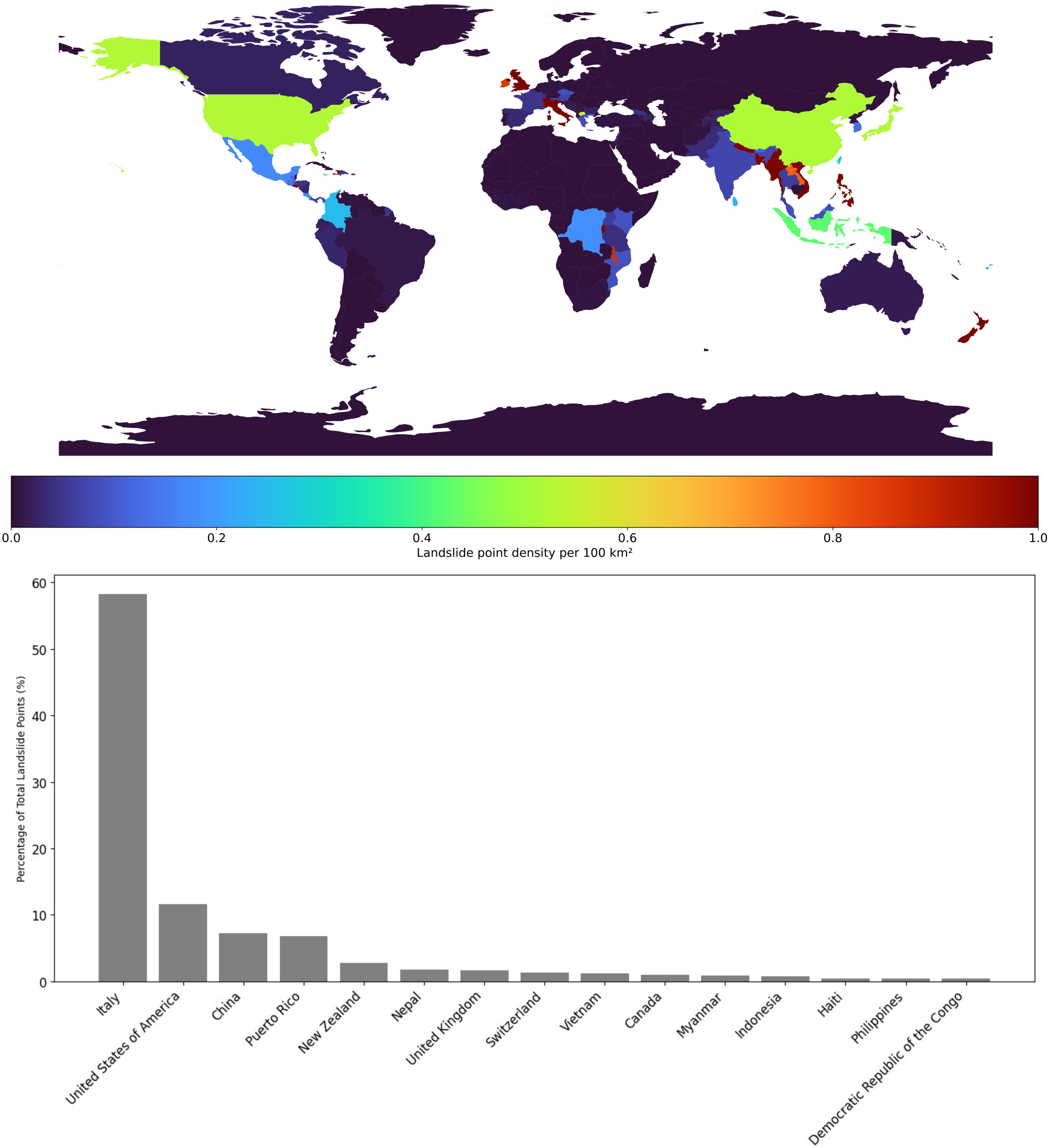

Figure 3. UGLC Landslide Point Density Per Country

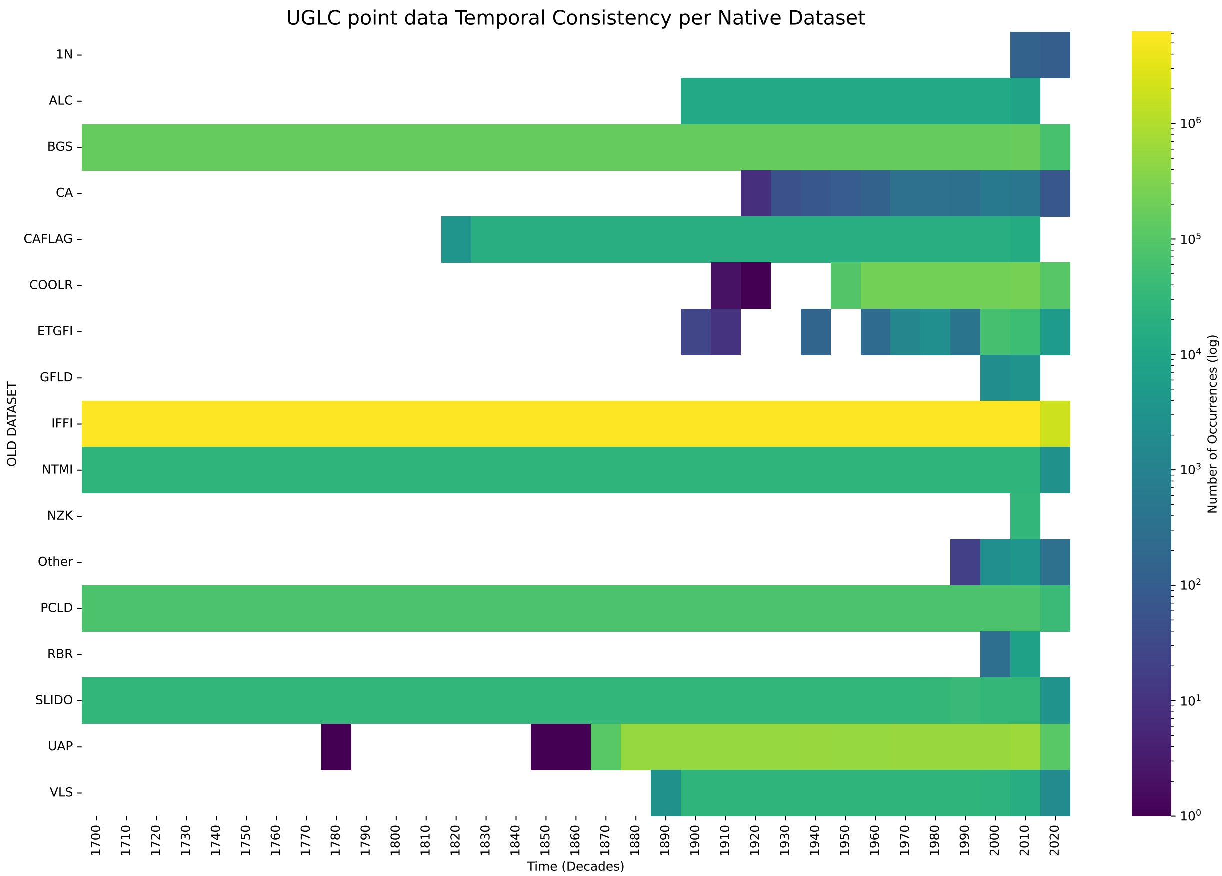

Figure 4. UGLC point data Temporal Consistency per Native Dataset

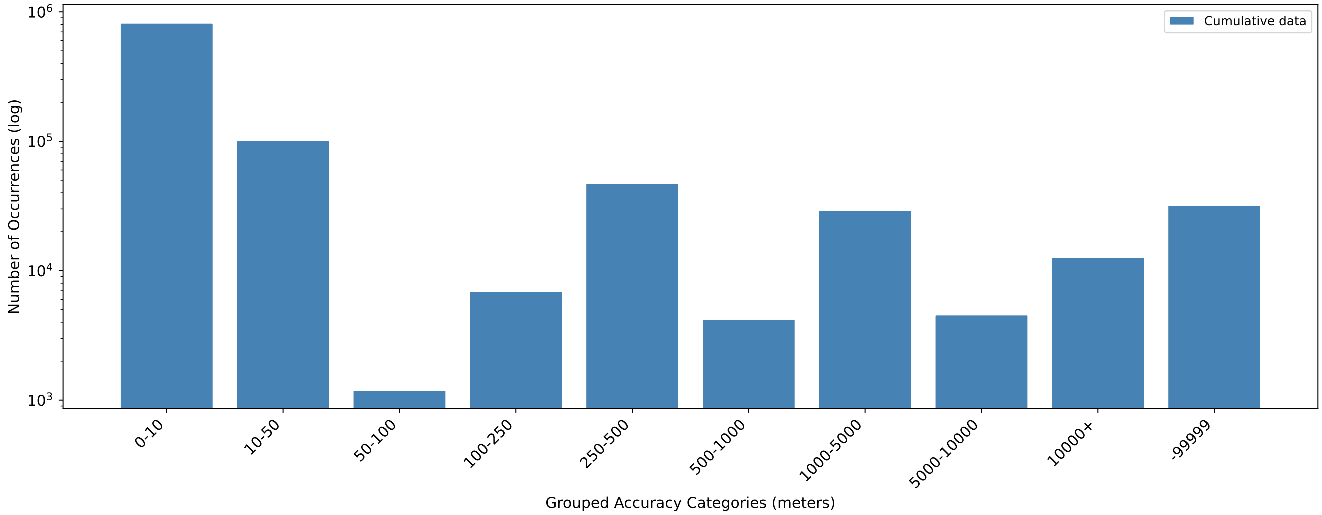

Figure 5. UGLC Point catalogue Data Accuracy Distribution

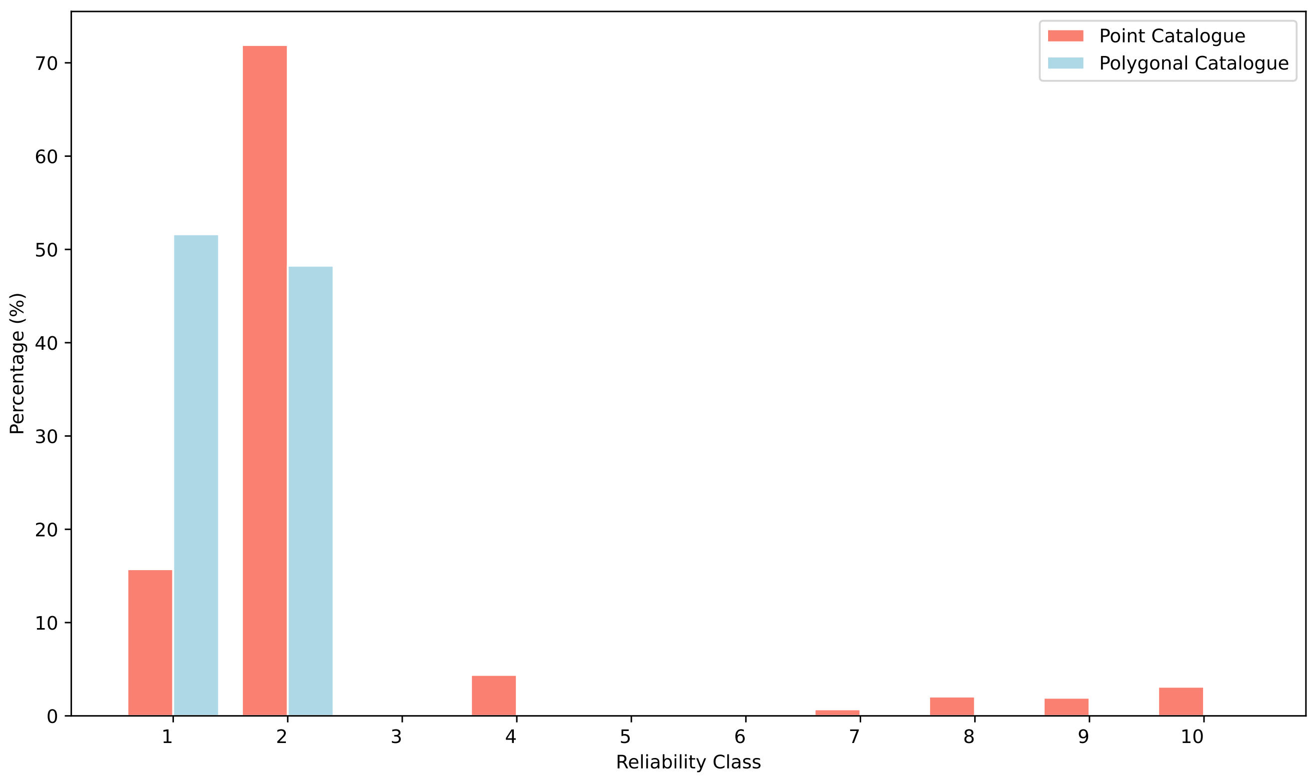

Figure 6. UGLC reliablity class distribution for Point and Polygon Catalogues

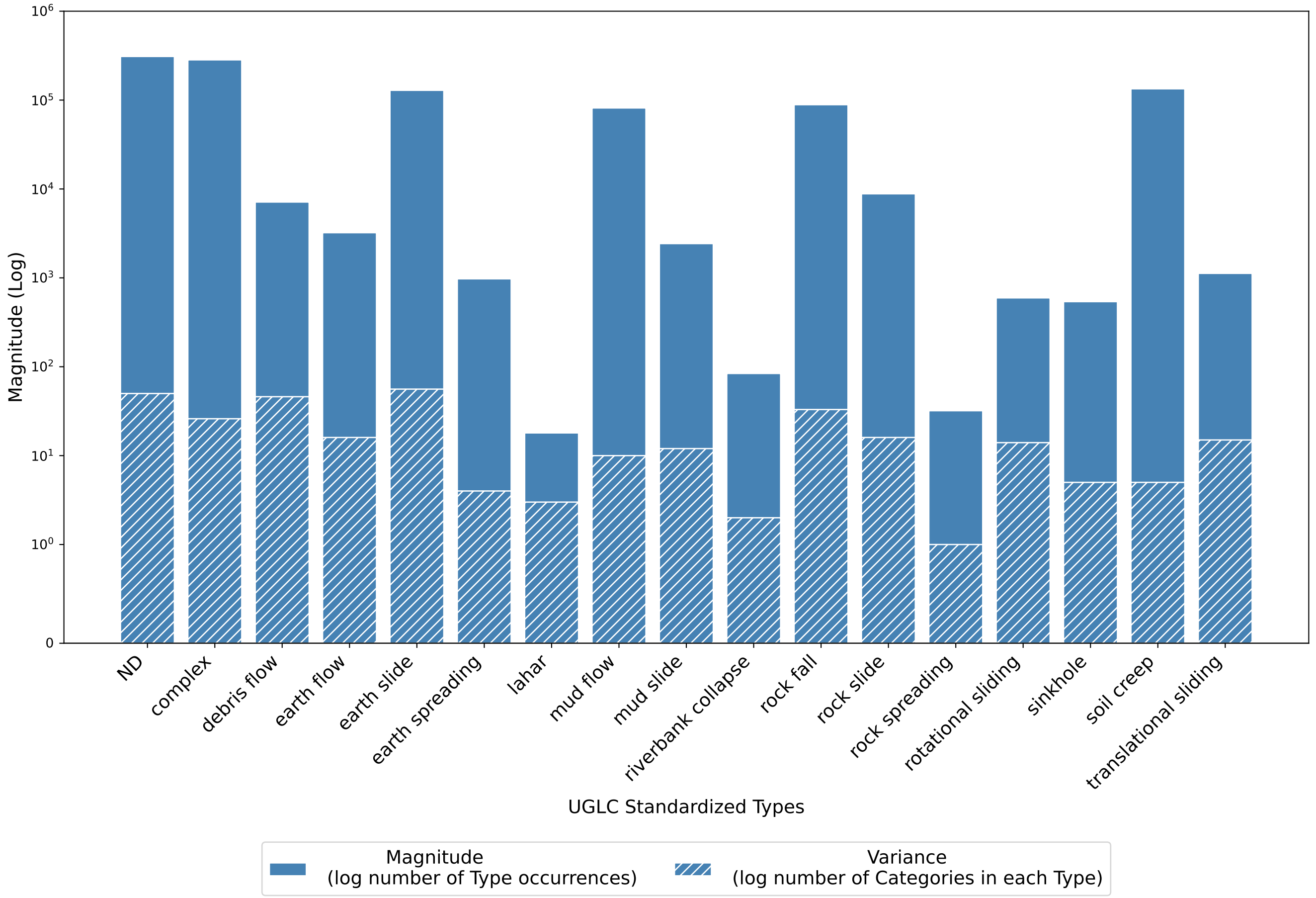

Figure 7. UGLC Standardised Type Distribution Magnitude vs Variance

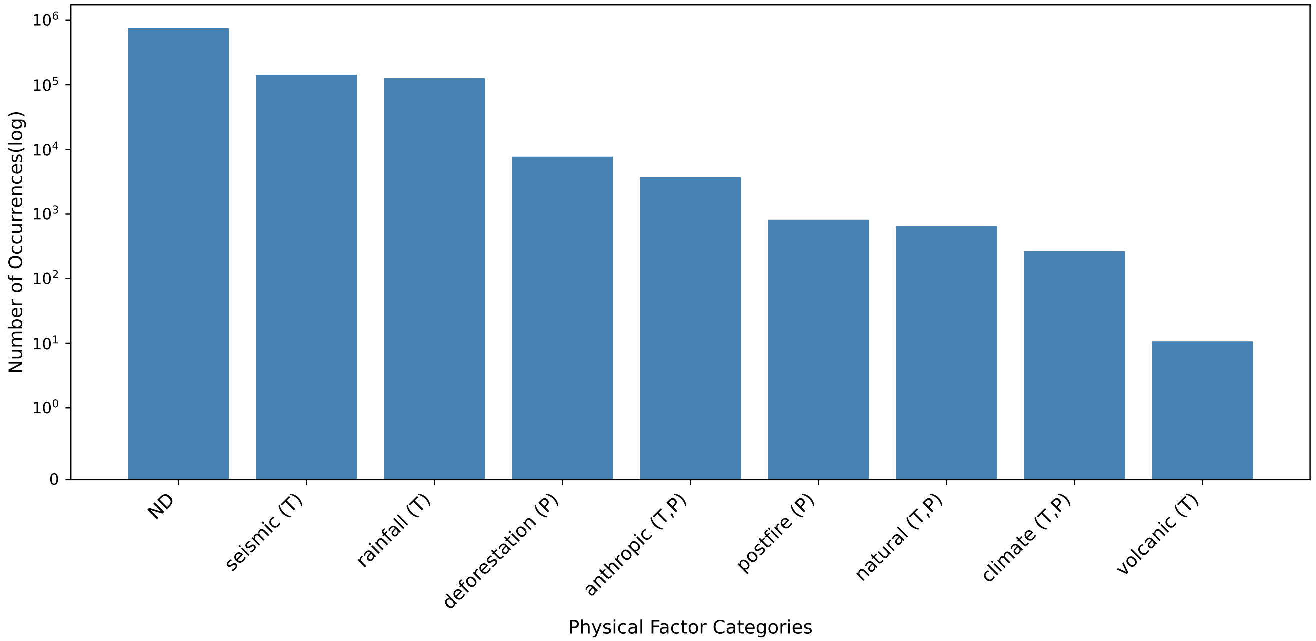

Figure 8. UGLC point data Physical Factors Distribution, distinguished in triggering factors (T) and preparatory factors (P).

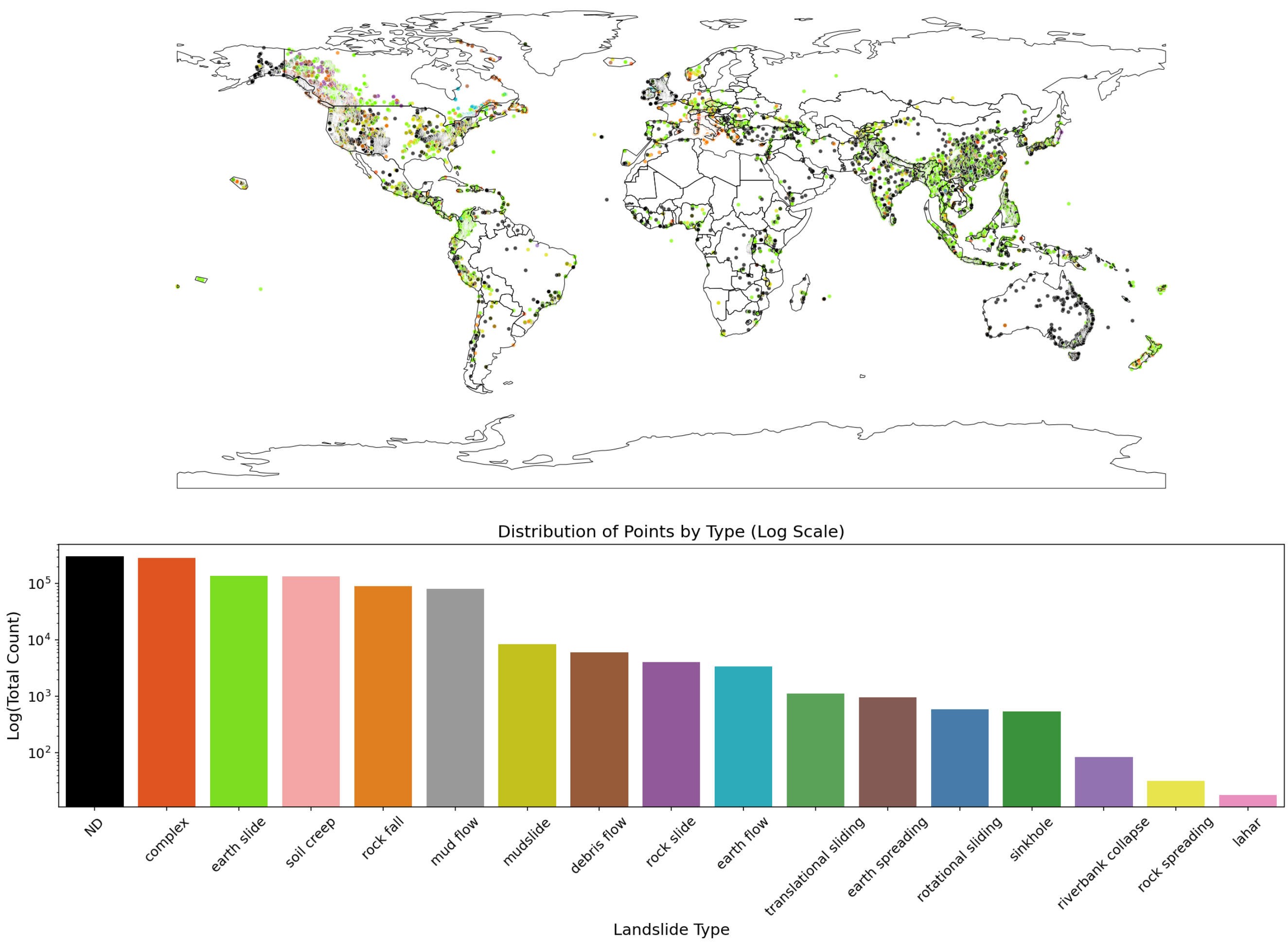

Figure 9. UGLC Landslide Points Type Distribution

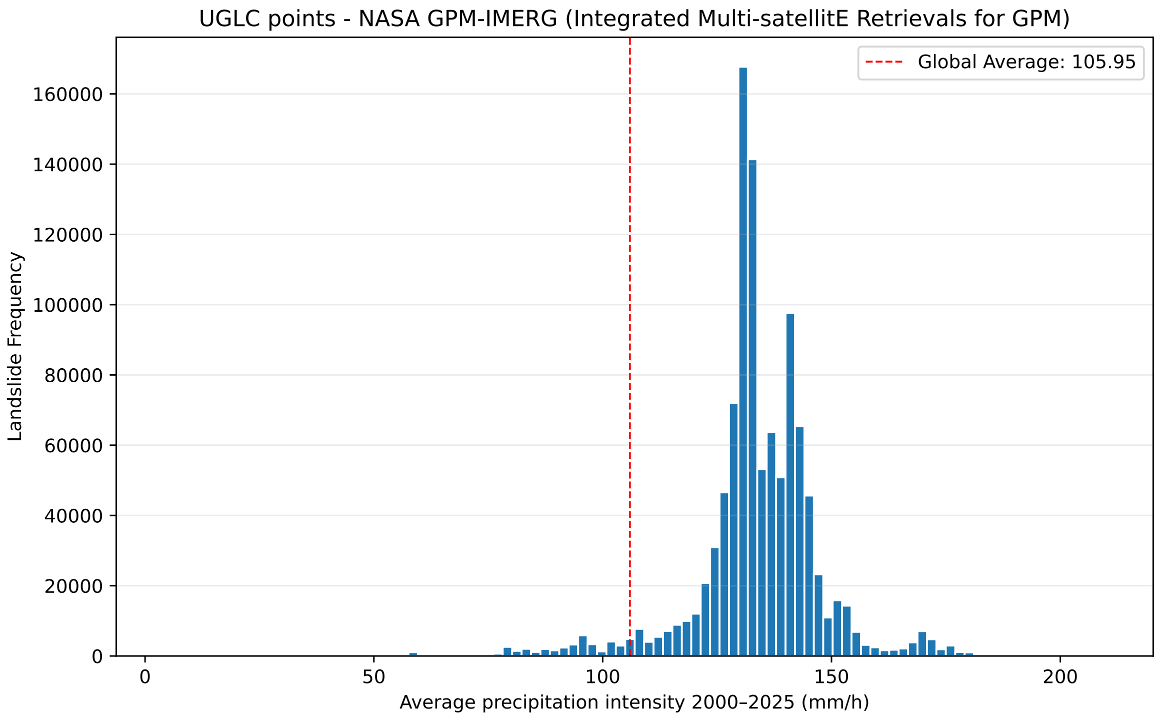

Figure 10. UGLC Landslide Points compared with mean global rainfall intensity derived from NASA GPM-IMERG data from 2000 to 2025).

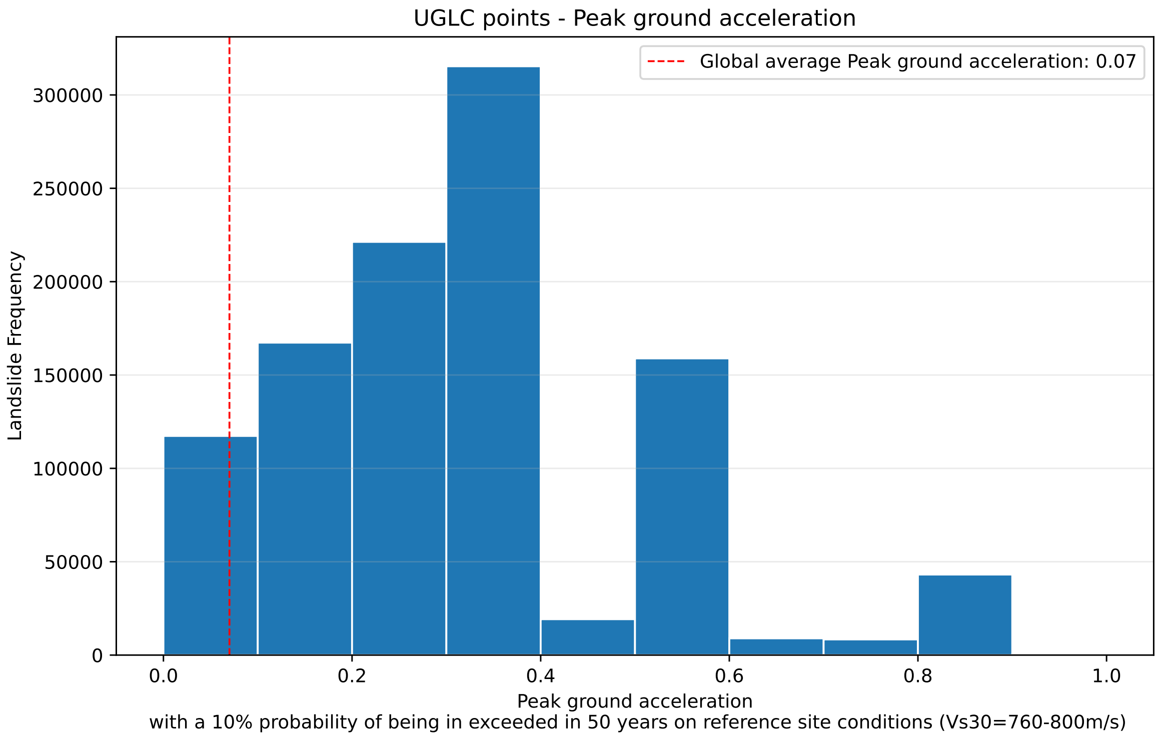

Figure 11. UGLC Landslide Points compared with global seismic hazard expressed as peak ground acceleration (PGA) from the Global Earthquake Model (GEM), corresponding to a 10% probability of exceed in 50 years under reference site conditions (Vs30 = 760–800 m/s).

Appendix A

Point datasets involved in UGLC (Table 1).

(A1) Cooperative Open Online Landslide Repository [COOLR]

COOLR is a NASA-maintained global database that collects landslide events.

Integrates the Global Landslide catalogue (Kirschbaum et al., 2010) with inventories from collaborating agencies and direct citizen reports, offering both report-based data (Landslide Reporter Catalog) (Juang et al., 2019) and automatically or manually detected events ("SALaD" Semi-Automatic Landslide Detection) (Kirschbaum et al., 2010, 2015)

Dataset Site.

(A2) Global Fatal Landslide Occurrence from 2004 to 2016 [GFLD]

This dataset presents a spatio-temporal analysis of fatal non-seismic landslides worldwide, covering the period from January 2004 to December 2016. The spatial distribution of landslides is heterogeneous, with Asia representing the dominant geographical area. The study also considers trends in human activity-triggered landslides, contributing to the discussion of climate versus human disturbances as drivers of landslide incidence (Froude and Petley, 2018).

Dataset Site.

(A3) ITAlian rainfall-induced LandslIdes CAtalogue [ITALICA]

ITALICA catalogue contains information on rainfall-induced landslides that occurred in Italy between January 1996 and December 2021. This dataset provides highly accurate spatial and temporal details of landslides, making it particularly useful for defining rainfall conditions that may trigger future hydrogeological instability in Italy (Peruccacci et al., 2023).

Dataset Site.

(A4) Landslide Inventories across the United States version2 [UAP]

This database provides an openly accessible centralised map of all existing information on landslide occurrence throughout the US. The database allowed for the first objective evaluation of previous national-scale landslide susceptibility products. The compilation can ultimately inform other research and more general hazard assessments for disaster management plans, transportation routes, and potentially insurance or other private industries (Mirus et al., 2020).

Dataset Site.

(A5) Australia Landslide catalogue [ALC]

This dataset provides a comprehensive record of landslide and flood events documented by Geoscience Australia and contributing scientific organisations. It includes naturally occurring landslides, those with significant human influence or direct human triggers, as well as flood events that caused substantial erosion or involved mud and debris flows. The dataset offers detailed geotechnical, geological and morphological information on individual landslides, compiled from relevant studies and

publications by various authors spanning the period from 1900 to 2016 (Australia, 2016).
Dataset Site.

(A6) Preliminary Canadian Landslide Database [PCLD] This database includes 9064 features with assigned landslide and material (surficial vs. rock) types. Where known, the date of occurrence, trigger, contributing factor, and reference are provided. The landslides have been identified primarily using Google Earth and publicly available lidar. Previously published landslide databases have also been incorporated and referenced. The Landslide Type attribution should be considered preliminary. (Brideau et al., 2024).
Dataset Site.

(A7) Shallow Landslide Inventory for 2000-2019 [RBR] This Landslide Inventory consists of 7944 landslide points detected with Google Earth image interpretation. Only recent landslides were considered, obtained by comparing images of different dates. The analysis is limited to shallow landslides, excluding deep rocky landslides with a maximum depth of a few metres. The depth of landslides was estimated through field observations and visual analysis of high-spatial-resolution satellite images (between 30 and 60 cm), for the period between 2000 and 2019. The inventory was compiled from Google Earth images as an update to the dataset. (Depicker et al., 2020, 2021).
Dataset Site.

(A8) Map of co-seismic Landslides for the 7.8 Kaikoura earthquake [NZK]
The 14 November 2016 Kaikōura earthquake generated thousands of landslides, which were initially mapped using low-resolution satellite imagery and later through an aerial survey. Version 2.0, completed in 2019, includes 29557 landslides, and a future version 3.0 is planned to include even more details and attributes (Massey et al., 2021).
Dataset Site.

(A9) Mass Movements Information System (SIMMA) of the Colombian Geological Service [CA]
This database consists of 2506 points, updated until 2023. It was created through the photo interpretation and digitisation of background images. The Colombian Mass Movement Information System (SIMMA) database was used to support digitisation in a GIS environment. The created database includes not only spatial and temporal location information, but also known geotechnical attributes of landslides, such as type, activity (supported by comparison with multispectral imagery), frequency, total extent, activity and average area of each type of landslide (Herrera-Coy et al., 2023).
Dataset Site.

(A10) National Landslide Database - Index data [BGS]
The British Geological Survey (BGS) has developed the National Landslide Database to investigate geological hazards associated with landslides in the UK. It is the country's most comprehensive source of landslide data, with more than 17000 recorded

events. Data were collected using various methods, including automated search in other databases and media, new mapping technologies, and citizen science engagement through social media. The database will also be used to produce a national UK landslide susceptibility map to model the impact of rainfall on soil stability. The information collected contributes to the reduction of the risk of natural disasters (Pennington et al., 2015).
Dataset Site.

(A11) Landslide Events Data [NTMI]

This dataset started as part of the initial work of the Irish Landslides Working Group, a comprehensive database of landslide events in Ireland was established. The earliest records of landslide events date back to 1488. This dataset has been fundamental to the landslide mapping work, particularly the Landslide Susceptibility Map. The database contains 2811 points (Ireland, 2012).
Dataset Site.

(A12) Vermont Geological Survey's preliminary landslide inventory [VLS]

A collection of landslide locations from the Vermont Geological Survey's preliminary landslide inventory, public geoform verified landslides, and other technical reports. The inventory includes documented historical sites and landslides verified by field visits or remote sensing. The points do not indicate the extent of the landslide (Survey, 2021).
Dataset Site.

(A13) Statewide Landslide Information Database for Oregon [SLIDO] The dataset is a collection of centroids of published or known historical landslides (circa 1928-2018 in Oregon). All available information on these landslides, including damage and loss, is included in the attribute table. This dataset is a collection of regionally significant or typical landslide locations with detailed, publicly available, site-specific studies (DOGAMI, 2024).
Dataset Site.

(A14) French Landslide Observatory (2015-2027) [1N]: This dataset contains data dedicated to the analysis of endogenous landslide seismicity on several independent sites studied by the French Landslide Observatory - OMIV (e.g. seismic sources related to physical processes triggered by the deformation of unstable slopes) and to the production of catalogues of landslide seismic sources. The sites studied are either continuously active large landslides or rocky cliffs affected by recurrent rockfalls. The sites are located in mountainous or coastal environments (Malet et al., 2015).
Dataset Site.

(A15) The CAmpi Flegrei LAndslide Geodatabase [CAFLAG] The CAmpi Flegrei LAndslide Geodatabase (CAFLAG) consists of 2302 landslide events that occurred along the volcanic slopes of the Campi Flegrei caldera (Italy), mostly in the past century. Landslide information was collected from multiple sources, including research articles, national catalogues, geological

maps, and fieldworks (Esposito and Matano, 2023).
Dataset Site.

(A16) Earthquake-Triggered Ground-Failure Inventories [ETFGI] The Earthquake-Triggered Ground-Failure Inventory Repository is designed to provide open access to data on earthquake-triggered ground failures, such as landslides and liquefaction. It contains inventories for a variety of soils, climates, and seismic conditions provided by USGS and external authors. The archive presents the original digital files of the inventory, where available, and an integrated database of standardised attributes. The goal is to support the development of robust and transportable soil failure models for seismic hazard analysis (Schmitt et al., 2017).
Dataset Site.

(A17) Inventory of landslide phenomena in Italy [IFFI] The Inventory of Landslide Phenomena in Italy—IFFI, produced by ISPRA and the Autonomous regions and provinces — counts landslides according to standardised and common methods. Due to the scale of the cartography used (1: 10.000) and the number of parameters associated with it, the IFFI Inventory is the most complete and detailed landslide database in Italy. The working methodology is based on the collection of historical and archival data, interpretation of aerial or satellite photographs, ground surveys, a landslide data sheet structured in three levels of progressive depth for the archiving of information, and a cartographic representation that includes: a landslide phenomenon identification point (PIFF) placed at the apex of the landslide, a polygon when the landslide is mapped at the adopted survey scale, and/or a line in the case of very elongated phenomena (e.g. rapid flows) (Trigila et al., 2021).
Dataset Site.

Appendix B

Polygonal datasets involved in UGLC (Table 2).

(B1) Patagonian Andes Landslides Inventory [PALI] The Patagonian Andes Landslides Inventory (PALI) is a comprehensive initiative to develop an automated landslide detection model adapted to the complex and dynamic landscapes of the Patagonian Andes. This effort addresses the significant gap in the availability of landslide inventories for this remote and geologically active region, where extreme climatic conditions and rugged topography challenge traditional mapping techniques.
The project will use state-of-the-art Deep Learning (DL) techniques to process and analyse satellite imagery to identify and classify landslides with high precision. Using advanced remote sensing data and machine learning algorithms, PALI aims to establish a robust methodological framework for landslide detection that can be applied to other mountain regions facing similar data scarcity issues.
In addition, this research evaluates the feasibility and efficiency of DL-based approaches for mapping landslides in the Andes and evaluates their performance compared to traditional inventory methods. The Andean datasets generated will provide

valuable information on landslide occurrence, distribution, and triggers, contributing to improved risk assessment and risk mitigation strategies (Morales et al., 2022).
Dataset Site.

(B2) Inventory of landslides triggered by the 2015 Mw 6.0 Sabah earthquake [MAL] An inventory of 5198 slope movements has been mapped over an area of 810 km^2. The inventory includes landslides triggered by the 6.0 magnitude earthquake that occurred in Sabah on 4 June 2015. The spatial analysis of the landslides was performed using landslide density (LND) and landslide area percentage (LAP) by dividing the study area into regular 1 km^2 cells. The ESI-07 seismic intensity for each cell was estimated by applying published area-volume relationships, demonstrating that the epistemic uncertainty associated with the chosen equation has little impact on the final result. Pre-existing landslides were identified using multi-temporal Google Earth imagery acquired between May 2008 and April 2015, and mapped as polygons that encompass source and deposit areas (Ferrario, 2022).
Dataset Site.

(B3) Inventories of landslides triggered by the 2019 Cotabato - Davao del Sur (Philippines) seismic sequence [PH] Between October and December 2019, the provinces of Cotabato and Davao del Sur (Philippines) were struck by four earthquakes with a magnitude > 6.0. The sequence started with a Mw 6.4 event on 16 October (EQ1), followed by three earthquakes of Mw 6.6, 6.5 and 6.8 on 29 October (EQ2), 31 October (EQ3) and 15 December 2019 (EQ4). The landslides in the dataset were manually mapped onto PlanetScope images at a resolution of 3 metres (Ferrario et al., 2023).
Dataset Site.

(B4) Haiti Landslide Dataset [HLD] Evaluation of post-earthquake landslide potential regarding the geological hazard of the Tiburon Peninsula. This study identified 4.893 landslides as a result of the earthquake and subsequent heavy rain from Tropical Cyclone Grace. Satellite imagery was used to map the earthquake-related landslide (Martinez et al., 2021).
Dataset Site.

(B5) Polygon inventory of 12.920 Asia Summer Monsoon (ASM) Triggered landslides in Nepal [ASM] This is a polygon .shp file of 12.920 Asia Summer Monsoon (ASM)-triggered landslides across central-eastern Nepal from 1988 - 2018. This inventory includes the landslide locations, perimeters (Length field), areas and pre-/post-monsoon season satellite image dates used to map each landslide. Note that landslides were not mapped in the years 2011 and 2012 due to scan line errors in Landsat 7 imagery. These landslides were mapped for a variety of purposes, from performing landslide susceptibility assessments to investigating landslide processes and preconditioning (University, 2024).
Dataset Site.

(B6) Danish landslide inventory [GEUS] This is the first comprehensive national landslide inventory for Denmark, derived from a 40 cm resolution DEM from 2015, supported by several 12.5 cm resolution orthophotos. The inventory was carried out using a manual mapping approach based on expert experience, with a quality control mechanism to assess its completeness. In total, 3.202 landslide polygons were mapped in Denmark, achieving a completeness of 87% (Luetzenburg et al., 2022).
Dataset Site.

(B7) Utah Landslide Inventory Polygons [UTH] Landslide Inventory Polygons depict landslide activity and debris flows at scales of 1:24.000 or better and were captured using Lidar, stereo aerial photography, other data, and field reconnaissance by the UGS. Landslide Inventory Polygons represents the most recent landslide mapping efforts of the UGS. Data represents a compilation of existing mapping of landslides at 1:100.000 pre-2007 and new landslide-specific data depicting a more detailed inventory at 1:24.000 post-2008 (Survey, 2018).

The Utah Landslide Inventory Polygons is a dataset from the Utah Geological Survey (UGS) that maps landslide activity throughout the state. It captures landslides and debris flows using LiDAR, stereo aerial photography, field reconnaissance, and other data sources. The inventory provides detailed mapping at scales of 1:24,000 or better, offering high-resolution insights.

The pre-2007 data were mapped at 1:100,000, providing a broader overview, while the post-2008 data features more precise mapping at 1:24,000. The most recent update occurred in 2018, ensuring reliable information. The dataset aids in land use planning, risk assessment, and hazard mitigation.

The data is accessible via the Utah Geospatial Resource Center (UGRC) through the Geoscience Data and Open Data Portal. Additional resources on landslide hazards are available on the UGS Landslide Hazards page.
Dataset Site.

(B8) Japan landslide dataset for semantic segmentation [JLD] This database contains images used for semantic segmentation of landslide scars from a fully convolutional neural network U-Net. (Bragagnolo et al., 2020).
Dataset Site.

(B9) Cooperative Open Online Landslide Repository Polygons [COOLR] This dataset is a global dataset of landslide events provided by NASA. The Cooperative Open Online Landslide Repository (COOLR) contains data collected through citizen science contributions via Landslide Reporter, stored as the Landslide Reporter Catalogue, along with NASA's Global Landslide Catalogue, compiled since 2007. It represents a polygonal dataset of landslide occurrences worldwide (Kirschbaum et al., 2010, 2015).
Dataset Site.

(B10) Earthquake-Triggered Ground-Failure Inventories Polygons [ETFGI] The Earthquake-Triggered Ground-Failure Inventory Repository is a polygonal dataset that provides open access to data on earthquake-triggered ground failures, such as landslides and liquefaction. It contains inventories of various soil types, climates, and seismic conditions contributed by both

USGS and external authors. The repository includes the original digital files of these inventories, where available, along with a standardised attribute database. Its goal is to support the creation of robust and transferable soil failure models for seismic hazard assessment (Schmitt et al., 2017).
Dataset Site.

(B11) Inventory of landslide phenomena in Italy [IFFI] The Inventory of Landslide Phenomena in Italy (IFFI, produced by ISPRA and the Autonomous Regions and Provinces) counts landslides according to standardised and common methods. Due to the scale of the cartography used (1:10.000) and the number of parameters associated with it, the IFFI Inventory is the most complete and detailed landslide database in Italy. The working methodology is based on the collection of historical and archival data, on the interpretation of aerial or satellite photographs, on ground surveys, on a landslide data sheet structured in three levels of progressive depth for the archiving of information, and on a cartographic representation that includes: a landslide phenomenon identification point (PIFF) placed at the apex of the landslide, a polygon when the landslide is mapped at the adopted survey scale, and/or a line in the case of very elongated phenomena (e.g., rapid flows) (Trigila et al., 2021).
Dataset Site.