



A global-scale applicable framework of landslide dam formation susceptibility

Abstract The formation and failure of landslide dams is an important and understudied, multi-hazard topic. A framework of landslide dam formation susceptibility evaluation was designed for large-scale studies to avoid the traditional dependence on landslide volume calculations based on empirical relationships, which requires comprehensive local inventories of landslides and landslide dams. The framework combines logistic regression landslide susceptibility models and global fluvial datasets and was tested in Italy and Japan based on landslide and landslide dam inventories collected globally. The final landslide dam formation susceptibility index identifies which river reach is most prone to landslide dam formation, based on the river width and the landslide susceptibility in the adjacent delineated slope drainage areas. The logistic regression models showed good performances with area under the receiver operating characteristics curve values of 0.89 in Italy and 0.74 in Japan. The index effectively identifies the probability of landslide dam formation for specific river reaches, as demonstrated by the higher index values for river reaches with past landslide dam records. The framework is designed to be applied globally or for other large-scale study regions, especially for less studied data-scarce regions. It also provides a preliminary evaluation result for smaller catchments and has the potential to be applied at a more detailed scale with local datasets.

Keywords Landslides · Landslide dams · Global scale · Susceptibility · Fluvial datasets · River hazards

Introduction

Landslide dams (LDams) represent a river blockage hazard caused by landslides obstructing river channels; LDams are frequently reported in international literature, particularly in mountainous regions with narrow river channels and steep hillsides in upstream catchment areas (Costa and Schuster 1988; Scarascia Mugnozza et al. 2006; Della Seta et al. 2017; Fan et al. 2020; Wu et al. 2022). Most LDams can form and collapse within a short period, ranging from a few hours to a month, potentially causing significant flooding consequences. These floods result from upstream backwater floods above the blockage, and/or the breach or failure of the LDams, as well as the long-term effects on local geomorphological and hydrological conditions as a result of interactions between the hillslopes and fluvial systems (Costa and Schuster 1988; Korup 2002, 2004; Scarascia Mugnozza et al. 2006; Fan et al. 2012, 2020; Della Seta et al. 2017; Wu et al. 2022). To predict and manage the LDam-related risks, it is therefore imperative to predict which specific river reaches are most likely to experience these events.

Evaluating the statistical probability of LDam formation on a local to regional scale has been the focus of many studies, which can be summarised into three major approaches: 1) damming probability classification based on expert opinion; 2) river blockage index establishment based on the empirical relationships of the landslide (deposit) volume and other related parameters; 3) the combination of probabilities related to LDam formation processes (source–pathway–receptor), including slope failure processes, landslide mass movement, and mass deposition in the channel.

The first approach, damming probability classification, identifies the final probability of LDam formation as qualitative measurements, and it relies on the expert experience. For example, van Westen et al. (2020) established several standards for classifying landslide damming potential in Dominica, including the evidence of LDam records, potential LDam occurrence probability, and other relevant geomorphological criteria of LDam formation.

In the second approach, the river blockage indexes is a measurement of LDam formation probability or stability directly estimated from related parameters, such as landslide volume, valley width, landslide velocity, catchment area, and dam height (see Table 1). Examples of these indexes are summarised in Table 1. Argentin et al. (2021) compared six different river obstruction and LDam stability indices for analysing the LDam records in the Austrian Alps and found their results are not consistent when compared to previous research, which is similar to the results in Eastern European Alps (Dufresne et al. 2018). Fan et al. (2020) compared various dam formation and stability indexes and found none of them is satisfactory for their newly compiled global-scale LDam dataset. Among all the parameters used to establish such an index, landslide (or LDam) volume is the most frequently applied. From Table 1, fluvial characteristics such as drainage area and river width are also identified as important variables.

The third approach is the LDam formation probabilities approach, which attempts to generate a final probability by combining the effects of various related processes. Fan et al. (2014) calculated the dam formation landslide (DFL) susceptibility from three specific LDam formation-related processes: i) whether the slope is prone to slope failures; ii) whether the landslide mass can travel far enough to reach the fluvial channel; and iii) whether the landslide volume is sufficient to form a LDam. Another method is proposed by Tacconi Stefanelli et al. (2020), who evaluated the LDam formation probability of landslides by using damming predisposition and damming probability and combined the separate LDam formation processes into a single susceptibility index.

These efforts exploring the LDam formation probability either require comprehensive records with landslide dimension data or

Table 1 River blockage indexes for estimating the probability of LDam formation or stability

River blockage index	Geomorphological parameters used other than landslide or LDam volume	Reference
Annual constriction ratio	Landslide velocity, valley width	Swanson et al. (1986)
Dimensionless blockage index	Catchment area, dam height	Ermini and Casagli (2003)
Blockage index	Catchment area	Canuti et al. (1998), Casagli and Ermini (1999)
Morphological obstruction index	Valley width	Tacconi Stefanelli et al. (2016, 2018)
Backstow index	Landslide height (maximum crest height of the LDam)	Korup (2004)
Catchment ruggedness-based indices (CRBI)	Catchment area, mean slope of the catchment, Melton ruggedness number, the mean height of the dam, the width of the dammed valley, run-out distance	Shafeiganjeh et al. (2022)

rely heavily on empirical formulas based on local conditions and expert knowledge. It is challenging to prepare data in the absence of prior events, especially for large-scale research, for which detailed inventories of landslides/LDams are usually unavailable. Moreover, most previous research has focused predominantly on the surrounding hillslopes to predict the damming probability of potential landslides or current landslide reactivation, rather than identifying the LDam formation probability for the prone river reaches. Most current LDam and landslide records are only recorded as point locations, which makes it difficult to calculate the landslide dimension data for large-scale research that is needed to identify which river reach is most prone to having LDams form. Besides, the river blockage index derived from regional-scale studies may not apply to other datasets and its application accuracy may not be reliable. Dufresne et al. (2018) found that the previous LDam stability thresholds were applicable to some large rockslide dams in the European Alps. Cencetti et al. (2020) showed that such an index is hard to generalise as it was generated from specific local geomorphological and hydrological conditions. Struble et al. (2021) also found that the dam stability index from another study did not fit with their LDam records in Western Oregon and so as the scaling relationship between the upstream catchment area and landslide dam size.

This research addresses LDams as a global-scale multi-hazard by developing a framework of LDam formation susceptibility evaluation, based on slope drainage units that can be applied globally by combining a landslide susceptibility evaluation with global fluvial datasets to address the research gaps identified above. The framework output is the LDam formation susceptibility for the river reaches themselves, not just the slopes above. To test the framework effectiveness, the comprehensive data available from Italy (1,027,802 records, 58.7% of total available global landslide records) and Japan (359,384 records, 20.5% of total available global landslide records) are used to develop the framework given the limited numbers of robust datasets from elsewhere (for more details of landslide records distribution across all the countries and regions, a visualisation is provided in Section 1.2 in the Supplementary document). These two countries are chosen, as

the landslides are widespread in the countries and well studied. Another reason is that their national-scale landslide records are available as open access to the public. The similarity of land areas, climate, geomorphology, geology, and tectonic conditions partially contributes to the control of certain large-scale constraints (Table 2). The susceptibility evaluation results using collected landslide and LDam records have been also validated. Therefore, such a developed framework uses global-scale climate, geomorphological and fluvial datasets, and can be applied in the traditionally data-scarce areas to understand which river reaches globally are most prone to LDam formation.

Data sources

Landslide and LDam datasets

A total of more than 1 million landslide records and 779 LDam records, most of which are geolocated, were collected globally from both landslide and LDam datasets, covering different spatial scales. The details of the datasets used are provided in Table 3. For LDam records, they also included the information of the DFL and thus can be used as a subset for further landslide data training. For most landslide records collected globally, there is currently no more detailed information of landslide types, scale etc. These records were utilised to identify the appropriate input variables for a landslide susceptibility evaluation model applied in different regions, creating training datasets for establishing the model and validating the susceptibility results.

Global fluvial datasets

Global fluvial datasets have been developing rapidly in the past decade (Lehner and Grill 2013; Yamazaki et al. 2014; Allen and Pavelsky 2018; Linke et al. 2019; Yamazaki et al. 2019; Frasson et al. 2019; Feng et al. 2022). According to the previous global-scale LDam dataset research by Wu et al. (2022), global fluvial datasets provide a more reliable and consistent data source for determining river width

Table 2 Environmental parameters of Italy and Japan

	Italy	Japan
Land area (km ²)	302,073	377,973
Mean annual rainfall (mma ⁻¹)	More than 3000 (European Alps, Palladino et al. 2018), to lower than 800 (Sicily, Melillo et al. 2016)	1000–4500 (Saito et al. 2014)
Mean annual temperature (°C)	13.5	12.36
Geology and landforms	Mountain areas dominated by metamorphic and sedimentary rocks in the Alps, marly-limestones and flysch in the Apennines. Volcanic and metamorphic rocks mainly outcrop in Sardinia and Calabrian mountains.	Mountain areas dominated by recent volcanics with Neogene sedimentary rocks forming lower ground
Tectonics	Compressional tectonics are represented in some parts of the alpine regions while extensional tectonics dominate the central to southern Apennines.	Compressional tectonics

compared to estimating it through a simple empirical relationship between drainage area and river width. This was demonstrated by Frasson et al. (2019), who showed significant variations in results between catchments across different regions when compared to directly applying a single relationship.

Multiple global fluvial datasets were included in several stages in this research: MERIT Hydro, a global fluvial dataset containing the data of river width at a resolution of 90 m (Yamazaki et al. 2019), was used for the river width data for the LDam formation susceptibility evaluation model establishment; HydroBASINS, the catchment boundaries and sub-basin delineations derived from HydroSHEDS data with 12 scales/levels (Lehner and Grill 2013), was utilised to further delineate the slope drainage area; HydroATLAS, a global compendium of hydro-environmental characteristics for all sub-basins of HydroBASINS (Linke et al. 2019), was used to explore related basin characteristic variables for the landslide susceptibility evaluation model establishment.

Methods

The LDam formation susceptibility framework consists of four main stages, as shown in Fig. 1 and outlined here, but followed below by a more detailed description of each stage (with further detail in the Supplementary material Section 1.1). Firstly, all necessary data was collected, and the study areas were prefiltered to identify locations susceptible for LDam formation, specifically where local conditions are prone to LDam formation, and a previous landslide record exists within the catchment. Secondly, there are two processing steps that proceed in parallel: 1) further delineating slope drainage units as the main calculation unit by using Pfafstetter, based on the current HydroBASIN sub-catchments and the forest and buildings removed Copernicus digital elevation model (FABDEM, Hawker et al. 2022); 2) performing a landslide susceptibility evaluation based on the collected landslide records, including variable exploration and selection, logistic regression modelling, validation, and optimal threshold extraction. Thirdly, the delineated

slope drainage units are then categorised into river units and adjacent slope units based on their locations and the median landslide susceptibility index extracted from each slope drainage unit to represent the landslide occurrence probability of each delineated slope drainage unit. Fourthly and finally, the landslide susceptibility index (LSI) around a specific river reach was derived by first accumulating the landslide occurrence probability on all adjacent slope drainage units around the reach, which was further combined with area of each unit and the river width data from global fluvial datasets to calculate the spatial probability of LDam formation on each river reach.

Initial area prefiltering for LDam formation study

Prefiltering the study area for the LDam formation susceptibility evaluation beforehand is essential to minimise unnecessary computational workload in large-scale modelling. Dikshit et al. (2019) proposed a rainfall threshold of 350 mm over 30 days for landslide susceptibility in the Bhutan Himalayas. Variations in geomorphological data due to the location of landslides or LDam records were documented (Jibson and Harp 2012; Tacconi Stefanelli et al. 2018), and areas with existing landslide and LDam records were prefiltered to mitigate data bias from regional difference. Moreover, Wu et al. (2022) identified unique dimensions and geomorphological features of LDams, such as the landslide height/length ratio, and suggested a threshold for LDam formation based on the relationship between river width and landslide volume, indicating limited damming potential in wider river reaches.

LDam formation zones must meet prerequisites of (i) potential landslide occurrence and (ii) the presence of a river with blockable width in the surrounding area. Prefiltering criteria include landslide and LDam records within a 19-km radius (Wu et al. 2022), climate conditions conducive to active rivers or landslides, and exclusion of continental coastlines to correct for potential underrepresentation in global river datasets (Yamazaki et al. 2019). The schematic figure of the prefiltering method and the results are shown in Fig. 2 with more details of the

Table 3 Datasets of landslide and LDam records

Name	Time coverage	Scale (area)	Number of landslide records/events	Contributors
FraneItalia	From January 2010 to 2017	National (Italy)	5438 (single) + 1787 (areal)	Calvello and Pecoraro (2018)
Dufresne (unpublished inventory)	-	Global	179	Dufresne et al. (2021)
CAmpi Flegrei LAndslide Geodatabase (CAFLAG)	1828–2017	Local (Campi Flegrei caldera, Italy)	2302	Italian National Research Council (CNR) (Esposito and Matano 2021)
Landslide Inventories across the USA	1900–2019	National (USA)	64,433	Jones et al. (2019)
Global Landslide Catalog (GLC)	2007–2019	Global	14,532	Kirschbaum et al. (2015)
High Mountain Asia Landslide Catalog Voo1(HMS_LS)	1956–2018	Regional (Asia)	12,755	Kirschbaum (2019)
Italian Catalogue of Earthquake-Induced Ground Failures (CEDIT)	1169–2019	National (Italy)	2077	Martino et al. (2022)
Digital Archive for Landslide Distribution Maps	1981–2014	National (Japan)	359,387 mass movement polygons	National Research Institute for Earth Science and Disaster Prevention of Japan (2014, accessed in 2022)
Landslide Inventories from An Open Repository of Earthquake-Triggered Ground-Failure Inventories	Since 1900	Global (event based)	356,497 geolocated data	Schmitt et al. (2017)
Rockslides and Rock Avalanches of Central Asia	-	Regional (Middle Asia)	1016	Strom and Abdrakhmatov (2018)
IFFI project (Italian Landslide Inventory)	1116–2017	National (Italy)	620,808	Trigila et al. (2010)
Landslides in Dominica	-	National (Dominica)	10,551	van Westen and Zhang (2018)
River Augmented Global Landslide Dams (RAGLAD)	Since eighth century	Global	779	Wu et al. (2022)

prefiltering procedures illustrated in Supplementary document Section 1.4. The methods were performed based on the level-12 HydroBASINS sub-catchments products, which have an average area of 130.5 km² because the level-12 sub-catchments are the highest resolution.

Slope drainage unit delineation

Selecting a proper mapping unit is crucial to map the spatial probability of hazard susceptibility modelling (Guzzetti et al. 2005).

Reichenbach et al (2018) categorised previous research into seven types of mapping units, finding that grid units (86.4%), slope units (5.1%), and unique condition units (4.6%) were most commonly used. However, comparisons between grid units, slope units, and subbasin units are still being evaluated. Grid units are predominantly used in statistical hazard susceptibility models, but slope units, effective for regional mapping, are still being refined for consistent performance (Jacobs et al. 2020). Erener and Duzgun (2011) found that a model mapped with slope units had a better

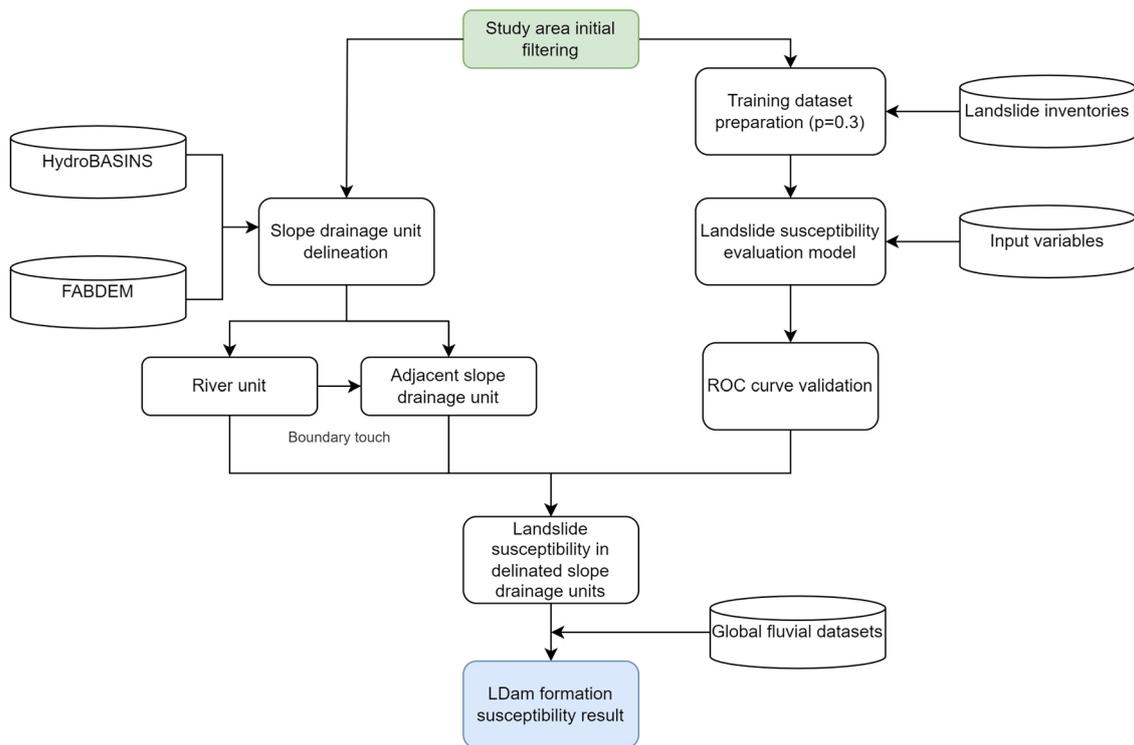


Fig. 1 Schematic figure of research methods for establishing the LDam formation susceptibility framework by combining the landslide susceptibility model and global fluvial datasets (a visualisation of this framework with more details is shown in Section 1.1 of the Supplementary document)

performance compared to those based on grid cells. Martinello et al. (2022) found that all three slope unit delineation methods they tested maintained excellent predictive performance compared to pixel-based models. Palau et al. (2020) observed that the effectiveness of gridded landslide susceptibility mapping decreases with resolution, while subbasin methods performed better at 100-m resolution. The transferability of some automatic unit delineation methods for mapping susceptibility requires further validation, as van Westen et al. (2020) found that the landslides in Dominica did not fit the automatic slope unit delineation method from Alvioli et al. (2020) because of inconsistencies. All mapping units have their pros and cons, so the selection of mapping units relies on multiple factors, such as the study's purpose, computation cost, and model performance.

For this study, the classic hydro-morphological unit is applied, named as a slope drainage unit, based on an automatic subbasin coding method, Pfafstetter (Verdin and Verdin 1999). This unit combines hillslope processes and river processes, essential for LDam formation studies, where sediment and flow accumulate within each unit at a common outlet. Although HydroBASINS product offers subcatchments delineated with a 90-m resolution digital elevation model (DEM) (Lehner and Grill 2013), higher-resolution delineations were performed using the FABDEM at 30 m (Hawker et al. 2022) via the pyflwdir package (Eilander et al. 2021). The reliability of these delineations was confirmed by comparing the FABDEM-derived units with high-resolution LiDAR-derived catchments in Dominica, showing an 81.3% overlap (Trigg et al. 2023) (see Fig. 3).

The main discrepancies occurred along the coast, which is less critical for LDam formation modelling.

Following the validation, slope drainage units were further separated into river units and adjacent slope drainage units for LDam formation susceptibility calculations. The river units are those slope drainage units that intersect with river reaches from the global fluvial datasets and the adjacent slope units are those units that sit immediately adjacent to the boundaries of the river units (see Fig. 4). The approach provides a geomorphology-based buffer zone instead of a simple distance-based buffer zone for defining areas susceptible to LDam formation.

Landslide susceptibility evaluation model

Landslide susceptibility describes the spatial probability of landslide occurrence in a given location based on a set of specific geomorphological and environmental conditions (Guzzetti et al. 2005). Because input data and calibration for these physically based models are not practical for large-scale regional studies, especially global scale, statistical-based methods are more widely applied in the large-scale landslide susceptibility evaluation (Nadim et al. 2006; Stanley and Kirschbaum 2017; Lin et al. 2017).

The logistic regression model is commonly applied in previous landslide susceptibility evaluation studies and was demonstrated to be efficient at various spatial scales from local to global scale (Ayalew and Yamagishi 2005; Rossi et al. 2010; Dou et al. 2015; Budimir et al. 2015; Lin et al. 2017; Raja et al. 2016). For large-scale

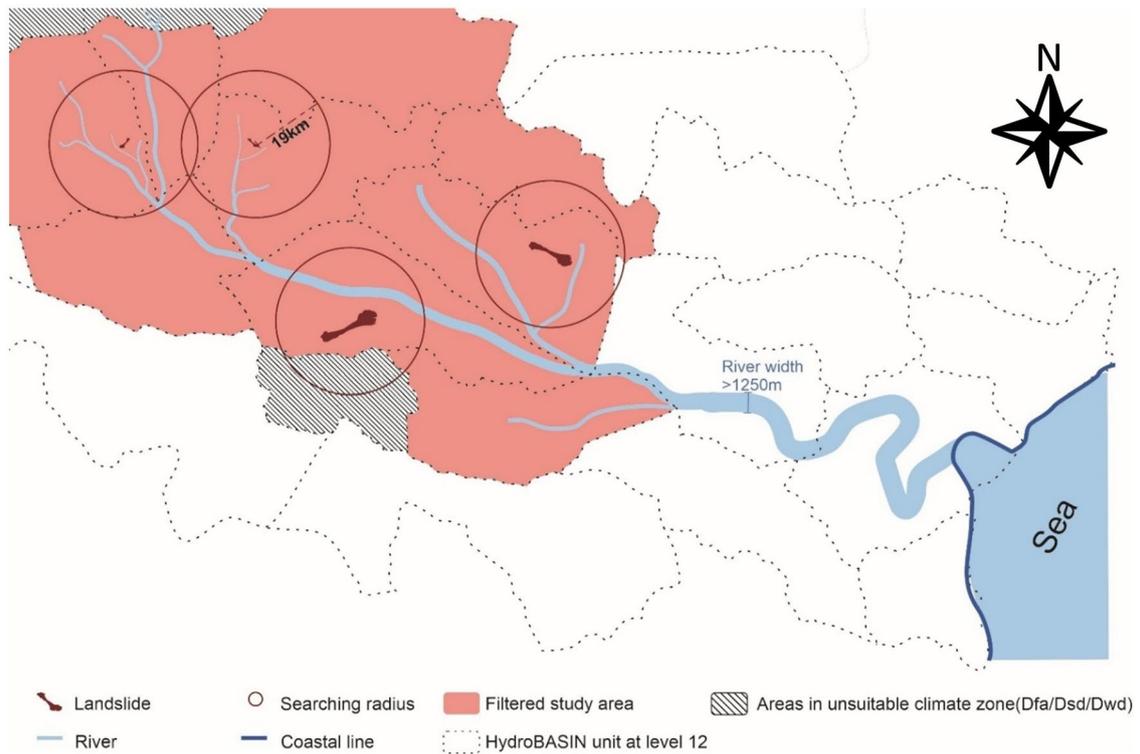


Fig. 2 Schematic figure of the prefiltering of the LDam formation study area: the areas near landslides and rivers with blockable width, excluding coastal areas and areas in unsuitable climate zones. The 19-km search radius is a likely maximum envelope for DFL travel (Wu et al. 2022)

landslide susceptibility modelling, the further delineation of landslide types and triggers is normally not considered during the modelling methods (Nadim et al. 2006; Lin et al. 2017). The logistic regression model predicts the probability of a landslide occurrence

at specific locations based on the relationship between landslide occurrences and related variables. The general formulas of the logistic regression model for landslide susceptibility evaluation, based on grid cell units, are described in Eqs. (1) and (2):

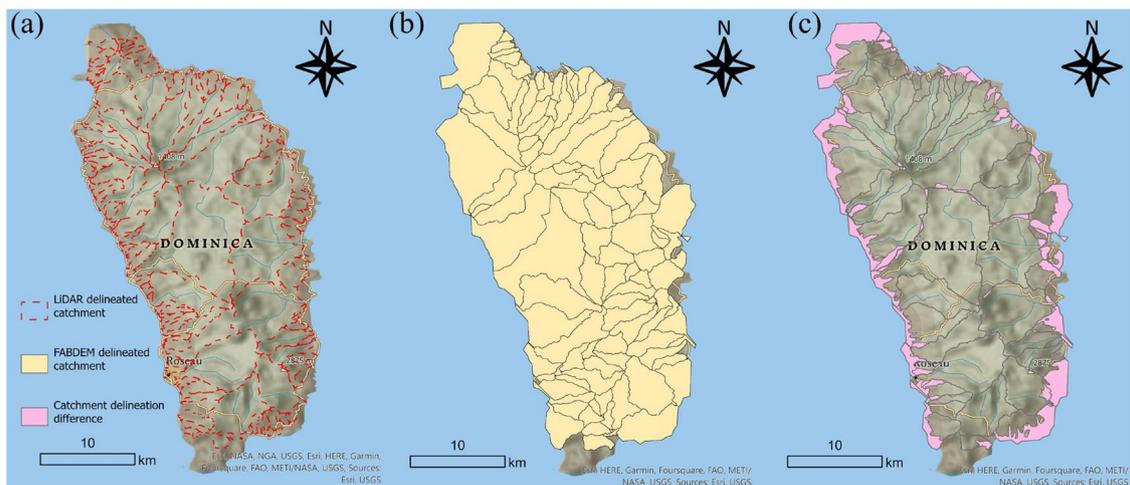


Fig. 3 Validation of FABDEM delineated slope drainage units against catchments derived from Dominica: **a** river catchment delineation result based on LIDAR data (Trigg et al. 2023); **b** slope drainage unit delineation result by Pfafstetter method based on FABDEM data; **c** boundary difference areas of the two datasets are highlighted in pink

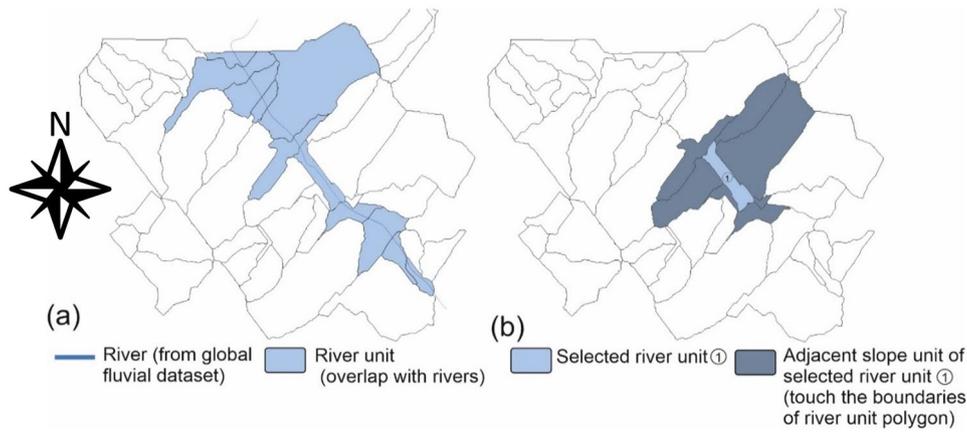


Fig. 4 Example figure showing the identification of the river unit and adjacent slope drainage units from the delineated slope drainage units: **a** river reach and intersected river units; **b** adjacent slope units to a single river unit

$$P(\text{landslide occurrence}) = \frac{1}{1 + e^{-z}} \quad (1)$$

$$z = \beta_0 + \beta_1 \times x_1 + \beta_2 \times x_2 \dots \dots \beta_n \times x_n \quad (2)$$

The output value P (landslide occurrence) in Eq. (1) of a logistic model ranges from 0 to 1, typically described as the landslide susceptibility index (LSI), which shows the spatial probability of future landslide occurrence at a given location. If the LSI value approaches 1, it means that the given location is more likely to have a landslide occurrence; if the LSI is close to 0, it means that the given location has much less probability of landslide occurrence. z is a dependent variable that reflects landslide occurrence based on all input variable x_n , β_n is the coefficient of a specific variable measuring the importance of the final evaluation result based on the past landslide records (the coefficient is the constant when $n = 0$).

According to Martinello et al. (2020), grid-cell units appear to be the more suitable measurement unit for modelling, while dimensionally appropriate slope units seem to be the most effective way to generate accurate landslide susceptibility maps. Therefore, a few studies have combined these two types of units together to achieve the final landslide susceptibility evaluation result (Domènech et al. 2019; Martinello et al. 2020). In our study, the final LSI evaluation result is modelled and assessed based on the grid cell first. The result of the grid cell approach was then aggregated to identify the LSI based on the delineated slope drainage unit by assigning the median LSI value within each slope drainage area to present the spatial probability for the whole unit.

Variable exploration for landslide susceptibility evaluation

Previous research can provide a foundation for selecting appropriate variables to evaluate landslide susceptibility by identifying potentially useful characteristics. Variables that were applied in previous landslide susceptibility studies mainly came from the research fields related to the conditions required for landslide occurrence, such as geology, hydrology, land cover, and geomorphology (Reichenbach et al. 2018). The variables were explored by accessing the correlation between multiple variables to landslide

occurrence using global-scale HydroALTAS dataset (Linke et al. 2019), details of which can be found in Supplementary material Section 1.3.

The final 11 selected variables for landslide susceptibility evaluation are river area, elevation, slope, curvature, air temperature, precipitation, landcover classes, potential natural vegetation classes, clay fraction in soil, lithological classes, and soil erosion rate. The vegetation type and land cover can be considered two different variables for landslide susceptibility evaluation according to the literature reviews (Guzzetti et al. 1999; Reichenbach et al. 2018). All the numerical variables were transformed with a quantile transformer to make them dimensionless and to improve the performance of the predictive models, as this data transformation is a robust pre-processing scheme and makes the variables measured at different scales more directly comparable (Pedregosa et al. 2011).

Data sources of landslide susceptibility evaluation variables

The data sources for the selected variables mainly kept with the same data sources as HydroALTAS or the updated version of original datasets after the HydroALTAS released (Linke et al. 2019). The data sources of DEM and land cover variables were replaced with higher-resolution data sources, including the 30-m resolution FAB-DEM (Hawker et al. 2022), 30-m resolution FROM-GLC (Finer Resolution Observation and Monitoring of Global Land Cover, Gong et al. 2013), and 1-km resolution Global Maps of Potential Natural Vegetation (Hengl et al. 2018). All these data sources are open access, and their information is provided in Table 4.

Spatial resolution for landslide susceptibility evaluation modelling process is determined by considering the detailed degree of the mapping details of landslide and LDam records and the collected data. A 30-m resolution was selected for grid-cell based landslide susceptibility evaluation modelling as this was the highest underlying resolution of most of the utilised global data sources. This resolution represents the potential probability of landslide occurrence well, as more than 50% landslides that blocked the river have an area less than 175,000 m² (Wu et al. 2022), which is approximately 20 cells at a 30-m resolution.

Table 4 Data source information of input variables for landslide susceptibility

Index	Category	Variable	Special resolution	Data source (reference)	Updated from HydroALTAS data sources
1	Hydrology	River area	500 m	HydroALTAS (Linke et al. 2019)	N
2	Geomorphology	Elevation	30 m	FABDEM (Hawker et al. 2022)	Y
3		Slope	30 m	FABDEM (Hawker et al. 2022)	Y
4		Curvature	30 m	FABDEM (Hawker et al. 2022)	Y
5	Climate	Air temperature	30 arc-second (1 km)	WorldClim 2.1 (Fick and Hijmans 2017)	Y-version updated of the same data source
6		Precipitation	30 arc-second	WorldClim 2.1 (Fick and Hijmans 2017)	Y-version updated of the same data source
7	Landcover	Land cover classes (all landcover type)	30 m	FROM-GLC (updated version in 2017) (Gong et al. 2013)	Y
8		Potential natural vegetation classes (vegetation type)	1 km	Global Maps of Potential Natural Vegetation at 1-km resolution (Hengl et al. 2018)	Y
9	Soil and geology	Clay fraction in soil	1 km	Soilgrid1km (Hengl et al. 2014)	N
10		Lithological classes	30 arc-second	GLiM—Global Lithological Map (Hartmann and Moosdorf 2012)	N
11		Soil erosion	100 m	High resolution cropland global soil erosion (GloSEM 1.3) (Borrelli et al. 2022)	Y-version updated of the same data source

Training dataset preparation

The first assumption of statistical-based methods for susceptibility evaluation is that the past is the key to the future (Guzzetti et al. 2005), so a training dataset based on landslide records needs to be prepared for modelling. We took all the landslide records from collected datasets as the landslide samples for two detailed study regions, Italy and Japan. There were more than 1 million landslide records in Italy, with 300 thousand landslide records in Japan.

Only a few researchers focused on exploring the sampling method of landslide susceptibility evaluation (Kornejady et al. 2017; Shao et al. 2020). The ratios of non-landslide/landslide sample greatly affect the occurrence probability of landslides. Shao et al. (2020) found that the landslide susceptibility model prediction is almost consistent with the actual ratio of landslide area based on the seismic landslides that occurred in the 2013 Lushan earthquake when the ratio is 1, which means the number of non-landslides is equal to the number of landslides. Therefore, we kept

the non-landslide sample with the same number as the landslide samples. The non-landslide sample was selected by the spatially balanced sampling method on the area within the initial filtered area but excluded the landslide areas.

Validation and optimal threshold extraction

The ROC (receiver operating characteristics) curve was applied to validate the model performance of logistic regression model for the modelling result in Italy and Japan, respectively. The ROC curve is a quantitative measure to evaluate the performance of the evaluation model by calculating the value of the area under the curve (AUC) and has been shown as a reliable tool for the landslide susceptibility evaluation model (Vakhshoori and Zare 2018). The ROC curve provides an analysis based on true positive rate (TPR) and false positive rate (FPR) at any classification threshold of the delineated result. The ROC curve has been widely applied in the validation of landslide susceptibility models and measurement of model

applicability (Lin et al. 2017; Raja et al. 2016). If the area under the ROC curve (AUC) value is above the value of 0.5, which is the value of the random classifier, it shows that the model has better performance than the random classifier. A higher AUC value indicates that the evaluation model is more reliable. The most ideal AUC value of ROC curve would be 1, which means in any given threshold, all the TPR is 1 while the FPR is 0.

Not all the LSI values indicated landslide occurrence. However, it is difficult to identify a unified threshold to define which value means a landslide potential occurrence simply based on the LSI value for the landslide susceptibility result, because the landslide inventories that are applied as training datasets cannot include all the landslide events that actually occurred. Therefore, selecting a proper threshold based on ROC analysis depends on the study's purposes (Cantarino et al. 2018). An optimal threshold for LSI activation value, calculated by finding the balanced point between TPR and true negative rate (TNR), was applied as the threshold to extract the activated mask, those areas with an activated LSI value that means landslide occurrence. In other words, the minimum value of the difference between TPR and TNR (TNR is equal to the value of 1-FPR, false positive rate) was calculated as the optimal threshold.

Model for LDam formation susceptibility

Assessing the LDam formation susceptibility evaluation is to predict the LDam occurrence probability on specific river reaches. For LDam formation, landslide volume is a significant variable that contributes to the damming probability, while its accuracy is hard to measure (Tacconi Stefanelli et al. 2020). Previous research on LDam formation susceptibility applied the empirical relationship between landslide area and landslide volume based on the local landslide inventories to achieve the landslide volume (Fan et al. 2014; Tacconi Stefanelli et al. 2020). Such a power law correlation between landslide area and landslide volume has been widely acknowledged although the coefficients of the empirical relationship depended on regions, landslide types, or slope materials (Simonett 1967; Hovius et al. 1997; Guzzetti et al. 2009; Larsen et al. 2010).

Many LDam formation indices have used landslide volume and river/valley width as important components (Swanson et al. 1986; Tacconi Stefanelli et al. 2016, 2018). A relationship between landslide volume and river width also showed a potential threshold for LDam formation according to a global-scale LDam dataset (Wu

et al. 2022). Therefore, we estimate the LDam formation susceptibility at specific river reaches by combining the LSI, landslide area, and river width, as shown in Eq. (3). We achieved the LSI result of the areas near specific river reach by combining all the LSI value from the slope drainage unit nearby.

$$P(\text{LDam formation}) = \frac{\log \sum_{i=1}^n LSI_n \times SA_n}{\log RW} \quad (3)$$

The LSI_n is the median landslide susceptibility index of the adjacent delineated slope drainage unit n , SA_n is the surface area of the slope drainage unit n , and RW is the river width at specific river reach, which can be directly accessed on MERIT Hydro (Yamazaki et al. 2019). The method did not rely on any specific empirical relationships, such as the relationship between landslide volume and landslide area. By combining all the delineated slope drainage units that are near the specific river reach, it is likely that these areas encompass all the regions where an adjacent potential landslide occurrence could potentially block the river. This is because over 94% of landslides that have dammed rivers from a global-scale LDam dataset have had a length of less than 3 km (Wu et al. 2022), while the delineated slope drainage unit in this study has a mean length of 3 km and a mean width 1.5 km.

Results

Landslide susceptibility evaluation results and validations

The input variables for landslide susceptibility showed different relative contributions for the two study areas. The top three contributing variables in Italy are topographical slope degree (coefficient equal to 3.3264), soil erosion (2.7336), and clay fraction of soil (2.0425), while those for Japan were the specific type of lithology (acid plutonic rocks, -1.7111), landcover (shrubland, 2.1241), and potential natural vegetation classes (cool-temperate rainforest, -2.0414). Some variables, such as temperature, precipitation, clay fraction in soil, and lithology, have contrary contributions to the final landslide susceptibility evaluation result in Japan and Italy. The varying results in variable contribution imply that local landslide inventories are crucial for accurately assessing landslide susceptibility and capturing localised contributions of variables, even when the same input variables and data are used. More details

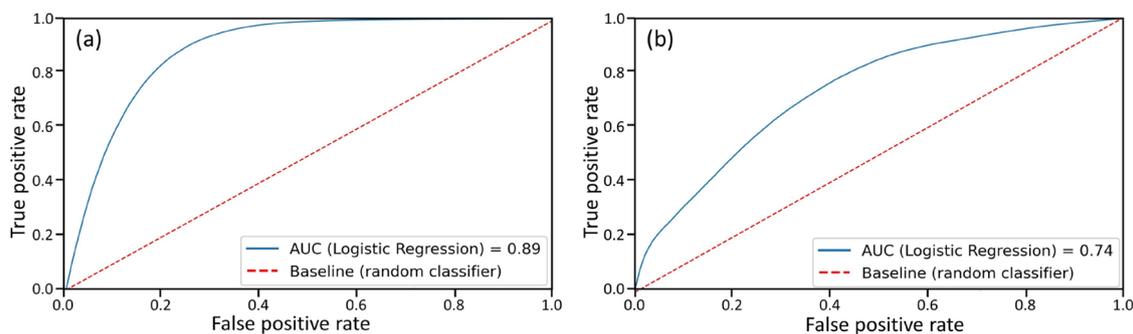


Fig. 5 The ROC curve of logistic regression model for landslide susceptibility evaluation in Italy (a) and Japan (b)

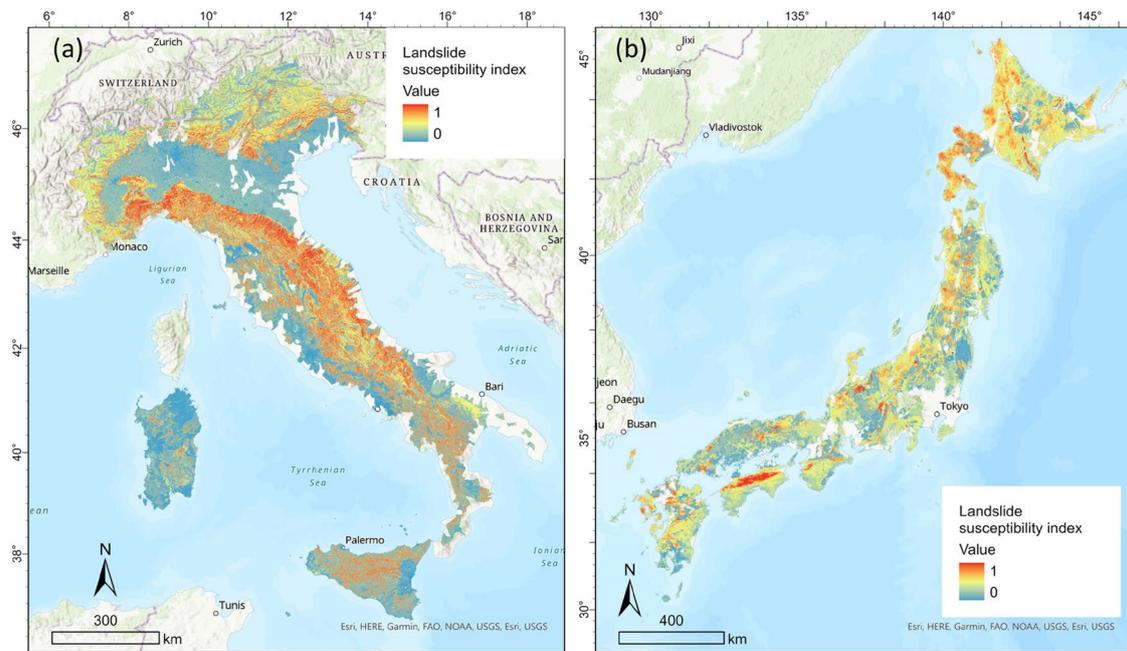


Fig. 6 Landslide susceptibility evaluation result in Italy (a) and Japan (b)

show the importance of variables contributing to landslide susceptibility results in Italy and Japan can be found at the table in Supplementary material Section 1.6.

Landslide susceptibility evaluation results for Italy and Japan were also validated by AUC value under the ROC curve. The fact that the logistic regression model of Italy landslide susceptibility evaluation presents an AUC value of 0.89 indicates the model has a great performance (Fig. 5(a)). The AUC value of logistic model performance in Japan is 0.74, which is lower, but still an acceptable value for a model performance (Fig. 5(b)).

After the ROC curve validation, the landslide susceptibility result can be used to identify where a landslide is more likely to occur given the LSI value. The mountainous regions along the European

Alps and Apennine in Italy exhibit the highest susceptibility to landslides, as indicated in Fig. 6(a). This observation aligns with the most landslide-susceptible classification area in Italy from the continental-scale landslide susceptibility evaluation result reported by Günther et al. (2014). Similarly, In Japan, the mountainous areas in the Hida Mountains, northern Shikoku, Kyoshu, Ōu Mountains, and Hokkaido are the areas prone to landslide occurrences (Fig. 6(b)), which correspond to area with the highest landslide density from a national-scale study in Japan by Paudel et al. (2016).

The optimal LSI threshold in Italy is 0.53 (Fig. 7(a)), while the threshold in Japan is 0.34 (Fig. 7(b)). The area with LSI values above these optimal thresholds was identified as the activated areas that have potential landslide occurrence.

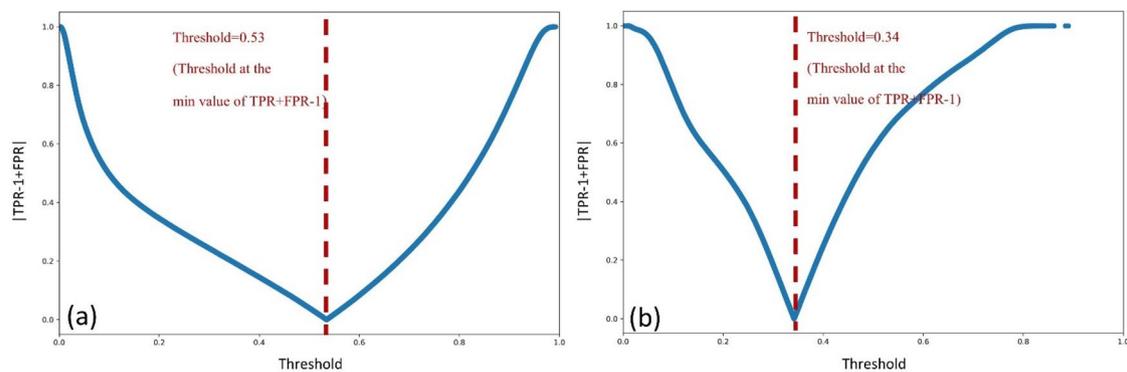


Fig. 7 The optimal threshold for LSI result modelled in Italy (a) and Japan (b)

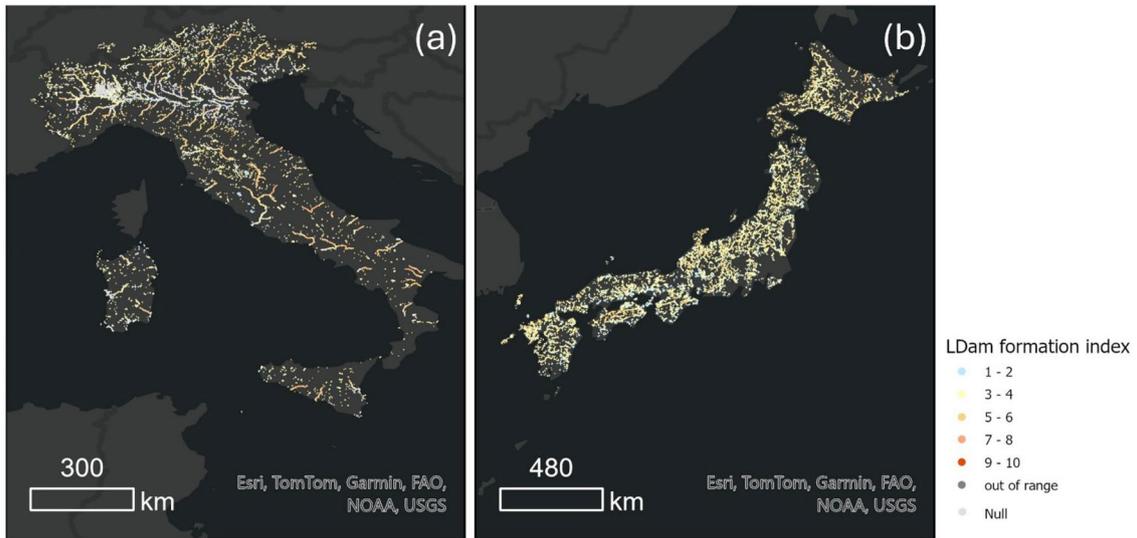


Fig. 8 Landslide dam formation index result in Italy (a) and Japan (b) (each dot with LDam formation index represented the centre of a river reach presented as 90-m resolution grid in MERIT Hydro; Figs. S1–8 in supplementary presents two instances of visualising the LDam formation index on a smaller scale)

Landslide dam formation susceptibility evaluation results with validations

The LDam formation index result of Italy is shown in Fig. 8(a). Approximately 39% of river reaches in Italy were identified with a null value of the LDam formation index due to a river width too large for LDam formation or due to the absence of the probability of landslide occurrence in the adjacent areas; these river reaches are mainly located in the plains. Likewise, Fig. 8(b) displays the LDam formation index results for Japan, indicating that around 31% of the river reaches have a null value. To provide more detail on the LDam formation index at smaller scales, the delineated LDam formation index along two river longitudinal profiles, whose river width and river channel presentation were delineated from MERIT hydro, can

be found at Supplementary material Section 1.7. This confirms that the river reaches with LDam records have relatively high values of the LDam formation index, as would be expected.

To validate the effectiveness of the LDam formation index, we conducted LDam formation index comparisons at all river reaches from global fluvial datasets and validated these indexes at the location of LDam records. Those LDam records which are located in lakes/reservoirs were excluded from the comparison as the LDam index would be low due to the lake presence, such as at inundation lakes or reservoirs behind old LDams. The LDam formation indexes for the reaches with LDam records in Italy ranged from 2.385 to 20.460, with a median value of 4.404 (Fig. 9(a)). Similarly, the LDam index for the reaches with LDam records in Japan ranged from 1.745 to 6.431, with a median value of 3.732 (Fig. 9(b)). A similar

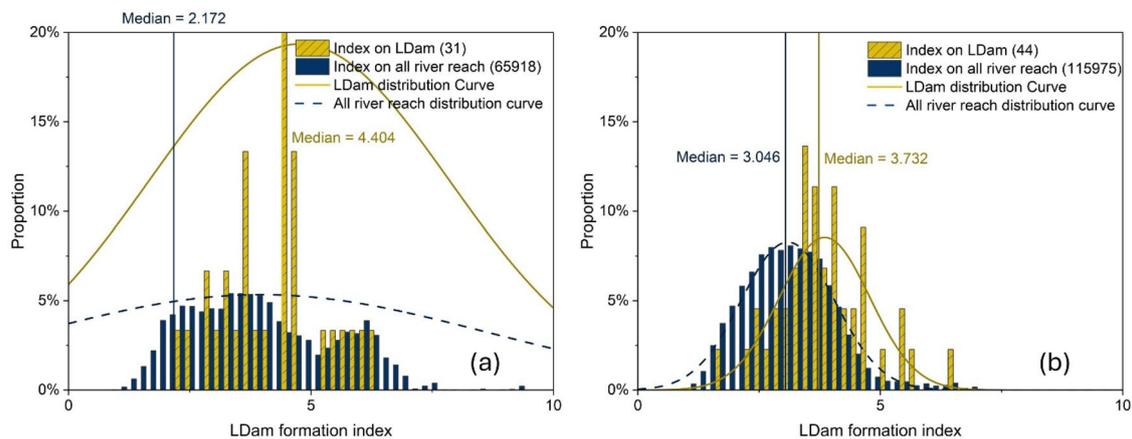


Fig. 9 Validation of LDam index value at LDam records and all river reaches present in MERIT Hydro for the results in Italy (a) and Japan (b) (scaling to the LDam index from 0 to 10; full size figure can be found at Supplementary material Section 1.7; numbers are the sample size)

method had been applied on rockfall susceptibility evaluation validation (Fernandez-Hernández et al. 2012). The LDam records have a relatively higher value of the LDam formation index compared to those of all the river reaches with valid indexes. The sample size of LDam records with valid LDam formation index depends on the existence of river reaches represented in global fluvial datasets, especially those on the upstream catchment areas (Italy: 13%; Japan: 32% before excluding those LDam records currently located in the lakes).

Discussions

Applicability of framework

The proposed LDam formation framework can be directly applied to other areas. Most of the LDam and landslide datasets that applied in this study, as well as global fluvial datasets, are open access, which enables the application of this methodology to other study areas using the same data sources. Another advantage of applying this framework is that it relied very little on information from empirical relationships, such as landslide volume estimation. With less dependence on site-specific empirical relationships, the pre-assessment of LDam formation on a large scale could be immediately conducted without the involvement of comprehensive landslide databases.

The framework can also provide a preliminary result for the research on a more local scale. The performance of this framework in two large-scale study regions, Italy and Japan, showed its applicability to other large-scale studies for exploring LDam formation, as most of the LDam formation index values at the LDam records are relatively high compared with all the valid indexes on the river reaches. By comparing our LDam formation susceptibility evaluation result with other research on more local scale on the Arno River basin (Fig. 10), we found the river reaches with relatively high

value of our LDam formation index shown in the main Arno River between Florence to Arezzo, which is consistent to the areas with high damming susceptibility of landslides. However, it should be noted that some tributaries (especially those around Mandrioli) were not represented in the global fluvial datasets in our analysis and thus some areas with high dam formation landslide occurrence probability does not show any valid LDam formation index. Further smaller-scale analysis could also use this framework with similar input data, if data resources on a local scale with more details and finer resolution are available.

Uncertainties

Three main uncertainties could affect the LDam formation evaluation framework according to the methods and results: (1) the potential diversity of landslide susceptibility result, (2) the focus of river channel representation and LDam record, and (3) their inconsistency.

Landslide susceptibility results were impacted by the data sources and modelling procedures. Limitations in large-scale landslide datasets, such as the data quality and comprehensiveness, always become a large uncertainty source for data-driven models, especially for the large-scale ones. Landslide types were undistinguished in this study as the detailed information was not available for most large-scale and open-access landslide datasets, but they crucially affected the landslide run-out distance, area, and volume (Corominas 1996; Larsen et al. 2010; Fan et al. 2014). In a literature review Reichenbach et al. (2018) noted similar uncertainties in landslide susceptibility evaluation, with only a few of 565 peer-reviewed articles from 1983 to 2016 addressing specific landslide types, and even then, used varying classification systems. Furthermore, Lin et al. (2017) reported that specific landslide types are rarely applied in large-scale landslide susceptibility evaluations. Even when using identical input variables and employing similar data processing procedures for landslide susceptibility evaluation,

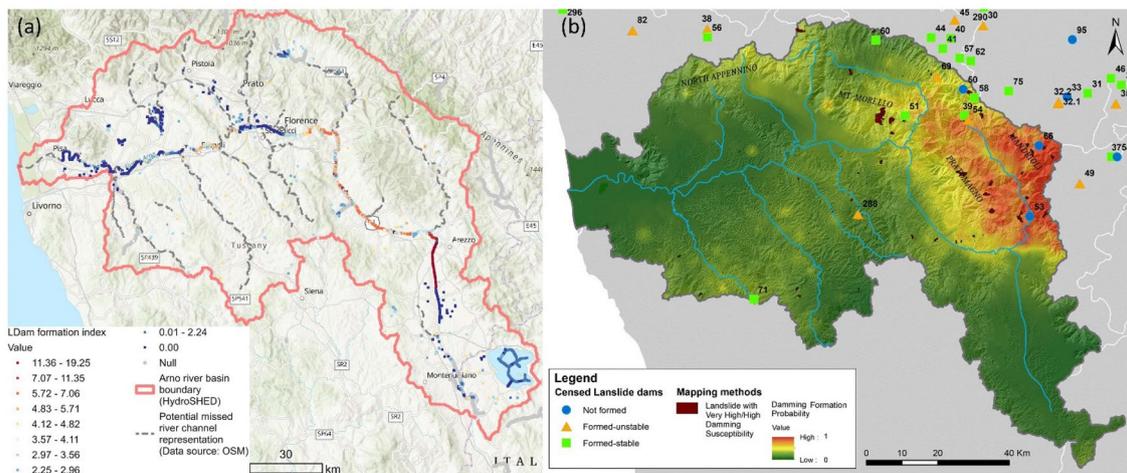


Fig. 10 Comparison of LDam formation results in Arno River basin, Tuscany, Italy (~8200 km²): **a** LDam formation index on river reaches derived from global fluvial datasets conducted by this study (each dot represented a 90-m river reach on MERIT hydro; grey dotted line represented those river channel without representation on global scale fluvial dataset); **b** Damming formation probability of landslides based on local-scale datasets (Tacconi Stefanelli et al. 2020)

it is inevitable to encounter varying contributions of variables to the evaluation result. Moreover, uncertainties could rise due to the diversity of the types within specific categorised variables, such as the vegetation and lithology types. However, such uncertainty could have relatively little effect on the following LDam formation analysis because the landslide susceptibility result, derived from different methods, data, and scope, in a specific area could still show largely similar results (Stanley and Kirschbaum 2017). Within the realm of such uncertainty, it is essential to acknowledge the ongoing deformations that affect the slopes involved in mass rock creep processes (Chigira and Kiho 1994), which can further evolve in rock avalanches. It is important to note that the occurrence of such rock avalanches is not primarily determined by force-driven failure mechanisms but rather by viscosity-driven processes (Marmoni et al. 2023). The case of the Scanno rock avalanche in Italy which caused the damming of the homonymous lake along the Tasso River is a nice example to show this impact (Esposito et al. 2013). The event was dated with a back analysis by Della Seta et al. (2017), while the role of creep processes in providing ultimate scenarios of LDam formation was also demonstrated. Such a process is particularly efficient in high-mountain areas as in the case of the Zagros Mountain (Iran), where giant rock avalanches caused

forming a LDam with enormous volume result in a lake-system evolution over thousands of years before the present (Delchiaro et al. 2019). However, such ongoing deformation could be under-represented by the landslide susceptibility result given the relatively coarse resolution of the data source (Fig. 11).

The inconsistency in representing river channels and the focus on LDam records can result in the omission of numerous streams or upstream rivers that could have a large potential to be dammed in the result of LDam formation susceptibility assessment. More than 80% of LDam records are typically located on the upstream catchment with drainage areas less than 500 km² (Wu et al. 2022), where these have steeper hillslopes for drainage area delineation and thus can lead to a limited representation for upstream rivers. The catchments at this scale were classified as streams or small rivers according to Bernhofen et al. (2021)'s classification. For global fluvial datasets, the representation of river channels relies on the quality and resolution of input data, such as DEM and climate data (Dottori et al. 2016), and the computational efficiency of global flood models (Bernhofen et al. 2021). The different sizes of the threshold for river channel representation can lead to the flood exposure difference (Bernhofen et al. 2021) and can directly influence the output of the LDam

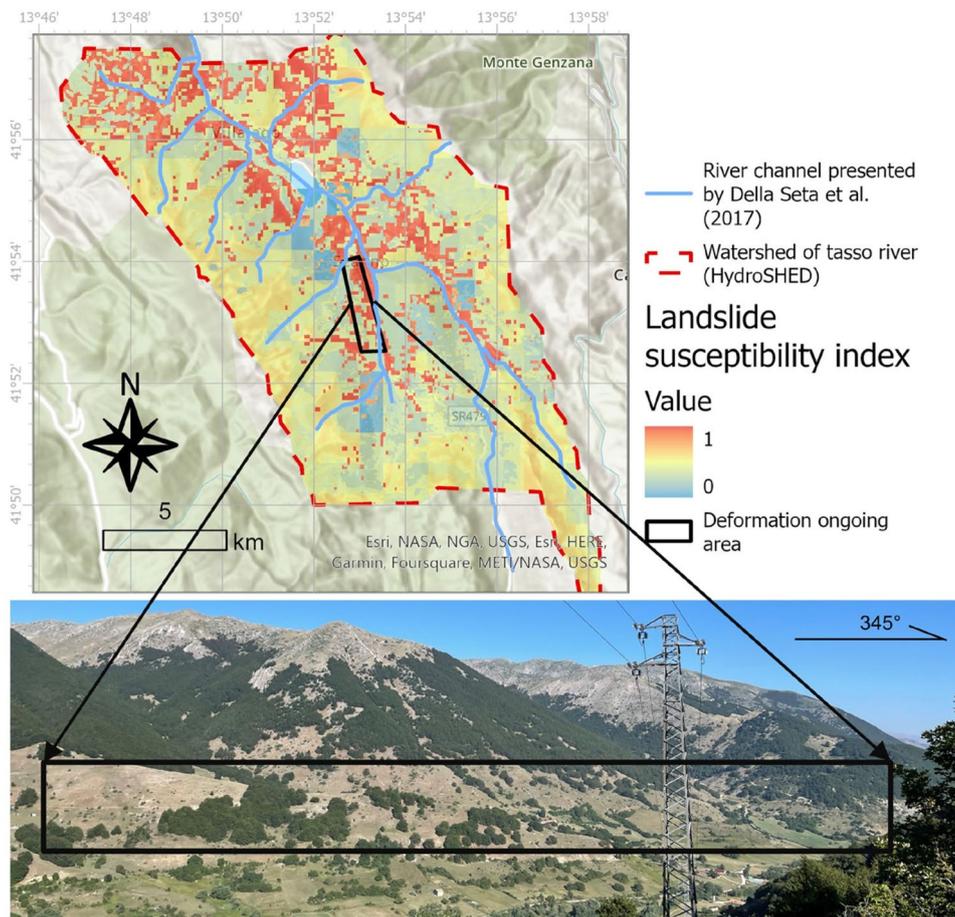


Fig. 11 Landslide susceptibility evaluation result in Tasso River, Abruzzo, Italy, and an ongoing deformation area, because of a mass rock creep process, surveyed on the field

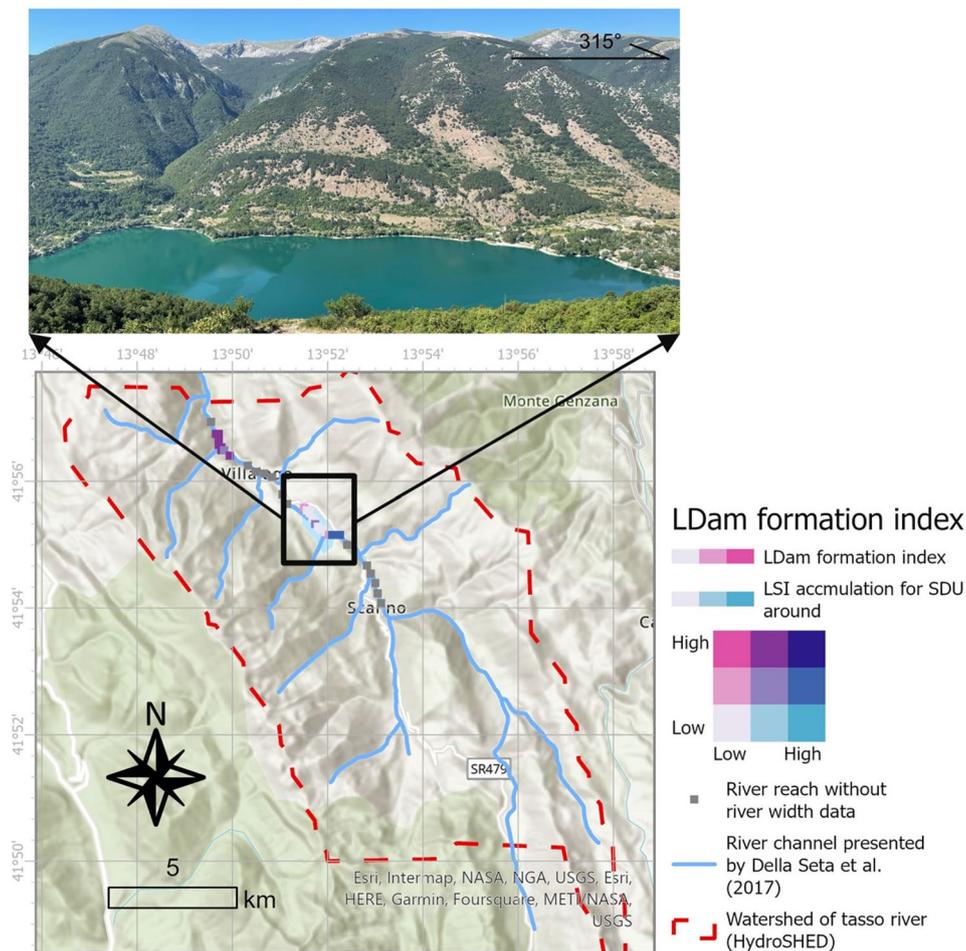


Fig. 12 LDam formation susceptibility evaluation result in Tasso River, Abruzzo, Italy, and the field picture of Scanno rock avalanche impounded lake (Lago di Scanno/Scanno Lake) (SDU, delineated slope drainage unit; LSI, landslide susceptibility index; the square represented the centre of a river reach shown as 90-m resolution grid in MERIT Hydro)

formation index through the river width data. Currently, global fluvial datasets may have more complete and accurate data for larger rivers compared with the upstream rivers where LDams are prone to occur as a good representation of smaller rivers remains to be resolved (Yamazaki et al. 2019). Figure 12 shows the representation of the LDam formation index at river reaches from MERIT Hydro in the Tasso River basin, Italy. The river reach crossing the Scanno Lake (Lago di Scanno) was represented well, while the other narrow river reaches did not contain valid river width data or were not represented in the global fluvial dataset. We used a colour matrix presenting landslide susceptibility from the adjacent slope drainage units and LDam formation index on each river reach to show the contribution of landslide susceptibility result to the final LDam formation index. The comparison shows that the area around Scanno Lake is presently in a relatively stable condition since either landslide susceptibility or LDam formation index is low. Such a result is consistent with the slope failures, already occurred in past (Della Seta et al. 2017), which reduced the potential of new landslide detachment from the slopes which surround the Scanno Lake.

Conclusions

Previous research has attempted to assess the probability of LDam formation; however, they relied heavily on the local-scale datasets or the empirical relationships based on local data and focused more on one side of the physical LDam formation processes and the occurrence of landslides of sufficient magnitude to obstruct river (damming formation probability of landslides). The LDam formation susceptibility evaluation framework was developed by combining landslide susceptibility evaluation and global fluvial datasets to estimate the LDam formation susceptibility on river reaches. Most of the input data, including the landslide datasets that were collected globally, global fluvial datasets, and input data for landslide susceptibility, are openly accessible.

The framework was built based on the unit of the delineated slope drainage and tested in Italy and Japan with all the results being validated against the data of LDam records. The logistic regression model employed for landslide susceptibility evaluation in Italy demonstrated excellent performance, achieving an AUC value of 0.89. In Japan, the logistic model achieved an AUC value of 0.74, which is considered an acceptable model performance. The

results obtained by quantifying the LDam formation index demonstrated the effectiveness of such an index to present the probability of LDam formation at specific river reaches because the data distribution of the LDam formation index for reaches with LDam records shows higher values compared to other river reaches. Moreover, the locations of relatively high LDam index are consistent with the LDam record positions along the river longitudinal profiles and the areas with relatively high damming formation probability of landslides. These results indicate the framework's capability to assess LDam formation susceptibility at river reaches in different large-scale regions around the world as the data source and relationship that applied to establish this framework are all global scale.

Overall, the developed framework offers a valuable tool for assessing LDam formation susceptibility at river reaches globally. Its reliance on open access data, including landslide datasets and global fluvial datasets, enhances its accessibility and applicability across various regions, even though such large-scale datasets introduce certain uncertainties for final results. The framework's performance in Italy and Japan, along with the demonstrated effectiveness of the LDam formation index, further supports its potential for widespread implementation and use.

The results obtained by quantifying the LDam formation susceptibility in this study can be utilised by international or national institutions as a preliminary exploration and hazard prevention of LDam formation prior to actual landslides occurring. This will allow mitigation measures, such as improving land use planning, applying early warning systems on adjacent slopes or rivers, and undertaking targeted infrastructure projects to reduce the risk of slope failure, be implemented to mitigate the risk of a LDam event. Continued research and collaboration, especially the smaller-scale research, can further refine the framework, expand its applicability, and strengthen its contributions to global efforts in managing and mitigating LDam-related risks.

The research shifted focus in LDam formation probability mapping from primarily landslide-centric DFL occurrences to exploring river reaches susceptible to LDam formation. This transition involved modifying procedures from calculating landslide volume and run-out based on empirical relationships to quantifying LDam formation probability, considering both fluvial characteristics and the likelihood of landslides in vulnerable areas. Instead of using a fixed uniform distance threshold from hillslopes to rivers, the new LDam susceptibility framework employs a geomorphological-based buffer zone derived from slope drainage delineation and spatial proximity. The framework utilised landslide susceptibility evaluation in a new way, a well-studied aspect that can be efficiently updated when new landslide susceptibility or GFDs are available.

Acknowledgements

Special thanks go to Dr. Anja Dufresne for providing the global-scale rockslide and rock avalanches dataset, and Prof. Alexander Strom for providing the rockslides and rock avalanches dataset in central Asia.

Author contribution

HW design the study, conduct the major jobs of data collection and analysis, and draft the manuscript; MT, WM, and RF contribute to forming the research objectives, supporting the data analyses, and interpretations; MT provide the Dominica catchment boundaries; HW, SM, CE, GMM, and GSM conducted a field visit to Tasso River

basin; HW and MT contribute to the proofreading. All the authors had contributed to the revisions of the manuscript and discussions.

Funding

Hang Wu received the University of Leeds Scholarship in Association with the China Scholarship Council (grant no. 201906400001).

Availability of data and materials

For all the landslide and landslide dam datasets, the direct references linked to the datasets are provided. The delineated slope drainage areas for LDam formation can be found at <https://doi.org/10.5518/1346>.

Code availability

Not applicable for this research.

Declarations

Ethics approval Not applicable for this research.

Consent to participate Not applicable for this research.

Consent for publication Not applicable for this research.

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Allen GH, Pavelsky TM (2018) Global extent of rivers and streams. *Science* 361(6402):585–588
- Alvioli M, Guzzetti F, Marchesini I (2020) Parameter-free delineation of slope units and terrain subdivision of Italy. *Geomorphology* 358:107124
- Argentin AL, Robl J, Prasicek G, Hergarten S, Hölbling D, Abad L, Dabiri Z (2021) Controls on the formation and size of potential landslide dams and dammed lakes in the Austrian Alps. *Nat Hazards Earth Syst Sci* 21(5):1615–1637

- Ayalew L, Yamagishi H (2005) The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains. *Central Japan Geomorphology* 65(1–2):15–31
- Bernhofen MV, Trigg MA, Sleigh PA, Sampson CC, Smith AM (2021) Global flood exposure from different sized rivers. *Nat Hazards Earth Syst Sci* 21(9):2829–2847
- Borrelli P, Ballabio C, Yang JE, Robinson DA, Panagos P (2022) GloSEM: high-resolution global estimates of present and future soil displacement in croplands by water erosion. *Sci Data* 9(1):406
- Budimir MEA, Atkinson PM, Lewis HG (2015) A systematic review of landslide probability mapping using logistic regression. *Landslides* 12(3):419–436
- Calvello M, Pecoraro G (2018) FranelItalia: a catalog of recent Italian landslides. *Geoenviron Disasters* 5(13):1–16
- Cantarino I, Carrion MA, Goerlich F, Martinez Ibañez V (2018) A ROC analysis-based classification method for landslide susceptibility maps. *Landslides* 16(2):265–282
- Canuti P, Casagli N, Ermini L (1998) Inventory of landslide dams in the Northern Apennine as a model for induced flood hazard forecasting. In: Andah K (ed) *Managing Hydro-geological Disasters in a Vulnerable Environment*. CNR-GNDCI Publication 1900. CNR-GNDCI-UNESCO (IHP), Perugia, pp 189–202
- Casagli N, Ermini L (1999) Geomorphic analysis of landslide dams in the Northern Apennine. *Trans Jpn Geomorphol* 20(3):219–249
- Cencetti C, De Rosa P, Fredduzzi A (2020) Characterization of landslide dams in a sector of the central-northern Apennines (Central Italy). *Heliyon* 6(6):e03799
- Chigira M, Kiho K (1994) Deep-seated rockslide-avalanches preceded by mass rock creep of sedimentary rocks in the Corominas Akaishi Mountains, central Japan. *Eng Geol* 38(3–4):221–230
- Corominas J (1996) The angle of reach as a mobility index for small and large landslides. *Can Geotech J* 33(2):260–271
- Costa JE, Schuster RL (1988) The formation and failure of natural dams. *Geol Soc Am Bull* 100(7):1054–1068
- Delchiaro M, Della Seta M, Martino S, Dehbozorgi Nozaem MR, Nozaem R (2019) Reconstruction of river valley evolution before and after the emplacement of the Giant Seymareh Rock Avalanche (Zagros Mts., Iran). *Earth Surf Dynam* 7:929–947
- Della Seta M, Esposito C, Marmoni GM, Martino S, Scarascia Mugnozza G, Troiani F (2017) Morpho-structural evolution of the valley-slope systems and related implications on slope-scale gravitational processes: new results from the Mt. Genzana case history (Central Apennines, Italy). *Geomorphology* 289:60–77
- Dikshit A, Sarkar R, Pradhan B, Acharya S, Dorji K (2019) Estimating rainfall thresholds for landslide occurrence in the Bhutan Himalayas. *Water* 11(8):1616
- Domènech G, Alvioli M, Corominas J (2019) Preparing first-time slope failures hazard maps: from pixel-based to slope unit-based. *Landslides* 17(2):249–265
- Dottori F, Salamon P, Bianchi A, Alfieri L, Hirpa FA, Feyen L (2016) Development and evaluation of a framework for global flood hazard mapping. *Adv Water Resour* 94:87–102
- Dou J, Tien Bui D, Yunus AP, Jia K, Song X, Revhaug I, Xia H, Zhu Z (2015) Optimization of causative factors for landslide susceptibility evaluation using remote sensing and GIS data in parts of Niigata. *Japan Plos One* 10(7):e0133262
- Dufresne A, Ostermann M, Preusser F (2018) River-damming, late-Quaternary rockslides in the Ötz Valley region (Tyrol, Austria). *Geomorphology* 310:153–167
- Dufresne A, Siebert L, Bernard B (2021) Distribution and geometric parameters of volcanic debris avalanches. In: Roverato M, Dufresne A, Procter J (eds) *Volcanic Debris Avalanches - From Collapse to Hazards*. *Advances in Volcanology*. Springer Heidelberg, pp 75–90
- Eilander D, Van Verseveld W, Yamazaki D, Weerts A, Winsemius HC, Ward PJ (2021) A hydrography upscaling method for scale-invariant parameterization of distributed hydrological models. *Hydrol Earth Syst Sci* 25(9):5287–5313
- Erener A, Düzgün HSB (2011) Landslide susceptibility assessment: what are the effects of mapping unit and mapping method. *Environ Earth Sci* 66(3):859–877
- Ermini L, Casagli N (2003) Prediction of the behaviour of landslide dams using a geomorphological dimensionless index. *Earth Surf Processes Landforms* 28(1):31–47
- Esposito C, Bianchi-Fasani G, Scarascia-Mugnozza G (2013) Quaternary gravitational morpho-genesis of Central Apennines (Italy): Insights from the Mt. Genzana case history. *Tectonophysics* 605:96–103
- Esposito G, Matano F (2021) *Campi Flegrei Landslide Geodatabase (CAFLAG)*. 4TU.ResearchData. Dataset. <https://doi.org/10.4121/14440757.v2>. Accessed 20 Feb 2022
- Fan X, Dufresne A, Siva Subramanian S, Strom A, Hermanns R, Tacconi Stefanelli C, Hewitt K, Yunus AP, Dunning S, Capra L, Geertsema M, Miller B, Casagli N, Jansen JD, Xu Q (2020) The formation and impact of landslide dams – state of the art. *Earth Sci Rev* 203 <https://doi.org/10.1016/j.earscirev.2020.103116>
- Fan XM, Rossiter DG, van Westen CJ, Xu Q, Gorum T (2014) Empirical prediction of coseismic landslide dam formation. *Earth Surf Proc Land* 39(14):1913–1926
- Fan X, van Westen CJ, Xu Q, Gorum T, Dai F (2012) Analysis of landslide dams induced by the 2008 Wenchuan earthquake. *J Asian Earth Sci* 57:25–37
- Feng D, Gleason CJ, Yang X, Allen GH, Pavelsky TM (2022) How have global river widths changed over time? *Water Resour Res* 58(8):e2021WR031712
- Fernandez-Hernández M, Paredes C, Castedo R, Llorente M, de la Vega-Panizo R (2012) Rockfall detachment susceptibility map in El Hierro Island, Canary Islands. *Spain Nat Hazards* 64:1247–1271
- Fick SE, Hijmans RJ (2017) WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *Int J Climatol* 37(12):4302–4315
- Frasson RPD, Pavelsky TM, Fonstad MA, Durand MT, Allen GH, Schumann G, Lion C, Beighley RE, Yang X (2019) Global relationships between river width, slope, catchment area, meander wavelength, sinuosity, and discharge. *Geophys Res Lett* 46(6):3252–3262
- Gong P, Wang J, Yu L, Zhao Y, Zhao Y, Liang L, Niu Z, Huang X, Fu H, Liu S, Li C, Li X, Fu W, Liu C, Xu Y, Wang X, Cheng Q, Hu L, Yao W, Zhang H, Zhu P, Zhao Z, Zhang H, Zheng Y, Ji L, Zhang Y, Chen H, Yan A, Guo J, Yu L, Wang L, Liu X, Shi T, Zhu M, Chen Y, Yang G, Tang P, Xu B, Giri C, Clinton N, Zhu Z, Chen J, Chen J (2013) Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *Int J Remote Sens* 34(7):2607–2654
- Günther A, Van Den Eeckhaut M, Malet JP, Reichenbach P, Hervás J (2014) Climate-physiographically differentiated Pan-European landslide susceptibility assessment using spatial multi-criteria evaluation and transnational landslide information. *Geomorphology* 224:69–85
- Guzzetti F, Ardizzone F, Cardinali M, Rossi M, Valigi D (2009) Landslide volumes and landslide mobilization rates in Umbria, central Italy. *Earth Planet Sci Lett* 279(3–4):222–229
- Guzzetti F, Carrara A, Cardinali M, Reichenbach P (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology* 31:181–216
- Guzzetti F, Reichenbach P, Cardinali M, Galli M, Ardizzone F (2005) Probabilistic landslide hazard assessment at the basin scale. *Geomorphology* 72(1):272–299
- Hartmann J, Moosdorf N (2012) The new global lithological map database GLiM: a representation of rock properties at the Earth surface. *Geochem Geophys Geosyst* 13(12) <https://doi.org/10.1029/2012GC004370>
- Hawker L, Uhe P, Paulo L, Sosa J, Savage J, Sampson C, Neal J (2022) A 30 m global map of elevation with forests and buildings removed. *Environ Res Lett* 17(2):024016
- Hengl T, de Jesus JM, MacMillan RA, Batjes NH, Heuvelink GB, Ribeiro E, Samuel-Rosa A, Kempen B, Leenaars JG, Walsh MG, Gonzalez MR (2014) SoilGrids1km — global soil information based on automated mapping. *PLoS ONE* 9(8):e105992
- Hengl T, Walsh MG, Sanderman J, Wheeler I, Harrison SP, Prentice IC (2018) Global mapping of potential natural vegetation: an assessment of machine learning algorithms for estimating land potential. *PeerJ* 6:e5457
- Hovius N, Stark CP, Allen PA (1997) Sediment flux from a mountain belt derived by landslide mapping. *Geology* 25(3):231–234

- Jacobs L, Kervyn M, Reichenbach P, Rossi M, Marchesini I, Alvioli M, Dewitte O (2020) Regional susceptibility assessments with heterogeneous landslide information: Slope unit- vs. pixel-based approach. *Geomorphology* 356:107084
- Jibson RW, Harp EL (2012) Extraordinary distance limits of landslides triggered by the 2011 Mineral, Virginia. *Earthquake Bull Seismol Soc Am* 102(6):2368–2377
- Jones ES, Mirus BB, Schmitt RG, Baum RL, Burns WJ, Crawford M, Godt JW, Kirschbaum DB, Lancaster JT, Lindsey KO, McCoy KE, Slaughter S, Stanley TA (2019) Summary metadata – landslide inventories across the United States, U.S. Geological Survey data release. <https://doi.org/10.5066/P9E2A37P>. Accessed 20 Feb 2022
- Kirschbaum DB (2019) High Mountain Asia Landslide Catalog, Version 1. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. <https://doi.org/10.5067/5ST0TZCD9RQ3>. Accessed 23 Feb 2022
- Kirschbaum DB, Stanley T, Zhou Y (2015) Spatial and temporal analysis of a global landslide catalog. *Geomorphology*. <https://doi.org/10.1016/j.geomorph.2015.03.016>
- Kornejady A, Ownegh M, Bahremand A (2017) Landslide susceptibility assessment using maximum entropy model with two different data sampling methods. *CATENA* 152:144–162
- Korup O (2002) Recent research on landslide dams - a literature review with special attention to New Zealand. *Prog Phys Geog* 26(2):206–235
- Korup O (2004) Geomorphometric characteristics of New Zealand landslide dams. *Eng Geol* 73(1):13–35
- Larsen IJ, Montgomery DR, Korup O (2010) Landslide erosion controlled by hillslope material. *Nat Geosci* 3(4):247–251
- Lehner B, Grill G (2013) Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrol Processes* 27(15):2171–2186
- Lin L, Lin Q, Wang Y (2017) Landslide susceptibility mapping on a global scale using the method of logistic regression. *Nat Hazards Earth Syst Sci Discuss* 17(8):1411–1424
- Linke S, Lehner B, Ouellet Dallaire C, Ariwi J, Grill G, Anand M, Beames P, Burchard-Levine V, Maxwell S, Moïdu H, Tan F, Thieme M (2019) Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. *Sci Data* 6(1):1–15
- Marmoni GM, Martino S, Censi M, Menichetti M, Piacentini D, Mugnozza GS, Torre D, Troiani F (2023) Transition from rock mass creep to progressive failure for rockslide initiation at Mt. Conero (Italy). *Geomorphology* 473–108750. <https://doi.org/10.1016/j.geomorph.2023.108750>
- Martinello C, Cappadonia C, Conoscenti C, Agnesi V, Rotigliano E (2020) Optimal slope units partitioning in landslide susceptibility mapping. *J Maps* 17(3):152–162
- Martinello C, Cappadonia C, Conoscenti C, Rotigliano E (2022) Landform classification: a high-performing mapping unit partitioning tool for landslide susceptibility assessment—a test in the Imera River basin (northern Sicily, Italy). *Landslides* 19(3):539–553
- Martino S, Caprari P, Fiorucci M, Marmoni GM (2022) Italian Catalogue of Earthquake-Induced Ground Failures (CEDIT), University of Rome Sapienza. <https://gdb.ceri.uniroma1.it/index.php/view/map/?repository=cedit&project=Cedit>. Accessed 08 Sept 2022
- Melillo M, Brunetti MT, Peruccacci S, Gariano SL, Guzzetti F (2016) Rainfall thresholds for the possible landslide occurrence in Sicily (Southern Italy) based on the automatic reconstruction of rainfall events. *Landslides* 13:165–172
- Nadim F, Kjekstad O, Peduzzi P, Herold C, Jaedicke C (2006) Global landslide and avalanche hotspots. *Landslides* 3(2):159–173
- National Research Institute for Earth Science and Disaster Prevention of Japan (2014) Digital archive for landslide distribution maps. National Research Institute for Earth Science and Disaster Prevention of Japan. https://dii-opac.bosai.go.jp/publication/nied_tech_note/landslidemap/index.html. Accessed 23 Feb 2022 (in Japanese)
- Palau RM, Hürlimann M, Berenguer M, Sempere-Torres D (2020) Influence of the mapping unit for regional landslide early warning systems: comparison between pixels and polygons in Catalonia (NE Spain). *Landslides* 17(9):2067–2083
- Palladino MR, Viero A, Turconi L, Brunetti MT, Peruccacci S, Melillo M, Luino F, Deganutti AM, Guzzetti F (2018) Rainfall thresholds for the activation of shallow landslides in the Italian Alps: the role of environmental conditioning factors. *Geomorphology* 303:53–67
- Paudel U, Oguchi T, Hayakawa Y (2016) Multi-resolution landslide susceptibility analysis using a DEM and random forest. *Int J Geosci* 07(05):726–743
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830
- Raja NB, Çiçek I, Türkoğlu N, Aydın O, Kawasaki A (2016) Landslide susceptibility mapping of the Sera River Basin using logistic regression model. *Nat Hazard* 85(3):1323–1346
- Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F (2018) A review of statistically-based landslide susceptibility models. *Earth Sci Rev* 180:60–91
- Rossi M, Guzzetti F, Reichenbach P, Mondini AC, Peruccacci S (2010) Optimal landslide susceptibility zonation based on multiple forecasts. *Geomorphology* 114(3):129–142
- Saito H, Korup O, Uchida T, Hayashi S, Oguchi T (2014) Rainfall conditions, typhoon frequency, and contemporary landslide erosion in Japan. *Geology* 42(11):999–1002
- Scarascia Mugnozza G, Bianchi Fasani G, Esposito C, Martino S, Saroli M, Di Luzio E, Evans SG (2006) Rock avalanche and mountain slope deformation in a convex dip-slope: the case of the Maiella massif, Central Italy. In: Evans SG, Scarascia Mugnozza G, Strom A, Hermanns RL (eds) *Landslide from Massive Rock Slope Failure*. Springer
- Schmitt RG, Tanyas H, Nowicki Jessee MA, Zhu J, Biegel KM, Allstadt KE, Jibson RW, Thompson EM, van Westen CJ, Sato HP, Wald DJ, Godt JW, Gorum T, Xu C, Rathje EM, Knudsen KL (2017) An open repository of earthquake-triggered ground-failure inventories. U.S. Geological Survey data release collection. <https://doi.org/10.5066/F7H70DB4>. Accessed 23 Feb 2022.
- Shafieiganjeh R, Ostermann M, Schneider-Muntau B, Gems B (2022) Assessment of the landslide dams in Western Austria, Bavaria and Northern Italy (part of the Eastern Alps): data inventory development and application of geomorphic indices. *Geomorphology* 415:108403
- Shao X, Ma S, Xu C, Zhou Q (2020) Effects of sampling intensity and non-slide/slide sample ratio on the occurrence probability of coseismic landslides. *Geomorphology* 363:107222
- Simonett DS (1967) Landslide distribution and earthquakes in the Bewani and Torricelli Mountains, New Guinea, statistical analysis. In: Jennings JN, Mabbutt JA (eds) *Landform Studies from Australia and New Guinea*. Cambridge University Press, Cambridge, pp 64–84
- Stanley T, Kirschbaum DB (2017) A heuristic approach to global landslide susceptibility mapping. *Nat Hazards (dordr)* 87(1):145–164
- Strom A, Abdrahmatov K (2018) Rockslides and rock avalanches of Central Asia - distribution, morphology, and internal structure. Elsevier
- Struble WT, Roering JJ, Burns WJ, Calhoun NC, Wetherell LR, Black BA (2021) The preservation of climate-driven landslide dams in Western Oregon. *J Geophys Res Earth Surf* 126(4):e2020JF005908
- Swanson FJ, Oyagi N, Tominaga M (1986) Landslide dams in Japan. In: Schuster RL (ed) *Landslide Dams: Processes, Risk, and Mitigation*. ASCE, pp 131–145
- Tacconi Stefanelli C, Casagli N, Catani F (2020) Landslide damming hazard susceptibility maps: a new GIS-based procedure for risk management. *Landslides* 17(7):1635–1648
- Tacconi Stefanelli C, Segoni S, Casagli N, Catani F (2016) Geomorphic indexing of landslide dams evolution. *Eng Geol* 208:1–10
- Tacconi Stefanelli C, Villimek V, Emmer A, Catani F (2018) Morphological analysis and features of the landslide dams in the Cordillera Blanca. *Peru Landslides* 15(3):507–521
- Trigila A, Iadanza C, Spizzichino D (2010) Quality assessment of the Italian Landslide Inventory using GIS processing. *Landslides* 7:455–470

- Trigg MA, Dehghani MS, Kesete YY, Carr AB, Trigg SG, Zekkos D, Lopez D, Pertierra M, van Westen CJ, Jetten V, Ogden FL (2023) Realities of bridge resilience in small island developing states. *Mitigation Adapt Strategies Global Change* 28(1). <https://doi.org/10.1007/s11027-022-10035-4>
- van Westen CJ, Van den Bout B, Lombardo L, Jetten V (2020) Reducing debris and water flow risks from landslides for Infrastructure Resilience of the Commonwealth of Dominica, Report to the World Bank by Faculty of Geo-Information Science and Earth Observation (ITC) University of Twente, Enschede, the Netherlands (Unpublished report)
- van Westen C J, Zhang J (2018). Landslides and floods triggered by Hurricane Maria (18 September, 2017) in Dominica. Digital or Visual Products, UNITAR-UNOSAT. <http://www.unitar.org/unosat/node/44/2762>. Accessed 23 May 2021
- Vakhshoori V, Zare M (2018) Is the ROC curve a reliable tool to compare the validity of landslide susceptibility maps? *Geomatics, Nat Hazards Risk* 9(1):249–266
- Verdin KL, Verdin JP (1999) A topological system for delineation and codification of the Earth's river basins. *J Hydrol* 218(1–2):1–12
- Wu H, Trigg MA, Murphy W, Fuentes R (2022) A new global landslide dam database (RAGLAD) and analysis utilizing auxiliary global fluvial datasets. *Landslides* 19(3):555–572
- Yamazaki D, Ikeshima D, Sosa J, Bates PD, Allen GH, Pavelsky TM (2019) MERIT Hydro: a high-resolution global hydrography map based on latest topography dataset. *Water Resour Res* 55(6):5053–5073
- Yamazaki D, O'Loughlin F, Trigg MA, Miller ZF, Pavelsky TM, Bates PD (2014) Development of the global width database for large rivers. *Water Resour Res* 50(4):3467–3480

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10346-024-02306-9>.

Hang Wu (✉) · **Mark A. Trigg**

School of Civil Engineering, University of Leeds, Leeds, UK
Email: hangwu94@foxmail.com; hangwu@qq.com

William Murphy

School of Earth and Environment, University of Leeds, Leeds, UK

Raul Fuentes

Chair of Geotechnical Engineering and Institute of Geomechanics and Underground Technology, RWTH-Aachen, Aachen, Germany

Salvatore Martino · Carlo Esposito · Gian Marco Marmoni · Gabriele Scarascia Mugnozza

Department of Earth Sciences and Research Centre for Geological Risk – CERl, University of Rome “Sapienza”, Rome, Italy