Engineering Geology 272 (2020) 105639

Contents lists available at ScienceDirect

Engineering Geology

journal homepage: www.elsevier.com/locate/enggeo

Impact of mapping strategies on rockfall frequency-size distributions

Sandra Melzner[a,*], Mauro Rossi[b], Fausto Guzzetti[b]

[a] Geoconsult ZT GmbH, Hölzlstraße 5, Wals, Salzburg 5071, Austria
[b] Consiglio Nazionale delle Ricerche, Istituto di Ricerca per la Protezione Idrogeologica, Via Madonna Alta 126, Perugia 06128, Italy

ARTICLE INFO

Keywords:
Rockfall
Mapping methods
Inventory
Catalogue
Frequency-size distributions
Statistics
Design block

ABSTRACT

Rockfall frequency size distributions are used in Austria for the definition of a design block for the planning of technical rockfall protection. Rockfall size datasets are often incomplete. Here, we study fifteen catalogues of rockfall size in Austria, Italy, and the USA to analyse the impact of the data collection and mapping methods on the representativeness of the catalogues and on the estimates of frequency-size statistics. To describe and compare the catalogues of rockfall size, we first use Empirical Cumulative Distribution Functions (ECDFs), followed by parametric distribution estimates in the form of Probability Density Functions (PDFs), and Cumulative Distribution Functions (CDFs). We discuss the output of Kolmogorov-Smirnov tests, the position of the frequency-size distribution rollover, and the *p*-value and the standard errors associated to the distribution parameters estimates to determine the reliability of our model results. In addition, we analyse the variations in the modelled CDFs for different percentiles of the frequency-size distributions to describe and discuss the representativeness of the rockfall catalogues. Our results show that different mapping strategies may affect the estimates of frequency-size distribution of rock fall volume, a relevant information when evaluating the possible impacts of rockfall processes. We conclude offering recommendations for rockfall mapping, and the use and of a non-parametric statistical method being capable to deal with small datasets, which is very typical when dealing with rockfall data. Such recommendations help for a correct dimensioning of designing rockfall mitigation measures.

1. Introduction

A rockfall is a type of extremely rapid mass movement characterized by a potentially long travel distance from the release (source) to the deposition area. Local rock properties and slope geometry condition the rockfall behaviour, making it difficult to prepare rockfall inventories with consistent and comprehensive spatial and temporal information. Methods used to collect information on rockfalls depend on the size and complexity of the study area, and on the scope and the resources available for the investigation, and include (i) field mapping, (ii) systematic search of archives, chronicles, newspapers and technical and event reports, (iii) visual inspection and in-situ monitoring of rock cliffs, (iv) interpretation of remote sensing imagery or photogrammetry, and (v) rockfall dating techniques.

In the literature, different statistical methods were proposed to analyse spatial, temporal, and size information of rockfalls. Implicitly or explicitly, all methods require "completeness" and "representativeness" of the rockfall catalogues and series (Corominas et al., 2017a; De Biagi et al., 2017b, De Biagi et al., 2017b; Malamud et al., 2004; Rossi et al., 2010). However, determining the level of completeness or the representativeness of the catalogues and time series is not trivial, and this can jeopardize the significance of the statistical analyses performed on the rockfall records.

Published rockfall statistics are rare, covering mostly small areas (slope scale) and short time periods, and are focused mainly on volume-frequency analysis or for the definition of a "design block" (Agliardi et al., 2009; Brunetti et al., 2009; Corominas et al., 2017a, 2017b; Crosta et al., 2015; De Biagi et al., 2017b; Dussauge et al., 2003; Dussauge-Peisser et al., 2002; Guzzetti et al., 1994; Lambert and Bourrier, 2013; Lari et al., 2014; Macciotta, 2014; Malamud et al., 2004).

In this work, we analyse 15 catalogues with information on the size of rockfalls obtained using different mapping methods for seven study areas in Austria, Italy, and the USA. We compare the rockfall information, and we discuss the impact of the different mapping methods on the quality, representativeness, and completeness of the size distributions of the rockfalls.

Different statistical approaches are introduced and exploited to estimate frequency-size rock fall statistics. The advantages and limitations of the different approaches are presented and discussed providing a

* Corresponding author.
E-mail address: sandra.melzner@geoconsult.com (S. Melzner).

https://doi.org/10.1016/j.enggeo.2020.105639
Received 24 November 2019; Received in revised form 2 March 2020; Accepted 12 April 2020
Available online 15 April 2020
0013-7952/

guidance for their proper usage and promoting standard and comparable procedures to analyse frequency-size rock fall statistics for the definition of a design block, which is a fundamental part when it comes to the planning of technical rockfall protection measures.

2. Background

2.1. Definitions

In this work, we define a "primary rockfall" a rock that detaches from a rock wall or rock outcrop by sliding, toppling or falling, and a "secondary rockfall" the remobilization of a previously deposited (rockfall) boulder, typically on a slope. After detachment, a rock either falls along a rock wall or moves along a slope by bouncing and flying along ballistic trajectories, or by rolling and sliding. A rock may fragment into individual rockfall boulders impacting the ground and/or trees. The rockfall boulders proceed independently down the slope without interacting with another and having an effect on the propagation mechanisms. A rockfall boulder stops when it has lost enough energy.

The "rockfall release area" ("source", "detachment") is the area where the rockfall initiates. The "transport area" ("propagation", "transit") is the section between the "source" and the "deposition area", where most of the rockfall travel paths (trajectories) occur. The "deposition area" ("accumulation") is the area where most of the boulders and fragments stop and deposit. In places, separation of the release, transport and depositional area is uncertain, and the main processes (detachment, propagation, stopping) coexist (Fig. 1). As an example, along the transport area a number of boulders and fragments can come to rest or can remobilised as "secondary rockfalls", and in the depositional area individual boulders can travel shorter or longer distances. The maximum rockfall runout lengths ("maximum runout") is the stopping location of rockfall boulders, typically marked by "outlying rockfall boulders" in the valley floor (Fig. 1).

Information about rockfalls is stored in "inventories", "catalogues", and "records". These terms are often used synonymously in the literature. In this work, a "rockfall inventory" is a type of landslide inventory (Guzzetti et al., 2012) that contains geographical and typological information on the source, transport, and deposition area of rockfalls. Information about an individual rockfall or multiple rockfalls can be extracted from the rockfall inventory database in the form of different thematic "rockfall catalogues".

In this work, (i) a "catalogue of rockfall sizes" is a list of volumes of rockfalls or rockfall boulders, (ii) a "historical rockfall catalogue" lists

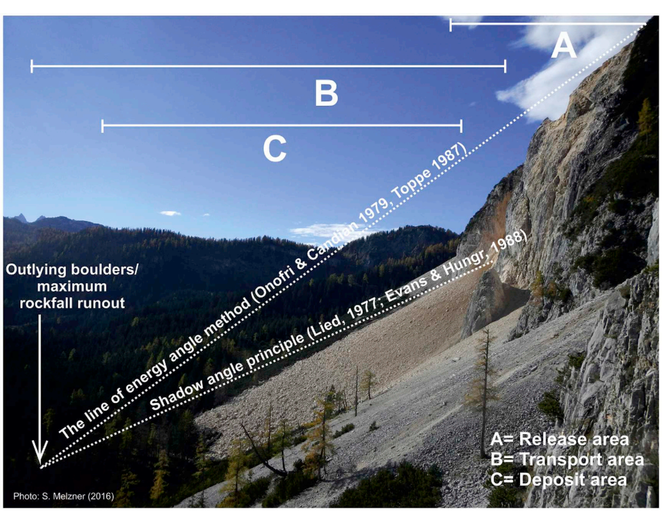

Fig. 1. Definition of the rockfall process terms. The separation of the release (A), transport (B) and depositional area (C) is fuzzy, and the main processes (detachment, propagation, stopping) coexist.

the time of occurrence of rockfalls, and (iii) a "historical catalogue of rockfall damage" is a list of information on the consequences of rockfalls over time. For some rockfalls, detailed temporal and spatial information and consequences may be existent, thus a combination of these catalogues is possible. A "historical rockfall record" or a "rockfall time series" show rockfalls over time for a given region, detailed spatial information is not always available.

We define a "rockfall scenario" as a potentially damaging rock block (boulder) size (volume), and its consequences. According to ONR 24810:2017 (2017), a "design block" refers to the return period of a block with a certain size (volume). "Rockfall magnitude" is used as a synonym for the rockfall volume, velocity or destructiveness of a rockfall, "rockfall intensity" is a proxy for rockfall magnitude, and is a measure of the destructiveness of a rockfall (Bunce et al., 1997; Crosta et al., 2015; Guzzetti et al., 2003; Hungr, 1997).

2.2. Quality of rockfall size catalogues

The collection of rockfall data is normally adjusted to the research objectives and the project framework i.e., the financial and temporal constraints, the project goals, and the size and settings of the study area. The characteristics and quality of the resulting rockfall catalogue on volumes depends on (i) the setting and characteristics of the study area (e.g., topography, geology, land use, forest cover), (ii) the accuracy of the base and thematic maps, (iii) the methods and techniques used, (iv) the source(s) of information, (v) the time available for the investigation, (vi) the experience of the investigators, and (vii) the available human, technological and economic resources. Depending on the scale of investigation, the individual rockfall features – e.g., release area, impact points, talus slopes and single deposited boulders – are represented by points, lines, or polygons.

Possible criteria for the evaluation of the "quality" of a catalogue of rockfall size can be related to the amount of data, level of detail and variability of information. The level of comprehensiveness of the catalogue can be described with "completeness" and the "representativeness" of the collected data and the variability of the geographical location of the single of multiple features encompassed by a rockfall: "Completeness" refers to the proportion of rockfalls contained in the catalogue respect to the total number rockfalls which have occurred. "Representativeness" refers to the degree of a given rock fall sample/subset to reflect the entire rockfall catalogue from which it is derived i.e., a representative rock fall sample should give unbiased statistical inference of what the population is like. "Thematic variability" refers to the amount of imprecision of the identification and classification of a rockfall or a given rockfall feature. "Geographic variability" refers to the amount of imprecision of the graphical representation of a rockfall feature to the real geographic position on the ground in the study area.

2.3. Rockfalls statistics and applications

Frequency-size analysis of rockfalls can be performed using non-parametric and parametric statistical approaches (Rossi et al., 2012, 2010) (Fig. 2).

Non-parametric approaches are based on the estimation of the frequency, or density, of rockfalls using histogram-based or kernel-based estimations. The methods may suffer from biases due to subjective choices of the width, breaks, and ranges of the size classes (the "bins", for histogram-based estimations) and of the bandwidth (for kernel-based estimations) (Fig. 2A). In many geological fields, histogram-based estimations are preferred to analyse frequency-size statistics (Fig. 2A). In our case, cumulating the histogram frequency count, and normalizing it by the total number of rockfalls in the record (Fig. 2B), we obtain the Empirical Cumulative Distribution Function (ECDF), which provides information on the probability to observe a rockfall smaller in size (e.g., volume) than an established threshold. The ECDF can be used to estimate the size of a rockfall corresponding to a given percentile

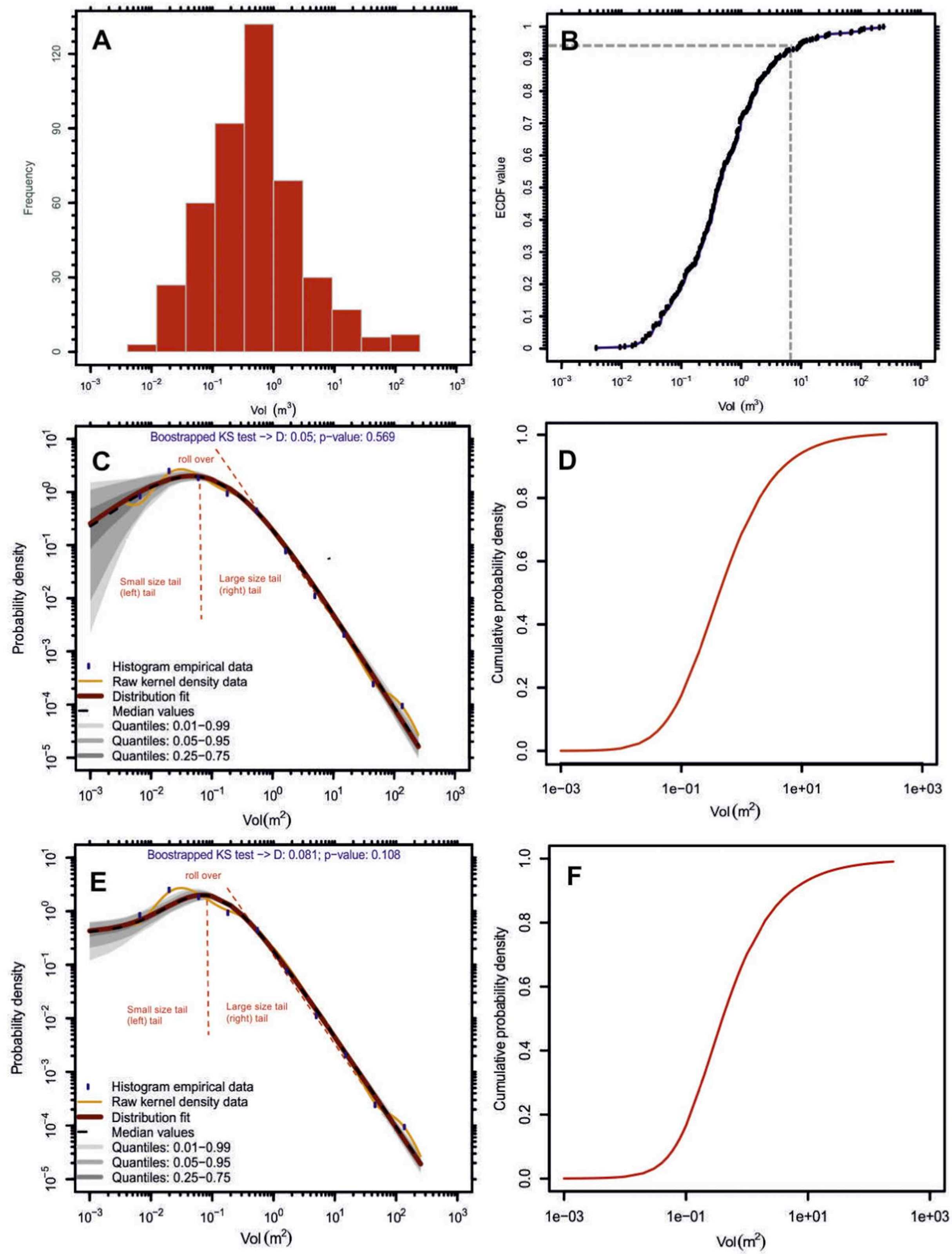

Fig. 2. Overview about statistical analysis methods. (A) Frequency size distribution; (B) Empirical cumulative size distribution function; (C) Probability density function of rockfall size- PDF, modelled as a double pareto; (D) Cumulative distribution function of size-CDF, modelled as a double pareto; (E) Probability density function of rockfall size- PDF, modelled as an inverse gamma; (F) Cumulative distribution functions- CDF, modelled as an inverse gamma; see Rossi et al., 2012 and appendix for equations.

e.g., the 95th percentile refers to 95% of the rockfalls smaller than a given volume (i.e., ONR 24810:2017, 2017).

Parametric statistical approaches require the *a-priori* selection of the (parametric) function to be used to model the Probability Density Function (PDF) and the Cumulative Density Function (CDF) of a given distribution. The three distribution functions (Pareto, Double Pareto, and Inverse Gamma) approximate the right tail of the empirical landslide size distribution (i.e., for the medium to very large sizes) with an inverse power-law (Fig. 2 C,E). When the landslide size decreases, the probability increases up to a maximum – the "rollover" – representing the modal (most frequent) landslide size value. For sizes smaller than the "rollover", the probability decreases with the size.

The cumulative distribution functions (CDFs) can be obtained integrating the PDFs (Fig. 2 D,F). By definition, the integral of the PDF has to be 1 (Rossi et al., 2012). Hence, the CDF varies between 0 and 1. In this work, the PDF (Fig. 2 C,E) shows the probability to observing a rockfall of a given volume $P(v_L)$, and the CDF (Fig. 2 D) shows the probability to observing a rockfall smaller than a given threshold volume, $P(V_L \leq v_L)$. Similarly to non-parametric distribution estimation i.e., the use of ECDF, CDF models can be used to estimate a rockfall volume exceeding or not a given percentile i.e., key information for the design of rockfall mitigation measures. Under the assumption that CDF and ECDF models are estimated from a representative rockfall volume sample, minor differences can be found in the derived percentile statistics.

3. Available data and analyses

3.1. Rockfall data sets

We used 15 catalogues listing rockfall volume in Austria, Italy, and the USA, to analyse the impact of the mapping methods on the quality, representativeness, and completeness of the size distribution of the rockfalls. Some of the catalogues are parts of comprehensive rockfall inventories comprising detailed information about other relevant rockfall features. Different mapping methods and strategies were adopted to collect the rockfall data in the different catalogues (Table 1).

3.2. Statistical analysis

We first use the Empirical Cumulative Distribution Functions (ECDF) to describe the individual catalogues of rockfall volumes (Fig. 2), and we then compare the results in terms of the impact of different factors including the mapping method and the dominant rock types. Factors influencing the volume and number of rockfalls include (i) lithology and structural geologic settings, (ii) position of the rockfalls on the slope, (iii) topographic factors, including slope geometry, or the presence of lakes and rivers at the bottom of the slope, and (iv) human-induced influences, including e.g., removal of boulders by local habitants (Fig. 3).

To estimate the parameters of the PDF and the CDF for the different catalogues, we used the statistical analysis tool developed by Rossi (2014). The tool adapts three parameter Double Pareto (DPS) function (Section 2.3) to analyse records of rockfall volumes. The selection of the DPS instead of the other two distribution functions (DP and IG) was based on the results of Kolmogorov-Smirnov (KS) tests, being the most appropriate model to describe the data (Table 2).

In this work, we do not consider the use of the Pareto (P) distribution (Barry, 1983; Brunetti et al., 2009) or the Generalized Pareto Distribution (GDP) (Barry, 2011; De Biagi et al., 2017a, 2017b), mostly due to the fact that their parameter's estimation requires a filtering (i.e., removal) of small rockfall volumes, hence being unable to properly characterize the size distribution and the related statistical estimations in the excluded volume range. The three DPS parameters α, β, t condition the shape of the probability distribution: α is the shape parameter that controls the slope of the large size tail. Larger α values correspond to steeper large size tails, which refers to a reduced proportion of large size values in the distribution above the rollover. The β parameter controls the slope of the small size tail and t controls indirectly the rollover position. Larger β correspond to steeper small size tails, which refers to a reduced proportion of small size values in the distribution below the rollover. To evaluate the significance of the modelled parameters (α, β, t), *t*-test was performed. In this test the t-statistics is used and the corresponding *p*-value (Table 2) is compared to different confidence levels to rank the significance of parameters. The following significance codes rank the significance of the parameters, from highly to not significant; [***] = p-value < 0.001, [***] = p-value < 0.01, [***] = p-value < 0.05, [**] = p-value < 0.1, [] = p-value ≥ 0.1. This information is helpful in determining/testing the correctness of the model choice in relation to the analysed dataset and its possible representativeness. Commonly, under the assumption of being using a representative dataset without measurements biases, in the case of one or more parameters resulting not significant, one possible conclusion is that it is appropriate to use a simpler model (i.e. the chosen model has too many parameters and the problem has been overparametrized). Additionally, bootstrap resampling was performed to evaluate the uncertainties associated to the probability density estimates (grey shades in Fig. 2 C, E). We determined variations of the (modelled) CDFs from the (empirical) ECDFs for three percentiles (50th, 75th, 95th) of the frequency-size distributions to describe and discuss the completeness/representativeness of the different datasets. Such variation was calculated as (CDF – ECDF) / ECDF × 100 (See Table 3).

In contrast to the non-parametric statistical approaches, the modelled distributions enable the estimation of sizes outside the mapped size-range of the ECDFs. In the comparison, we used CDF values calculated using the model parameters summarized in Table 2 shown in Fig. 4.

We determined the DPS model results analysing (i) the Kolmogorov-Smirnov (KS) tests, (ii) the presence of a rollover in the modelled data, and (iii) the values of the parameters *t* controlling the transition between the tails, α controlling the slope of the large tail, and β controlling the slope of the small tail, the corresponding standard errors and their related significance. The KS-Test, failed if the *p*-value was below 0.05 i.e., the selected significance level for the test. Standard errors (s.e.) are compared to the estimated values, expressing them as a percentage of the estimated values (s.e./estimated value) × 100); values of this index < 10% are assumed to be very good results, < 20% medium good results and > 50% are results of poor quality.

In addition to the previous analysis, we investigated the sensitivity of the CDF estimated from a selected catalogue to variations of sizes by filtering the sizes below and above given thresholds (Fig. 5). Thus, the resulting distributions give information about how sensitive the CDF model is to the filtering (i.e. the removal, not mapping) of smaller (Fig. 5A) or larger boulder (Fig. 5B) sizes.

4. Results

To identify the factors influencing the volume size distribution we considered factors such as (i) lithology and the structural setting (catalogues C_{UM}, C_{BB}, C_{OM}, C_{BT}, C_H, C_{YV}), (ii) the position of the rockfalls on the slope (catalogues C_{H4}, C_{H6} and C_{H7}), (iii) topographic factors, including slope geometry, and the presence of lakes and rivers at the bottom of the slope (catalogues C_{H2} and C_{H4}), and (iv) possible human influences, including e.g., removal of boulders by local habitants.

4.1. Empirical cumulative distribution functions of rockfall volume

The analysis of the ECDFs with respect to lithology (Fig. 3A and Tables 1–3) showed that the hard and very hard rocks of catalogues C_{OM} and C_{YV} form larger boulder sizes than the less hard rocks of the other catalogues. Catalogue C_{YV} also contains data about very large rockfalls and rock mass falls/rockslides, which were not considered in the other mapping strategies. The comparison of different datasets of carbonatic sedimentary rocks (Fig. 3B) revealed that variations in the ECDFs within and among the catalogues is evident particularly for larger percentiles. Considering the 80th percentile of all catalogues in Fig. 3B, the volumes range from 0.5 m^3 (C_{H7}) to 50 m^3 (C_{H3}), being

Table 1
Overview about catalogues of rockfall size.

Inventory	Catalogue	Country	Area size [km^2]	Number of data entries	Volume range	Time period	Rock type	Mapping strategy/data collection method/source of information
I_{UM}	C_{UM}	IT	950	133	0.004 < 200 m^3	1997	Carbonatic, sedimentary rocks	Rockfall boulders along roads, triggered by earthquake 1997; measuring of boulder volumes
I_{BB}	C_{BB}	AUT	17	464	0.00012 < 90 m^3	2008	Carbonatic, sedimentary rocks	Rockfall boulders in the maximum reach, basis for hazard zoning; measuring of boulder volumes
I_{BB}	C_{BB1}	AUT	17	381	0.02 < 90 m^3	2008	Carbonatic, sedimentary rocks	The catalogue "C_{BB1}" is a subset of "C_{BB}", Rockfall boulders in the maximum reach, basis for hazard zoning; measuring of boulder volumes
I_{BB}	C_{BB2}	AUT	17	56	0.027 < 40 m^3	2007	Carbonatic, sedimentary rocks	The catalogue of rockfall sizes "C_{BB2}" was compiled by another investigator in the same study area as catalogue "C_{BB}" and "C_{BB1}". For this catalogue, the assessment strategy focused on mapping rockfall boulders > 0.02 m^3 in the lower slope.
I_{OM}	C_{OM}	AUT	130	372	0.008 < 1400 m^3	2010	Metamorphic rocks	Rockfall boulders in the maximum reach, basis for hazard zoning; measuring of boulder volumes
I_{BT}	C_{BT}	IT	0.05	62	0.004 < 129.5 m^3	1997	Carbonatic, sedimentary rocks	Rockfall boulders on closed road, frequency assessment; measuring of boulder volumes
I_H	C_H	AUT	6,7	443	0.004 < 252 m^3	2014	Carbonatic, sedimentary rocks	Rockfall boulders in the maximum reach, basis for hazard zoning; measuring of boulder volumes, the catalogue "C_H" was split into four main structural geological sections for hazard assessment (see catalogues "$C_{H1\text{-}H5}$").
I_H	C_{H1}	AUT	1,6	162	0.004 < 252 m^3	2014	Carbonatic, sedimentary rocks	Devision of study area in structural sections, subset 1 of catalogue "C_H", Rockfall boulders in the maximum reach, basis for hazard zoning; measuring of boulder volumes
I_H	C_{H2}	AUT	x1,1	51	0.016 < 48 m^3	2014	Carbonatic, sedimentary rocks	Devision of study area in structural sections, subset 2 of catalogue "C_H", Rockfall boulders in the maximum reach, basis for hazard zoning; measuring of boulder volumes
I_H	C_{H3}	AUT	2,6	91	0.027 < 120 m^3	2014	Carbonatic, sedimentary rocks	Devision of study area in structural sections, subset 3 of catalogue "C_H", Rockfall boulders in the maximum reach, basis for hazard zoning; measuring of boulder volumes
I_H	C_{H4}	AUT	0,3	77	0.027 < 160 m^3	2014	Carbonatic, sedimentary rocks	Devision of study area in structural sections, subset 1 of catalogue "C_H", Rockfall boulders in the maximum reach, basis for hazard zoning; measuring of boulder volumes
I_H	C_{H5}	AUT	0,9	64	0.027 < 120 m^3	2014	Carbonatic, sedimentary rocks	The "C_{H5}" catalogue is a subset of the "C_{H3}" catalogue, covers 0.9 km^2 and comprises 64 rockfalls in the range 0.027 < V_R < 120 m^3. Rockfall boulders in the maximum reach, basis for hazard zoning; measuring of boulder volumes
I_H	C_{H6}	AUT	0,3	99	0.01 < 2.11 m^3	2014	Carbonatic, sedimentary rocks	Rockfall boulders in the maximum reach, basis for hazard zoning; measuring of boulder volumes
I_H	C_{H7}	AUT	0,3	27	0.01 < 0.72 m^3	2014	Carbonatic, sedimentary rocks	Rockfall boulders in the maximum reach, basis for hazard zoning; measuring of boulder volumes
I_{YV}	C_{YV}	US	3000	317	0.05–864	2011	Igneous rocks	Published and unpublished historical accounts and field studies

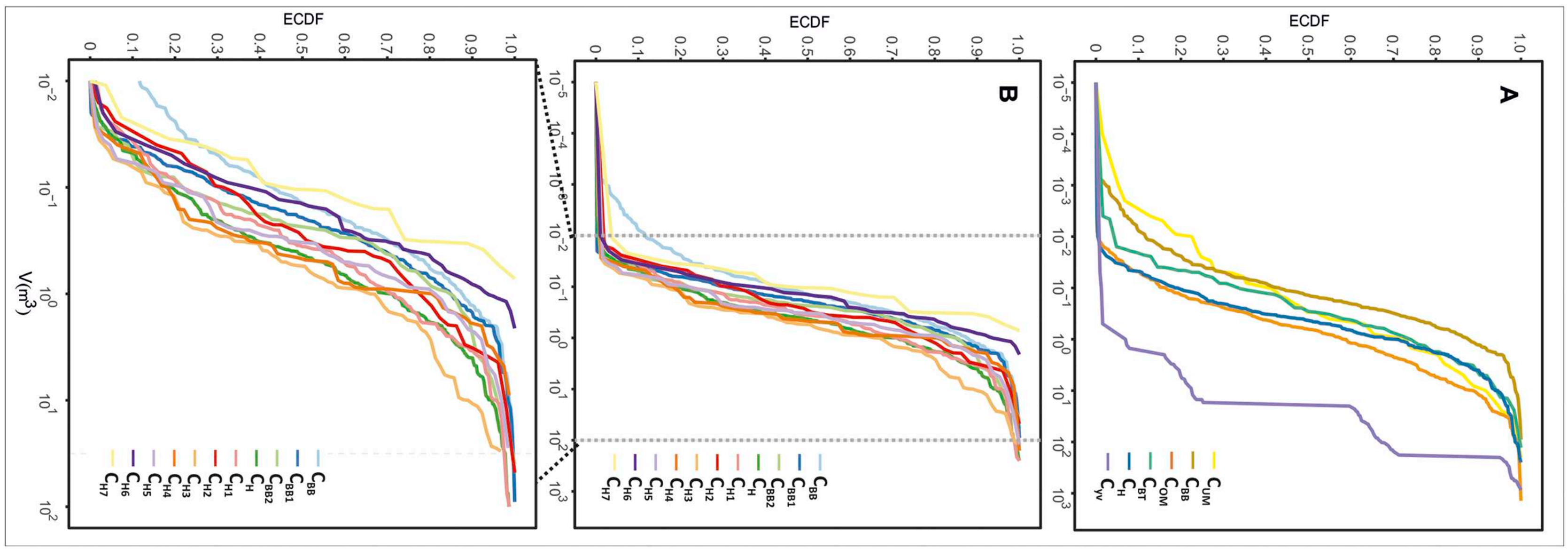

Fig. 3. Comparison of empirical cumulative distribution functions (ECDF) of all catalogues sizes, considering mapping different lithologies (A) and different mapping strategies (B).

Table 2
Comparison of the estimated parameters for PDF and CDF for the different catalogues of rockfall size using the Double Pareto Simplified (DPS) model. Significance codes rank the significance of the parameters, from highly significant [***] to not significant [*].

Catalogue	Rock types	Number of boulders	Range boulder size (m^3)	Slope large-sized tail (α value)	Slope small-sized tail (β value)	Transition between the tails (t value)	Rollover (m^3)	KS- test D/p-value
C_{UM}	Dolomite, limestone	133	0.004–200	0.64 ± 0.12 [***]	0.46 ± 0.45 [***]	0.46 ± 0.45 []	NA	0.075/0.808
C_{BB}	Dolomite, limestone	464	0.00012–90	1.09 ± 0.088 [***]	0.68 ± 0.06 [***]	0.26 ± 0.06 [***]	NA	0.063/0.298
C_{OM}	Paragneiss, orthogneiss	372	0.008–1400	0.71 ± 0.05 [***]	1.25 ± 0.25 [***]	0.25 ± 0.14 []	0.02	0.054/0.662
C_{BT}	Dolomite, limestone	62	0.004–129.5	0.68 ± 0.12 [***]	1.55 ± 0.92 []	0.05 ± 0.08 []	0.01	0.11/0.767
C_{H}	Limestone	443	0.004–252	0.80 ± 0.06 [***]	2.10 ± 0.63 [***]	0.10 ± 0.06 []	0.05	0.076/0.14
C_{YV}	Granite	317	0.05–864	0.63 ± 0.03 [***]	2.21 ± 0.35 [***]	2.00 ± 0.82 [*]	1.2	0.275/0
C_{H1}	Limestone	162	0.004–252	0.77 ± 0.09 [***]	1.94 ± 0.79 [*]	0.08 ± 0.79 []	0.04	0.073/0.697
C_{H2}	Limestone	51	0.016–48	0.87 ± 0.15 [***]	1.33 ± 0.57 [*]	0.14 ± 0.13 []	0.02	0.145/0.589
C_{H3}	Limestone	91	0.027–120	0.68 ± 0.06 [***]	6.16 ± 0.00 [***]	0.01 ± 0.01 [*]	0.07	0.117/0.5
C_{H4}	Limestone	77	0.027–160	1.34 ± 0.28 [***]	1.16 ± 0.33 [***]	0.53 ± 0.29 [.]	0.07	0.131/0.406
C_{H5}	Limestone	64	0.027–120	0.86 ± 0.15 [***]	6.18 ± 11.49 []	0.02 ± 0.07 []	0.08	0.109/0.794
C_{H6}	Limestone	99	0.01–2.11	1.25 ± 0.193 [***]	2.55 ± 1.32 []	0.08 ± 0.06 []	0.06	0.135/0.266
C_{H7}	Limestone	27	0.01–0.72	1.41 ± 0.52 [**]	2.06 ± 1.93 []	0.07 ± 0.10 []	0.04	0.148/0.888
C_{BB2}	Dolomite, limestone	56	0.027–40	0.88 ± 0.14 [***]	7.50 ± 12.96 []	0.02 ± 0.04 []	0.06	0.125/0.657
C_{BB1}	Dolomite, limestone	381	0.02–90	0.99 ± 0.06 [***]	3.07 ± 0.75 [***]	0.05 ± 0.02 [**]	0.05	0.049/0.714

[.] = p-value < .1, [] = p-value ≥ 0.1.
*** *p*-value < .001
** *p*-value < .01
* *p*-value < .05

Table 3
Variations of CDF to the ECDF, considering the 50th, 75th and 95th percentile. Variations are calculated in percent (CDF-ECDF/ECDF*100).

Catalogue	50th percentile			75th percentile			95th percentile		
	CDF [m^3]	ECDF [m^3]	Variation to ECDF [%]	CDF [m^3]	ECDF [m^3]	Variation to ECDF [%]	CDF [m^3]	ECDF [m^3]	Variation to ECDF [%]
C_{UM}	0.3	0.3	0.00	1.6	2.0	−20.00	32.0	22.0	45.45
C_{BB}	0.11	0.14	−21.43	0.4	0.4	0.00	3.1	2.1	55.00
C_{OM}	0.7	0.6	16.67	2.8	3.1	−9.68	35	28.4	23.24
C_{BT}	0.3	0.29	3.45	1.0	1.2	−16.67	14.5	13.2	9.85
C_{H}	0.4	0.4	0.00	1.4	1.4	0.00	12.7	10.8	17.59
C_{YV}	22.0	20.0	10.00	99.0	200.0	−50.50	1600	201.4	694.44
C_{H1}	0.3	0.36	−16.67	1.2	1.3	−7.69	11.9	9.8	21.43
C_{H2}	0.3	0.31	−3.23	0.9	0.92	−2.17	6.8	4.18	62.68
C_{H3}	0.6	0.55	9.09	2.1	1.8	16.67	27.2	27.5	−1.23
C_{H4}	0.5	0.48	4.17	1.1	1.0	10.00	4.4	3.39	29.79
C_{H5}	0.3	0.33	−9.09	0.9	0.87	3.45	7.1	5.6	26.79
C_{H6}	0.2	0.15	33.33	0.3	0.4	−25.00	1.4	1.01	38.61
C_{H7}	0.1	0.11	−9.09	0.2	0.31	−35.48	0.8	0.49	63.27
C_{BB2}	0.2	0.24	−16.67	0.7	0.6	16.67	4.8	7.86	−38.93
C_{BB1}	0.2	0.19	5.26	0.5	0.5	0.00	3.0	2.1	42.86

even larger for higher percentiles. For small volumes, the differences between the ECDFs are smaller. The analysis of the ECDF differences allow a meaningful interpretation of the impact of the mapping strategy.

4.2. Probability density and cumulative distribution functions of rockfall volume

Fig. 4 shows the modelled probability density functions (PDF), and the cumulative distribution functions (CDF) for all the considered catalogues of rockfall sizes modelled using DPS. For details on the estimated parameters see Tables 2 and 3.

We determined variations of the CDFs from ECDFs, for three percentiles (50th, 75th, 95th) of the frequency-size distributions, to describe and discuss the representativeness of the data sets (Table 3). For most of the rock types, when we consider the high percentile values, the CDFs (Fig. 4B, D, F, H) are higher than the ECDFs (Fig. 3A and B). This evidence supports the conclusion that the empirical data may not represent correctly the extremes i.e., the largest volumes in the catalogues. In these cases, modelled CDF should be preferred to ECDF based estimations. Only for C_{H3} and C_{BB2} the CDF values are lower than the ECDFs.

The subdivision of a study area in structural/geologic homogenous domains (Fig. 4C and Fig. 4D) proved to have an effect on the frequency-size distributions (C_H versus C_{H1}-C_{H5}). The orientation of joint mass structure influence highly the topographic settings and the slope form, and thus, often the rockfall size and runout lengths of rockfalls. The large volumes in the C_{H3} are the result of a nearly slope parallel dipping (dip-slope situation) of the bedding of the carbonatic rocks. In the case of C_{H1}, the slope-ward dipping bedding planes result in vertical cliffs, shorter runout-lengths and smaller rockfalls. The low probability density values for large volume rockfalls in catalogues C_{H2} and C_{H4} are probably due to the fact that the large boulders were deposited directly in the lake or torrent in the lower slope respectively and not mapped. Furthermore, all catalogues do not contain data about boulders of large rockfalls or rock mass collapses (rock mass fall/cliff fall), as it is the case for C_{YV}. The latter is a dominant process in the area of catalogue C_{H3}, and in two sections of catalogue C_{H1}.

The analysis of the comparison of the inventories mapped by two different investigators (Fig. 4E and Fig. 4F) indicates that the geographic position of the mapped boulders affects the frequency-size distributions, significantly. This is due to the fact that larger boulders have longer travel distances than smaller boulders. The higher percentage of larger sizes in catalogue C_{H4} is due to the fact that boulders where mapped in the lower slope (maximum reach of past events) by the first investigator, whereas the boulders catalogue C_{H6} were mapped in the middle of the slope by the second investigator. This supports the hypothesis that mapping strategies affect significantly the representativeness of catalogues of rock sizes. The difference in sizes are even higher if only the young boulders (C_{H7}) are considered.

In turn, analysis of the mapping focus of two different investigators (Fig. 4G and Fig. 4H) shows that both mapping strategies resulted in similar mappings (C_{BB} and C_{BB2}), if boulders < 0.02 m^3 are not considered (C_{BB1}). The high percentage of smaller boulders in C_{BB1} results in the lack of a distinct rollover (Fig. 4A and B), which is also the case for C_{UM}. C_{BB1} and C_{UM} with the lowest volume values, compared to those mapped in rest of the catalogues, not show a distinct rollover. This possibly suggests how mapping strategy affect statistical inference on size distribution. This may be important if the fragmentation of rockfalls along the slope is mapped. Rock fragmentation can be very high at the first impact at the base of the cliff of the rockfall (highest loss of energy) which may result in small fragments, as well for the very hard rocks of C_{YV}.

4.3. Variability of PDF and CDF

For most of the considered catalogues, results of the KS-Test (Table 2) show p values > 0.05, supporting the conclusion that the DPS is a model appropriate for the empirical rockfall volume distribution. Only for catalogue C_{YV} the DPS model performed poorly, with a $P = 0$. This reveals the sensitivity of the KS-test (which is known to be appropriate for continuous data) to discrete data, as in the case for the C_{YV} catalogue (Fig. 3A).

Comparing the position of the "rollover" for the examined datasets, the most frequent boulder volume is in the range $0.04 < V_L < 0.07$ m^3. Catalogues C_{OM}, C_{BT} and C_{H2} have the lowest rollover (0.2 m^3) and catalogue C_{H5} has the highest rollover (0.8 m^3). The catalogues C_{BB} and C_{UM} do not show a rollover in their frequency size distribution.

The standard errors (s.e.), expressed as a percentage of the estimated α values (slope of large sized rockfalls), reveal that the large size boulders distributions of catalogues C_{BB}, C_{OM}, C_H, C_{H3}, C_{BB1} with values lower than 10%, are well described by the DPS model, whereas for catalogues C_{UM}, C_{BT}, C_{H2}, C_{H4}, C_{H5}, C_{H6}, C_{BB2} the model is less accurate (values lower than 20%) and being even more inaccurate in case of catalogue C_{H7} (value greater than 30%). In contrast, the standard errors expressed as a percentage of the estimated β value are much higher than those estimated for α values, showing a less accurate model behaviour for small boulders sizes.

We determined variations of the CDFs from ECDFs, for three percentiles (50th, 75th, 95th) of the frequency-size distributions, to describe and discuss the representativeness of the data sets (Table 3). The

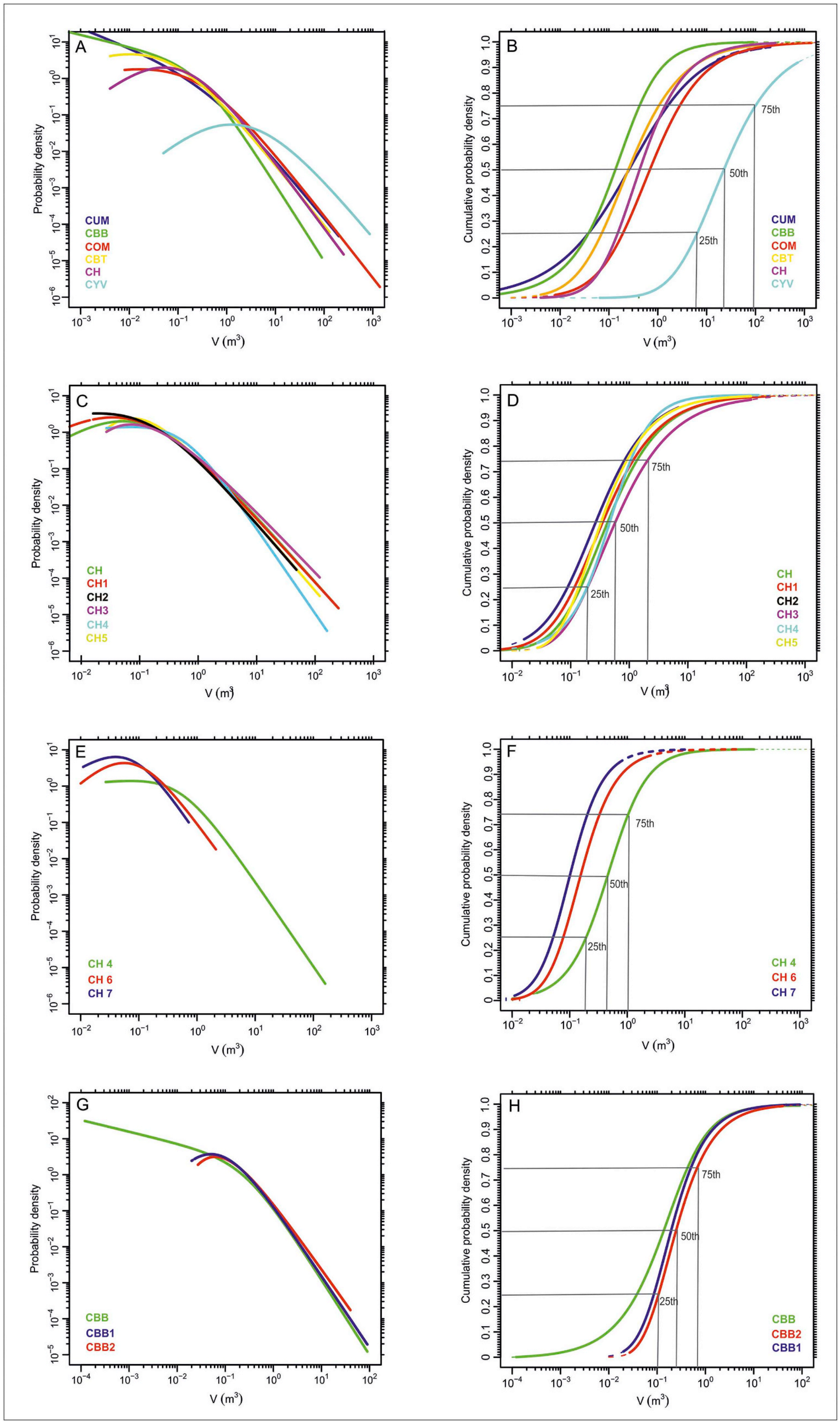

Fig. 4. Impact of mapping strategy on probability densities of rockfall sizes (A, C, E, G) and cumulative distribution function of rockfall size (B, D, F, H). Dashed curves in plots B, D, F, H show values of distribution function calculated outside the observed volume ranges approximating values of cumulative probability of 0 and 1. The three thin grey lines in the CDF plots (i.e., B, D, F, H) corresponding to 0.25 (25th percentile) 0.50 (50th percentile) and 0.95 (95th percentile).

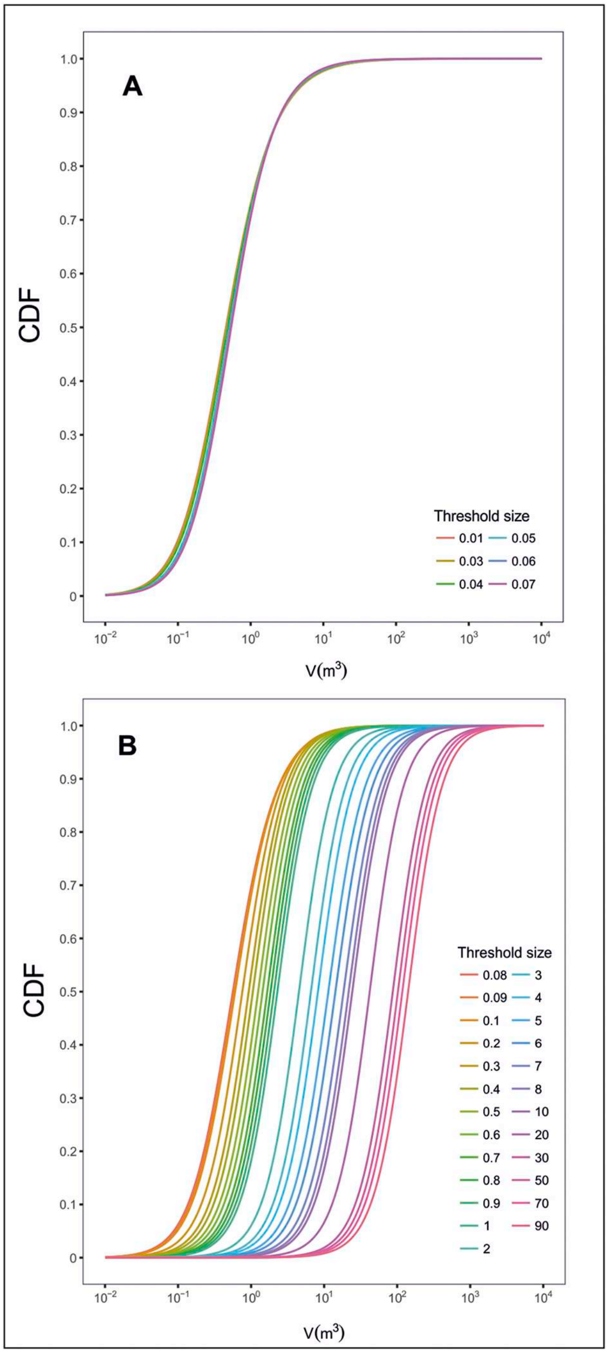

Fig. 5. Sensitivity of CDF estimations to variations in sizes: to not map small sizes (small sizes below roll-over, removal of small boulders) (A) and to not map large sizes (large sizes above roll-over, removal of large boulders) (B).

comparison of the CDFs with the ECDF (Table 3) show that most of the variations between modelled and empirical percentiles are within a (−/+) 50% range with respect to the empirical value.

As already previously highlighted, for most of the rock types when considering high percentiles, CDF are higher than the ECDF (supporting the conclusion that the empirical data may not represent correctly the extremes i.e., the largest volumes in the catalogues. In these cases, modelled CDF should be preferred to ECDF based estimations.

The sensitivity of the CDFs towards variations of minimum and maximum mapped boulder sizes was tested by (i) removing the sizes considering different thresholds below the roll-over (small sizes) (Fig. 5A), and (ii) removing the sizes considering different thresholds above the roll-over (large sizes) (Fig. 5B). The results show that the CDF estimations in general have a low sensitivity to the removal (i.e. the exclusion when mapping) of small sizes, whereas the sensitivity to the removal of large sizes is high.

5. Discussion and conclusions

We analysed fifteen catalogues listing information on rockfall size in Austria, Italy, and the USA with respect to the impact of mapping strategy i.e., choice of mapping method and data source. Results showed that specific factors have an impact on frequency-size distributions of the rockfalls. These factors should be considered when planning and conducting data collection campaigns i.e., the definition of a "boulder scenario" (or design block). The main results of this work can be summarized as follows:

Comparing different rock types, the variations of the Empirical Cumulative Distribution Functions (ECDFs) within and among the catalogues become evident for larger percentiles. This is important when using rockfall size percentiles for the definition of design blocks (i.e., ONR 24810). We recommend to use different percentiles (i.e., from 75th to 99th) to better portrait the statistical distribution of large size rockfalls for a more robust stastistical hazard zoning and for a more objective design of rockfall mitigation measures.

Catalogues C_{UM} and C_{BB} do not show a rollover. We maintain this is a result of the comparably large number of small rockfalls (C_{BB}). Catalogues $C_{H5,}$ $C_{H6,}$ C_{H7} and C_{BB2} show a large dispersion for small rockfalls (β-values). This reveals the role of the mapping strategy, which focused on the mapping of even very small boulders.

Comparing the position of the "rollover" for the datasets, it emerges that the most frequent boulder size values for most of the catalogues are between 0.04 and 0.07 m^3. Catalogues with very small rockfall sizes (< 0.002 m^3) do not show a distinct rollover in the frequency-size distribution. In such cases, the smallest is the most frequent boulder size, and then it should be used in place of "rollover" position for the definition of a block scenario.

Sensitivity analysis of the CDFs to variations of minimum and maximum rockfall sizes (i.e. which in a catalogue/inventory depend by the choice of the mapping strategy) reveals that the CDF estimations are slightly sensitive to the exclusion of small size boulders, whereas the sensitivity to the exclusion of large sizes is high.

The appropriateness of modelled rockfall size distributions was analysed using KS-Test and considering the variability of the estimated DPS distribution parameter (α, β, t), s and the associated *p*-value. For most rockfall catalogues, the KS-Test obtained considering the DPS distribution function was passed, and the resulting distribution parameter proved significant. This finding supports the use of the DPS distribution to model rockfall size here and elsewhere.

Mapping the return periods of rockfall boulders in the field require a large and representative dataset of rockfall sizes and an appropriate selection of the mapping strategy. For most rock types, considering the larger percentiles of the frequency size distributions, the modelled Cumulative Distribution Functions (CDF) values are higher than the ECDF values. We believe, that in these cases empirical data may not represent correctly large sizes due to mapping issues and modelled CDF should be preferred in place of ECDF based estimations, closing data gaps.

Recommendations for the definition of mapping strategies arising from these findings are:

- Selective mapping i.e., neglecting mapping of large (old) rockfalls should be avoided. Topographic factors, such as lakes and rivers on the bottom of the slope, or mapping of rockfalls in the middle of the slopes may lead to an underestimation of large rockfall sizes. The deposits of large rock mass falls/cliff falls should be analysed separately.
- Site-specific rockfall surveys for single structures, or subdivision of catalogues for specific analyses may result in a low number of

rockfalls, possibly with a reduced volume range. In such cases, the use of parametric distribution estimations using DPS functions, and after having verified that KS-test is passed, may help providing more robust volume statistics. In addition, the merging with other catalogues should be considered to ensure more robust size statistics estimation, leading to better boulder scenarios.

- Estimation of rockfall size in pre-defined size classes is not appropriate for probability density functions (PDF) and CDF models. Even if ECDF can cope with the data, biases due to the mapping and classification strategy exist.

As a final consideration, from a statistical perspective rockfall datasets, as other geological records, should be considered always incomplete. Incompleteness is not necessarily an issue if the datasets are representative. Indeed, appropriate inferential approaches, like non-parametric (i.e. ECDF-based) and parametric (i.e. using distribution models) approaches, may provide meaningful statistics even from incomplete but representative samples. Both inferential approaches present advantages and limitations. While ECDF provides estimates within the original sample size range, modelled CDFs are able to extend outside these bounds. Parametric models should be preferred because they are easier to use to derive distribution related statistic, but also because they provide an analytical framework helpful to analyse statistics differences for rockfall sizes outside the observed size sample range. Inferential statistical parametric and non-parametric methods such as included in the statistical tool by (Rossi, 2014) can be used and developed where missing, to be able to cope with small datasets and to potentially close data gaps. Procedures to account data accuracy and to evaluate the statistical significance and uncertainty of the results, need to be always applied to obtain reliable rock fall statistics, regardless the selected inferential approach. Under this view, information on the mapping method used to collect rockfall data, the type of source information, and references to the sources of information should be part of any rockfall database.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The Geological Survey of Austria is thanked for supporting S.M. during the paper completion also after leaving the Geological Survey. Special thanks to the National Park Service, CA, for the Yosemite rockfall dataset.

Appendix A. Mathematical background

In the following, mathematical formulations of the different distribution models used for parametric PDF and CDF estimations are provided. Greater details can be found in Rossi et al. (2014).

Double Pareto (five parameters)

$$pdf\left(\mathrm{x}|\ \alpha, \beta, \mathrm{t}, \mathrm{c}, \mathrm{m}\right) = \left[\frac{\beta}{t}\left(1-\left(\frac{1+\left(\frac{m}{t}\right)^{-\alpha}}{1+\left(\frac{c}{t}\right)^{-\alpha}}\right)^{\left(\frac{\beta}{\alpha}\right)}\right)\right]\left[\frac{\left(1+\left(\frac{m}{t}\right)^{-\alpha}\right)^{\left(\frac{\beta}{\alpha}\right)}}{\left(1+\left(\frac{x}{t}\right)^{-\alpha}\right)^{\left(1+\frac{\beta}{\alpha}\right)}}\left(\frac{x}{t}\right)^{(-\alpha-1)}\right]$$

Double Pareto (three parameters)

$$pdf\left(\mathrm{x}|\ \alpha, \beta, \mathrm{t}\right) = \left[\frac{\beta\left(\frac{\mathrm{t}}{\alpha}\right)}{\left(1+\left(\frac{\mathrm{x}}{\mathrm{t}}\right)^{-\alpha}\right)^{\left(1+\left(\frac{\beta}{\alpha}\right)\right)}\left(\mathrm{x}^{(\alpha+1)}\right)}\right]$$

Inverse Gamma (three parameters)

$$pdf\left(\mathrm{x}|\ \alpha, \eta, \lambda\right) = \left[\frac{\lambda^{2\alpha}}{\Gamma(\alpha)}\right]\left[\left(\frac{1}{\mathrm{x}+\eta^2}\right)^{(\alpha+1)}\right] exp\left[-\frac{\lambda^2}{x+\eta^2}\right]$$

References

ONR 24810:2017, 2017. Technical Protection Against Rockfall - Terms and Definitions, Effects of Actions, Design, Monitoring and Maintenance. (Technical Guide). Austrian Standards. Austrian Standards Institute, Vienna ONR 24810:2017-02-15.

Agliardi, F., Crosta, G.B., Frattini, P., 2009. Integrating rockfall risk assessment and countermeasure design by 3D modelling techniques. Nat. Hazard. Earth Syst. Sci. 9, 1059 Katlenburg-Lindau.

Barry, C.A., 1983. Pareto Distributions. International Co-operative Publishing House.

Barry, C.A., 2011. Chapter 7: Pareto and generalized Pareto distributions. In: Modeling Distributions and Lorenz Curves.

Brunetti, M.T., Guzzetti, F., Rossi, M., 2009. Probability distributions of landslide volumes. Nonlinear Process. Geophys. Gottingen 16, 179.

Bunce, C.M., Cruden, D.M., Morgenstern, N.R., 1997. Assessment of the hazard from rock fall on a highway. Can. Geotech. J. 34 (3), 344–356.

Corominas, J., Mavrouli, O., Ruiz-Carulla, R., 2017a. Magnitude and frequency relations: are there geological constraints to the rockfall size? Landslides 1–17. https://doi.org/10.1007/s10346-017-0910-z.

Corominas, J., Mavrouli, O., Ruiz-Carulla, R., 2017b. Rockfall occurrence and fragmentation. In: Advancing Culture of Living with Landslides. Springer, Cham, pp. 75–97. https://doi.org/10.1007/978-3-319-59469-9_4. Presented at the Workshop on World Landslide Forum.

Crosta, G.B., Agliardi, F., Frattini, P., Lari, S., 2015. Key issues in rock fall modeling, hazard and risk assessment for rockfall protection. In: Engineering Geology for Society and Territory - Volume 2. Springer, Cham, pp. 43–58. https://doi.org/10.1007/978-3-319-09057-3_4.

De Biagi, V., Napoli, M.L., Barbero, M., 2017a. A quantitative approach for the evaluation of rockfall risk on buildings. Nat. Hazards J. Int. Soc. Prev. Mitig. Nat. Hazards 88, 1059–1086.

De Biagi, V., Napoli, M.L., Barbero, M., Peila, D., 2017b. Estimation of the return period of rockfall blocks according to their size. Nat. Hazards Earth Syst. Sci. 17, 103–113. https://doi.org/10.5194/nhess-17-103-2017. Katlenburg-Lindau.

Dussauge, C., Grasso, J.-R., Helmstetter, A., 2003. Statistical analysis of rockfall volume distributions: implications for rockfall dynamics. J. Geophys. Res. Solid Earth 108, 2286. https://doi.org/10.1029/2001JB000650.

Dussauge-Peisser, C., Helmstetter, A., Grasso, J.-R., Hantz, D., Desvarreux, P., Jeannin, M., Giraud, A., 2002. Probabilistic approach to rock fall hazard assessment: potential of historical data analysis. Nat. Hazards Earth Syst. Sci. 2, 15–26.

Guzzetti, F., Cardinali, M., Reichenbach, P., 1994. The AVI project: a bibliographical and archive inventory of landslides and floods in Italy. Environ. Manag. 18, 623–633. https://doi.org/10.1007/BF02400865.

Guzzetti, F., Reichenbach, P., Wieczorek, G.F., 2003. Rockfall hazard and risk assessment in the Yosemite valley, California, USA. Nat. Hazards Earth Syst. Sci. 3 (6), 491–503.
Guzzetti, F., Mondini, A., Cardinali, M., Fiorucci, F., Santangelo, M., Chang, K., 2012. Landslide inventory maps: new tools for an old problem. Earth Sci. Rev. 112, 42–66.
Hungr, O., 1997. Some methods of landslide hazard intensity mapping. Landslide Risk Assess 215–226.
Lambert, S., Bourrier, F., 2013. Design of rockfall protection embankments: a review. Eng. Geol. 154, 77–88. https://doi.org/10.1016/j.enggeo.2012.12.012.
Lari, S., Frattini, P., Crosta, G.B., 2014. A probabilistic approach for landslide hazard analysis. Eng. Geol. 182, 3–14. https://doi.org/10.1016/j.enggeo.2014.07.015. Advances in Engineering Geology for Landslides and Slope Stability Problems: Part 1.
Macciotta, R., 2014. A Review of The Uncertainties in Rockfall Protection Design. https://doi.org/10.13140/2.1.3065.5684.
Malamud, B.D., Turcotte, D.L., Guzzetti, F., Reichenbach, P., 2004. Landslide inventories and their statistical properties. Earth Surf. Process. Landf. 29, 687–711. https://doi.org/10.1002/esp.1064.
Rossi, M., 2014. R Code for Prototype Software for Landslide Statistics (LSTATS). Deliverable of Lampre Project (Deliverable of Lampre Project).
Rossi, M., Witt, A., Guzzetti, F., Malamud, B.D., Peruccacci, S., 2010. Analysis of historical landslide time series in the Emilia-Romagna region, northern Italy. Earth Surf. Process. Landf. 35, 1123–1137.
Rossi, M., Cardinali, M., Fiorucci, F., Marchesini, I., Mondini, A.C., Santangelo, M., Ghosh, S., Riguer, D.E.L., Lahousse, T., Chang, K.T., Guzzetti, F., 2012. A Tool for the Estimation of the Distribution of Landslide Area in R. Presented at the EGU. pp. 9438.