Received: 10 February 2015 | Revised: 10 May 2017 | Accepted: 28 November 2017

DOI: 10.1111/jfr3.12333

CIWEM Chartered Institution of Water and Environmental Management | WILEY Journal of Flood Risk Management

ORIGINAL ARTICLE

Quantifying the damage susceptibility to extreme events of mountain stream check dams using Rough Set Analysis

B. Mazzorana[1,2,3] | H. Trenkwalder-Platzer[4] | M. Heiser[4] | J. Hübl[4]

[1]Faculty of Science, Instituto de Ciencias de la Tierra, Universidad Austral de Chile, Valdivia, Chile

[2]RINA—Núcleo de Investigación en Riesgos Naturales y Antropogénicos, Universidad Austral de Chile, Valdivia, Chile

[3]Millennium Nucleus CYCLO, The Seismic Cycle Along Subduction Zones, Valdivia, Chile

[4]Institute of Mountain Risk Engineering, University of Natural Resources and Life Sciences, Vienna, Austria

Correspondence
Bruno Mazzorana, Facultad de Ciencias, Instituto de Ciencias de la Tierra, Universidad Austral de Chile, Valdivia, Chile.
Email: mazzoranabr1@gmail.com

Funding information
Iniciativa Científica Milenio (ICM), Grant/Award number: NC160025

In this work, a Rough Set Analysis-based approach is proposed to quantify the damage susceptibility of check dams through specific indexes, which all require expert judgment to be quantified. The indexes are the Post-Event Damage Condition (*DamPost*), the induced Condition Change (*DamCh*), the Residual Condition (*RC*), and the Post-Event Functionality (*FPost*). Preliminarily, an existing data set, containing a quantification of the damage indexes, the associated characterisation of the flow process type of three torrential hazard events occurred in South Tyrol (Italy), the identification of construction material and the determination of age of the structure, was statistically analysed. To predict the damage indexes based on Rough Set Analysis, a general model, which considered all check dams regardless of their construction material, two specific models, for concrete and for masonry structures respectively, and a simplified version of the general model were set up. The derived rule bases exhibited satisfactory prediction accuracies only when the post-event functionality, *FPost*, was chosen. Prediction accuracies were 68% for the general model, 79% for the material category concrete, 60% for the material category masonry, and 86% could be obtained by simplifying the decision attribute to a binary form (functionality given or not).

KEYWORDS

flood damage Indexes, flood defence structures, integrated flood risk management, life-cycle management

1 | INTRODUCTION

Preventing the release of sediment from their sources in mountain catchments, interfering with the dynamics of sediment transport by stabilising the streambeds through the realisation of grade control structures and retaining solid material volumes transported during extreme events are widespread strategies to reduce risks in mountain areas (Bergmeister, Suda, Hübl, & Rudolf-Miklau, 2009).

On the contrary, it is ascertained that, without a release of sediments either from their sources or from their intermediate deposits and without maintaining sediment connectivity throughout the stream network, the reactivation of hydro-morphological and the associated ecological functionalities are physically unfeasible for supply-limited and highly altered mountain rivers (Rinaldi, Surian, Comiti, & Bussettini, 2011). In parallel, on several debris cones and alluvial fans a clear increasing tendency of wealth moving into flood prone areas could be retraced over the last decades, leading to a possible net exacerbation of risk (Fuchs, Keiler, Sokratov, & Shnyparkov, 2013; Mazzorana, Simoni, et al., 2014). Without a profound revision of land use management and without significantly reducing the vulnerability of the built environment, the persistence of both functional and reliable check dam structures is of highest priority (Suda, 2012). As outlined by Dell'Agnese, Mazzorana, Comiti, Von Maravic, and D'Agostino (2013), the determination of check dams damage susceptibility is an

essential requirement for the definition of adequate maintenance strategies. With the intent to promote an enhanced life cycle management of check dams, this need was recently recognised on an operative level also by the Platform on Natural Hazards of the Alpine Convention (Rimböck et al., 2014). With respect to natural hazard risk management, Mazzorana, Trenkwalder-Platzer, Fuchs, and Hübl (2014) observe that the aim to reduce potential hazards by consolidating the stream beds, and in particular by using grading structures that artificially may retain amounts of sediment, has to be judged carefully. Due to a limited technical lifetime of any constructive mitigation, in combination with the possibility of technical failure (residual hazard), the natural disposition factors gradually change. This gradual change, however, does not seem to be acknowledged by the local actors, for example, the population affected. On the contrary, land-use in the run-out areas of hazardous processes increased since the 1950s, and depending on the respective national and regional building laws, a considerable amount of value was concentrated in endangered areas (Fuchs et al., 2013). The failure to prevent completely damages by natural hazards generated a higher demand for protection in those areas heavily developed in recent decades. Therefore, starting from the 1960s, the respective agencies responsible for the protection against natural hazards continued to pursue the consolidation strategy throughout the European Alps by constructing new grade control structures, prevailingly as masonry works in a first stage and then progressively as concrete structures. To give an example, approximately 30,000 check dams have been constructed in South Tyrol, Italy, since 1900, and 16% of them were judged not to satisfy the required reliability and, consequently, technical efficiency requirements (Mazzorana, 2008). Due to these inherent deficiencies of pure consolidation strategies, a large number of open, filtering check dams has been constructed since the early 1970s. The functional efficiency of this type of structure was gradually refined (Üblagger, 1972) firstly by improving the mechanical sieving function and subsequently by modifying the design to obtain a cost-efficient dosing function (Armanini & Larcher, 2000). In many cases, however, the design of such systems was inherently weak due to (a) erroneous assumptions of full performance of the previously constructed consolidation structures and (b) procedural and content-related gaps in the adopted planning procedures (Mazzorana & Fuchs, 2010). From a reactive perspective, capillary monitoring activities have been carried out to both ascertain the condition and functionality of the realised constructions. At the beginning of the 21st century, large, but more or less systematic and homogeneous, check dam condition survey campaigns were conducted in many alpine regions (Suda, 2012). For hazard mapping and risk management purposes, it is essential to recognise the following three basic or fundamental characteristics of protection structures:

1. Structures forming the protection systems are of a dual nature because they are designed to mitigate natural hazards but on the other hand they are prone to be damaged throughout their lifecycle by the same processes they should mitigate (Vorogushyn, Merz, & Apel, 2009), thus reducing their performance over time.
2. Sudden unexpected collapse of check dams can result in increased hazards downstream due to the formation of dam-break surges and the release of large volumes of sediments.
3. Physical susceptibilities of single check dams contribute to the susceptibility of the entire risk mitigation system, but the latter is not equal to the sum of the susceptibilities of the single structures. In fact, because of the interrelationships between structural damage and event intensity, non-linear dynamics emerge making it difficult to predict the final state.

However, the need to develop predictive models to quantify the damage susceptibility of mountain stream check dams for extreme events has only recently be fully recognised (Mazzorana & Fuchs, 2010). Dell'Agnese et al. (2013) made extensive use of statistical methods to determine a damage index defined on pre- and post-event comparisons of check dam conditions and relevant impact variables. As a result of their study they proposed a vulnerability matrix for consolidation check dams. This matrix describes the average expected values for residual functionality (RF) of check dams as a function of structure characteristics (taking into consideration also the initial RF values) and event intensity and type. Event intensity is expressed as type of event, discharge, local energy slope, unit stream power, sediment size, flow width, and depth. The matrix is meant to represent only a preliminary tool to estimate the physical vulnerability of check dams, and, as such, it is intended as a starting point to plan the preventive maintenance of check dams. With the overall aim to enhance the detection of the damage generating mechanisms and to improve pro-active check dam maintenance strategies, a series of prediction models based on Rough Set Analysis techniques (Munakata, 2008; Rutkowski, 2008) are set up. In its essence, Rough Set Analysis is flexibly used to generate rule bases, which establish a relation (i.e., in form of if-then implications) between attributes of the system and a selected objective variable. To this end, as described in the next sections, an existing data set, which contains quantified process-response information with respect to selected check dam structures built in South Tyrol, Italy, is first analysed. In a successive step, as outlined in Section 3 below, the damage susceptibility prediction models based on Rough Set Analysis techniques are described and the obtained results are presented.

TABLE 1 Overview of the occurred events, the selected creeks, the dominant processes and the associated number of selected and investigated check dams

Creek	Date of the event	Process	Number of check dams
Hölderle creek	August 1, 2013	Debris flow	28
Holer creek	August 6, 2008 and August 21, 2008	Debris flow	77
Höllental creek	August 6, 2008 and August 21, 2008	Debris flow	18
Keltal creek	July 17, 2009	Debris flow	22
Rethen creek	July 29, 2009 and September 24, 2009	Debris flow	24
Gadria-Allitzer creek	July 24, 2009	Debris flow	14
Tanz creek	September 4, 2009	Intense bedload transport—debris flood	10
Tinne creek	September 4, 2009	Intense bedload transport	13
Ziel creek	August 6, 2008 and August 21, 2008	Debris flow	25
			231

2 | DATA ANALYSIS

2.1 | Data set, data structure, and damage indexes

A data set originally complied by Von Maravic (2010) and Dell'Agnese et al. (2013) containing quantified process–response information with respect to selected check dam structures, is the basis for both statistical data analysis and knowledge generation by means of rough set analysis techniques (compare next section). The data set contains information about (a) torrential hazard events occurred in 2008, 2009, and 2013 in South Tyrol, Italy and (b) knowledge about check dam structures and responses to process events impacts.

Table 1 presents an overview of the events considered along selected creeks, their dates of occurrence, the dominant process, the municipality were the event took place, and the number of selected and investigated check dams. Figure 1 provides an associated geographical overview of the selected creeks, the district names and the associated limits.

With reference to the sketch of the functional parts of a check dam as shown in Figure 2, the structure of the data set including all considered variables is shown in Table 2 (i.e., factorial and numeric variables).

To provide an overview about the value range of the main geometrical characteristics of the considered check dams (see Figure 2), the construction height $-h-$ varies from 1 to 9.5 m, whereas the construction width comprising the spillway and unanchored part of the wings $-b-$ ranges from 5.8 to 28.5 m. Since the main function of the surveyed check dams is consolidating the streambed and both the distance between two successive structures (10–100 m) and the difference between channel slope and equilibrium slope (0–5%) are relatively small, their retention volume is practically negligible. The estimated event-related peak discharges range from 80 to 200 m^3/s (Trenkwalder-Platzer, 2014).

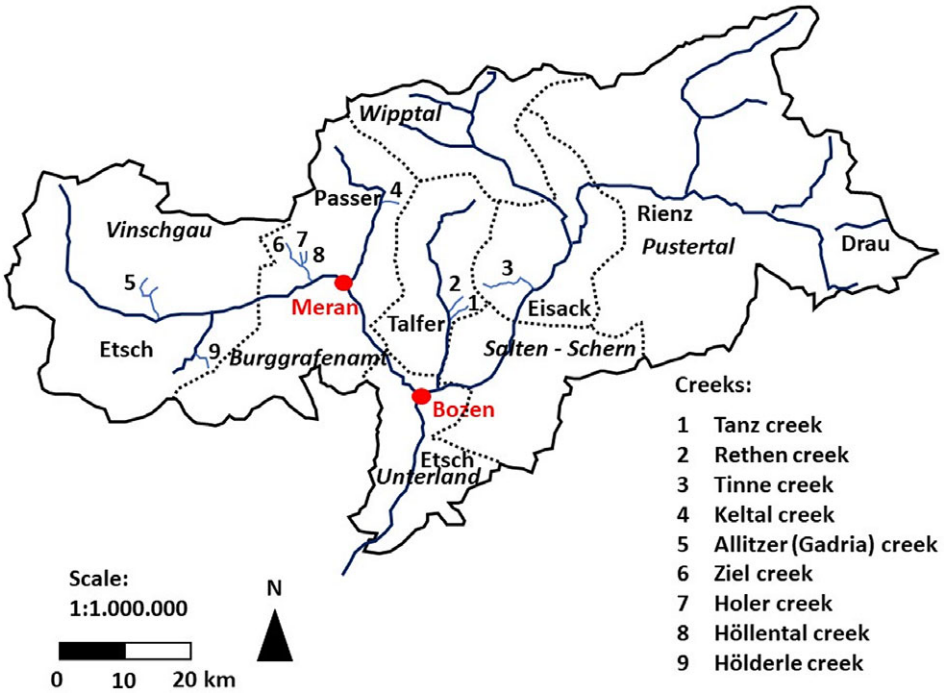

FIGURE 1 Geographical overview of the selected creeks in South Tyrol, Italy

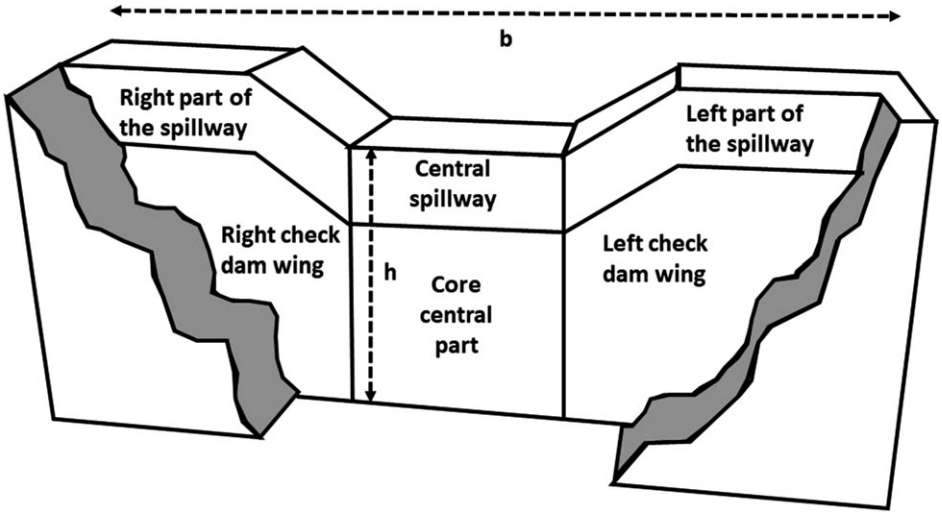

FIGURE 2 Subdivision of the check dam into functional parts, considered geometrical parameters

It has to be remarked that the set of damage indexes reported in Table 2 is not exhaustive. Comiti, Lenzi, and Mao (2010) extensively investigated the local scouring which may exceed the depth of the check dam foundations and decrease the global stability of the structure and Suda (2012) highlights critical seepage and filtration effects. The available data set did not allow a detailed event based analysis of these damage generation mechanisms, and their qualitative assessment should be mandatory.

The computation of the damage indexes is based, respectively, on the assessment of the vectors *PRE* and *POST*, which express the condition of the check dam in a pre- and post-event situation, respectively. These vectors contain as elements the expert based assessments of each part of the check dam on a factor–point scale (range: 1–5, see Table 2).

$$\begin{aligned} PRE &= (fl.pre, kl.pre, ml.pre, km.pre, kr.pre, fr.pre) \\ &= (pre_1, pre_2, pre_3, pre_4, pre_5, pre_6) = (pre_i;\ i = 1,\ldots,6) \end{aligned} \tag{1}$$

$$\begin{aligned} POST &= (fl.post, kl.post, ml.post, km.post, kr.post, fr.post) \\ &= (post_1, post_2, post_3, post_4, post_5, post_6) \\ &= (post_i;\ i = 1,\ldots,6) \end{aligned} \tag{2}$$

As outlined in Dell'Agnese et al. (2013) the vector of weights, whose elements reflect the relative importance for structural stability of the different parts of the check dam, is

$$\begin{aligned} W &= (w.fl, w.kl, w.ml, w.km, w.kr, w.fr) \\ &= (w_1, w_2, w_3, w_4, w_5, w_6) = (w_i; i = 1,\ldots,6) \end{aligned} \tag{3}$$

Based on a previous assessment of the vectors *PRE*, *POST* and *W*, the damage indexes expressing the check dam's pre- and post-event condition *DamPre* and *DamPost*, respectively, can be calculated as follows:

$$DamPre = \frac{\sum_i [w_i \cdot (pre_i)] - 1}{4} \cdot 100 \tag{4}$$

$$DamPost = \frac{\sum_i [w_i \cdot (post_i)] - 1}{4} \cdot 100 \tag{5}$$

The damage index change through the event—*DamCh*—and the residual condition—*RC*—are then evaluated through the following expressions:

$$DamCh = DamPost - DamPre \tag{6}$$

$$RC = 100 - DamPost \tag{7}$$

The introduction of these damage indexes is useful to monitor the evolution of the check dam condition throughout its life cycle as exemplified in Figure 3.

2.2 | Descriptive statistics

In Figure 4 the distribution with respect to the building material and to the construction age class is shown. The number of construction age classes was reduced and their upper and lower limits adjusted with respect to Dell'Agnese et al. (2013) to avoid classes with a relatively low number of structures and limit the complexity of the developed prediction models (compare Section 3 for details). Whereas the creeks Ziel and Rethen feature exclusively check dams built of concrete, the check dams in the Höllental creek and Hölderle creek are made of masonry. In the Allitzer (Gadria) creek and Holer creek an approximately equal distribution of concrete and masonry check dams is to be found. In the Tinne and Tanz creeks masonry structures prevail, whereas in the Keltal creek the opposite is true.

In the bar charts shown in Figure 5 the average values of the Damage Indexes *DamPre*, *DamCh*, and *DamPost* are reported separately for the different creeks (Section A), for the different structural parts of the check dam considering the entire data set (Section B) and for both construction material categories concrete and masonry (Sections C and D, respectively). The response, captured by the damage index *DamCh*, was particularly significant for the check dams

TABLE 2 Variables of the data set, typological description, categorisation, and value range

Variable name	Description	Type	Categories and range
Creek	Name of the creek	Factor: info available from the stream network	No subdivisions
Event	Identification number of the event	Factor: info available from the event database	No subdivisions
Condition before the event (xx.pre)			
fl.pre	Left wing of the check dam	Assessment on a factor—point scale (range: 1–5) based on photo interpretation from the structure database (Baukat30)	1: No visible damages; 2: Elements slightly damaged at the surface; 3: Parts with gaps, fissures or fractures; 4: larger portions severely damaged; 5: Structural integrity completely missing
kl.pre	Left part of the spillway		
ml.pre	Central part of the check dam		
km.pre	Central part of the spillway		
fr.pre	Right wing of check dam		
kr.pre	Right part of the spillway		
Condition after the event (xx.post)			
fl.post	Left wing of the check dam	Assessment on a factor—point scale (range: 1–5) based on photo interpretation from the structure database (Baukat30)	1: No visible damages; 2: Elements slightly damaged at the surface; 3: Parts with gaps, fissures or fractures; 4: larger portions severely damaged; 5: Structural integrity completely missing
kl.post	Left part of the spillway		
ml.post	Central part of the check dam		
km.post	Central part of the spillway		
fr.post	Right wing of the check dam		
kr.post	Right part of the spillway		
Flowtype	Dominant process type: Debris Flow, Debris Flood (Hyperconcentrated flow) or Bedload transport	Factor: Assessed from the event database (ED30)	1: Bedload transport; 2: Debris flood (Hyperconcentrated flow); 3: Debris Flow
h_0	Flow depth (m)	Numeric value measured during post-event documentation based on silent witnesses (e.g., water marks) debris flow levees etc.)	Continuous range $h_0 > 0$
w_0	Flow width at the free surface (m)		Continuous range $w_0 > 0$
Material	Construction material typology (concrete, masonry)	Factor: Deduced by photo interpretation from the structure database (Baukat30) and by post event documentation surveys	1: Concrete; 2: Masonry
History	Info on occurred damaging mechanism		1: First damaging event; 2: Pre-existing damages
Age	Age of the structure (years)	Construction age of the structure	1: <15; 2: 15–30; 3: 31–50; 4: >50
Geometry	Height of the check dam (m)	Numeric value: retrieved from the structure database (Baukat30) and during post event documentation surveys	Continuous range > 0
Inclination	Slope of the stream at the check dam location	Dimensionless numeric value: deduced from the digital terrain model - DTM - (resolution: 2.5 m); measured during post event with a laser distance meter	
Diameter	Estimated D90 at the check dam location	Numeric value: assessed during post event documentation surveys	
DamPre	Damage index before the event	Computed numeric value (%)	Continuous range: 0–100%
DamCh	Damage index change through the event		
DamPost	Damage index after the event		
RC	Residual condition		

(construction material: masonry prevailingly) in the Tanz creek, similarly for those in the Hölderle creek (construction material: masonry exclusively). In both cases, *DamPre* was relatively large. In the Rethen creek, *DamPre* was large as well, but, in comparison, *DamCh* was lower (construction material: concrete exclusively). In the Ziel, Tinne and Allitzer (Gadria) creeks, *DamPre* was rather low. These three creeks exhibited a similar response in terms of *DamCh*, despite different construction materials were used (concrete exclusively in the Ziel creek, prevailingly masonry in the Tinne creek and both construction materials in the Allitzer-Gadria creek).

Plot B in Figure 5 reveals, as intuitively expected, that the central parts of the check dam (i.e., mL and km, respectively) are more prone to be damaged in comparison to the lateral parts (i.e., kl, kr, fl, fr). This pattern emerges clearly for masonry check dams, whereas for check dams built in concrete it is less pronounced. In plot C, the empirical cumulative probability density distributions of all damage indexes are shown separately for concrete and masonry check dams. For the latter case, a comparably large damage increase—*DamCh*—could be detected, not only for check dams with large *DamPre* index values but also for a significant percentage of previously undamaged structures (low

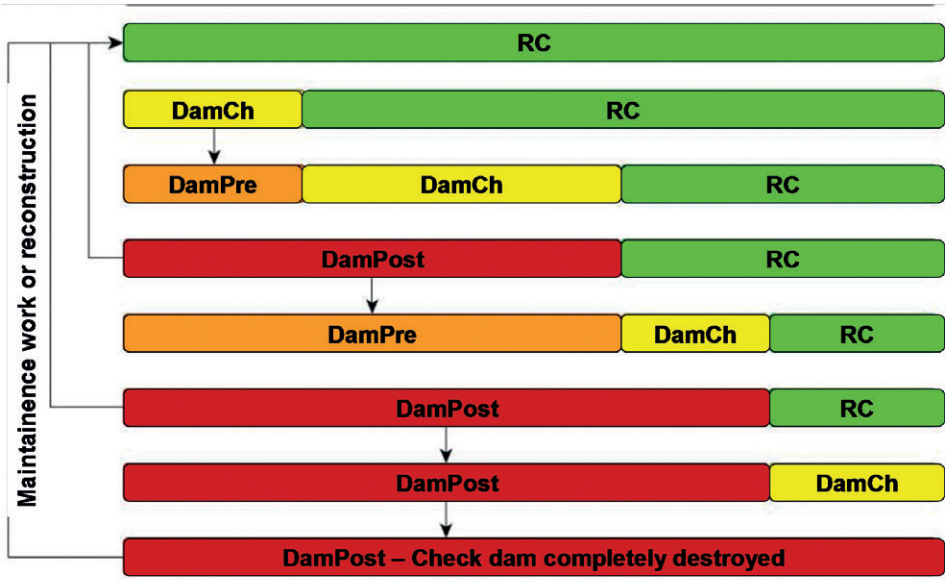

FIGURE 3 Representation of the evolution of the check dam condition using the defined damage indexes. *RC*, residual condition; *DamCh*, damage index change; *DamPre*, damage index value before the event; *DamPost*, damage index value after the event

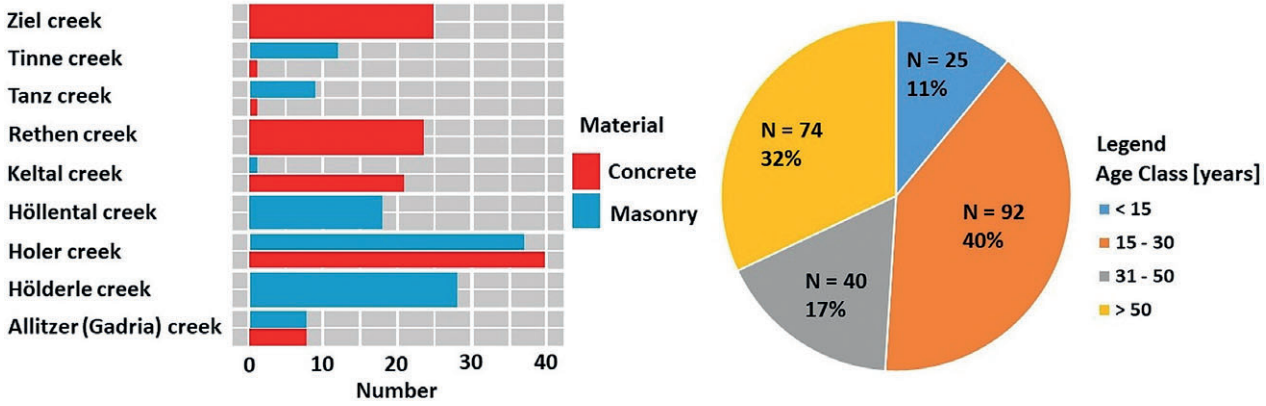

FIGURE 4 Left: Distribution of the number of check dams for each creek belonging to the different construction material categories (concrete, masonry); Right: Distribution of the check dams with respect to their age class

DamPre index values). This fact is appreciable taking note of the remarkable "shift" to the right of the *DamPost* cumulative probability distribution in the masonry check dam case. The same information content is represented through box plots in plot D.

In Figure 6 one can appreciate that for masonry check dams, which belong prevailingly to older construction age classes, the scatter of both *DamCh* and *DamPost* increases significantly compared to concrete check dams featuring a pre-existing damage state. The functional performance of a consolidation check dam related to its capacity to stabilise the stream bed and the adjacent hillslopes is judged through a proper functionality attribute, *FPost*. The functionality classification is as follows: class 1 (the Post-Event Functionality of the structure is unaltered), class 2 (the Post-Event Functionality of the structure is slightly reduced), class 3 (the Post-Event Functionality of the structure is significantly reduced), and class 4 (no residual Post-Event Functionality).

In Figure 7, the relationship between *FPost* and the average *DamPost* values is shown for each functionality class for the whole check dam set and for the check dam sets according to their construction type.

Multivariate linear regression models were applied for predicting the values of the adopted damage index set from various collection of predictor variable values (Johnson & Wichern, 2002). The maximum likelihood of the mean square error arising from the prediction of the values of the damage indexes was unacceptably large for all models applied. Therefore, the capabilities of a rule induction approach based on the rough sets theory were explored (compare Section 3).

3 | DAMAGE SUSCEPTIBILITY PREDICTION THROUGH ROUGH SET ANALYSIS

3.1 | Theoretical background

Rough Set Analysis is a computational intelligence technique, which has been recently developed to mine complex data sets featuring quantitative and qualitative attributes and to

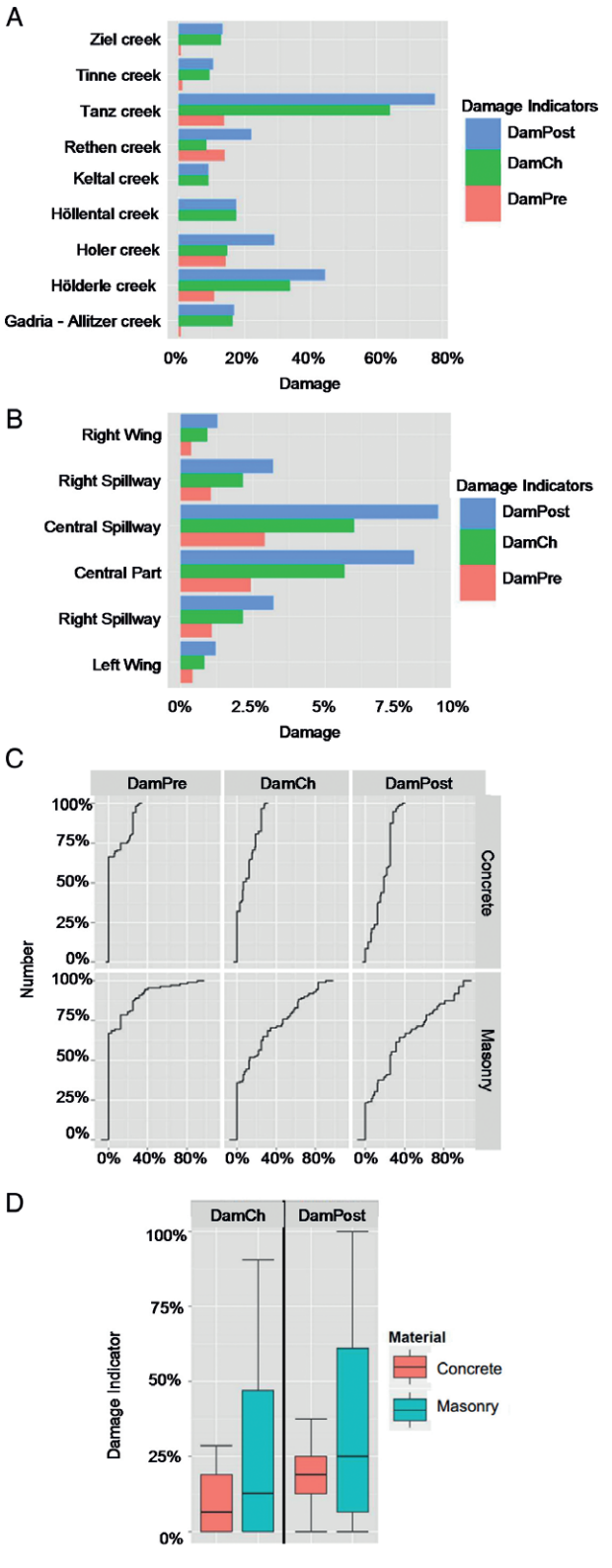

FIGURE 5 Calculated average values for the damage indexes *DamPre*, *DamCh* and *DamPost* for the events along the selected creeks (A), for the different check dam parts without discerning the construction material (B). Plot C: Empirical cumulative probability density distributions for all damage indexes for the considered construction materials, concrete and masonry, respectively. Plot D: Box plots for all damage indices for the considered construction materials

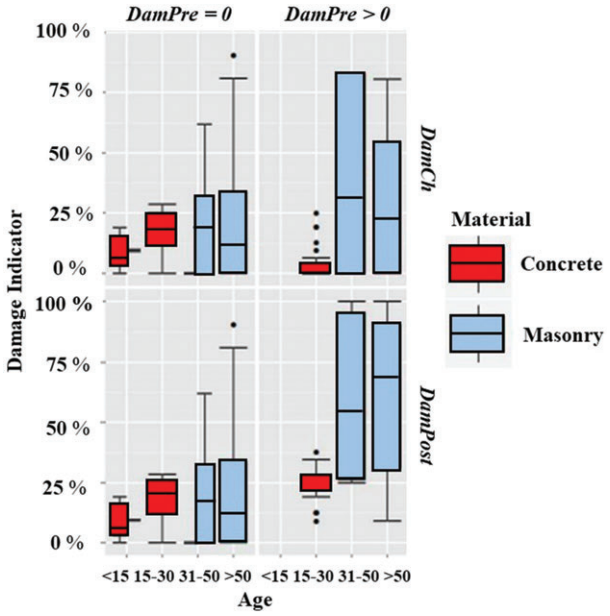

FIGURE 6 Box plots of the damage indexes *DamCh* and *DamPost* depending on the pre-existing damage condition (*DamPre* > 0 or *DamPre* = 0), differentiating between different construction materials and construction age classes

overcome the limited applicability of statistical methods to such data sets (Rutkowski, 2008). It was initially proposed in the 1980s by Zszislaw Pawlak (Pawlak, 1997). Since then the range of applications and software implementations in the technical, economic, and natural science domain, has been significantly extended (Munakata, 2008). In its essence, Rough Set Analysis is used to generate rule bases, which establish a relation (i.e., in form of IF-THEN implications) between attributes of the system and a selected objective variable (Olson & Delen, 2008). Non categorical data need to be discretised in classes by a manual way or by ad hoc discretisation-algorithms. Typically, data processed by Rough Set Analysis is organised in form of an information system

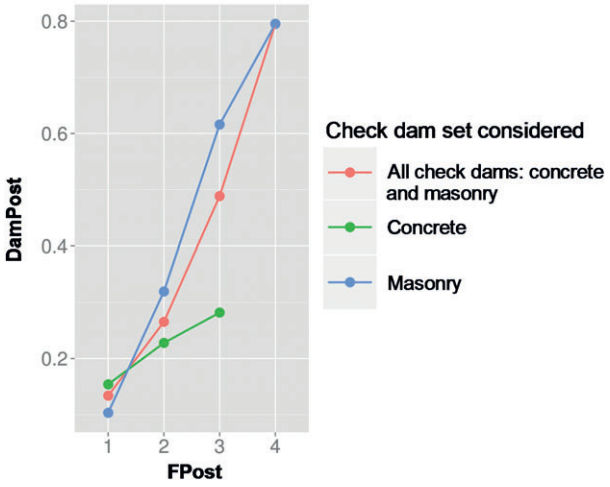

FIGURE 7 Relationship between *FPost* classes and the corresponding average *DamPost* values for the whole check dams set (red), for the check dams built of concrete (green) and masonry (blue)

$K = (U, C, D, V, \rho)$ where U indicates the entire data set, represented in form of a decision table, which links condition attributes C to the decision attributes D. V is the set of possible levels the key attributes $C \cup D$ can assume, and ρ is the information function defined as $\rho : CxD \rightarrow V$. The information function defines unambiguously the set of rules encoded in the given information system. The elimination of redundant information to provide more compact rules is achieved by identifying reducts, or subsets of key variables that still manage to preserve all the information within the decision table K. As outlined by Skowron and Rauszer (1992) the task of finding all reducts is an NP-complete problem. The search for reducts has been significantly enhanced by developing, for example, greedy algorithms (Johnson, 1974) and genetic algorithms (Düntsch & Gediga, 2000; Vinterbo & Øhrn, 2000). The identified reducts provide the basis for the determination of the rules of the rule base. Figure 8 visualises the core concepts of Rough Set Theory.

In a sharp definition of a set, every rule corresponds exactly to one outcome. For this reason, the boundary region equals 0 and all the rules are explicitly defined. On the contrary, a Rough Set describes a boundary region where a certain rule can have different meanings. In this case, the outcome of the rule can only be approximated by a lower and an upper approximation, quantified by the confidence factor alpha. The sum of uncertain rules describing X is visualised by the grey area in Figure 8. The blue area, instead, describes the number of rules with an exact outcome.

3.2 | Rule base generation

3.2.1 | Methodological approach

For determining the rules in the analysis of the present data set, the software ROSETTA (Rough Set Toolkit for Analysis of Data) was used. ROSETTA was developed in the period 1996–1998 at the Technical University of Trondheim on the kernel of RSES (Technical University of Warsaw). ROSETTA allows a spreadsheet based input of the data and a dynamic selection of the decision and the condition attributes. Continuous data is discretised applying one of the various algorithms available. With respect to the general structure of an information system $K = (U, C, D, V, \rho)$ and with the aim of conducting a Rough Set Analysis with the software ROSETTA, the pre-selected condition attributes C are shown in Table 3.

In this paper the rule base generation by Rough Set Analysis is presented, considering the post-event functionality, *FPost*, as decision variable. Other decision variable choices (i.e., *DamPost*, *DamCh*, *RC*) did not result into models and, hence, rule bases, featuring acceptable accuracies (Trenkwalder-Platzer, 2014). The attributes *FPost* with the classes 1, 2, 3, and 4 (compare Section 2), as well as a binary variable (i.e., YES, NO) expressing the post-event functionality fulfilment (or its absence), were chosen as decision attribute set D. Rearranging the whole data set according to the selected condition and decision attributes (i.e., the sets C and D) three distinct models reflecting particular condition attribute structures and featuring *FPost* as well as an additional simplified model were considered for Rough Set Analysis: (a) an overall general model containing the whole data set regardless of the material type, (b) a specific model for check dam structures built of concrete obtained by partitioning the whole data set according to this particular level of the condition attribute material, (c) a model for check dam structures built of masonry obtained by the corresponding data partition, and (d) a model containing the whole data set with the binary functionality decision attribute instead of *FPost*. For validation purposes, the

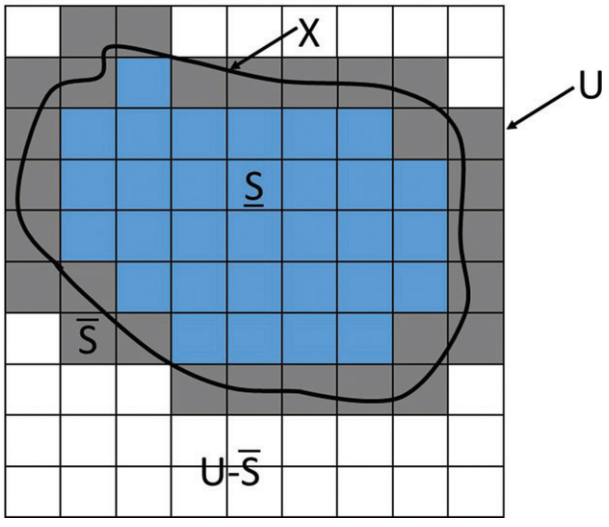

FIGURE 8 The set X can only be approximated by pixel sets. The following regions can be discerned: Universe U, Rough Set X, Lower Approximation S, Upper Approximation $\overline{S}$, Negative Region $NEG(X) = U - \overline{S}(X)$

TABLE 3 Attribute structure (condition attributes C) for Rough Set Data Analysis

Attribute	Description	Scaling
Flowtype (m)	Flowtype reflecting the flow process (i.e., bedload transport, debris flood and debris flow according to post-event documentation evidences)	Nominal
h_0 (m)	Flow depth (i.e., measured by a laser beam device according to flow marks measured in the field)	Discretised
w_0 (m)	Flow width (same measurement as for flow depths)	Discretized
Material	Type of construction material (i.e., concrete, masonry)	Nominal
Age	Year of construction	Ordinal
History	History of the structure (i.e., past damaging events)	Ordinal
Geometry (m)	Height of the structure	Discretised
Inclination	Channel slope at structure's location	Discretised
Diameter (m)	Maximum detectable grain size diameter in the reach containing the structure	Discretised
DamPre (%)	Estimation of the pre-damaged structural parts	Discretised

data set was divided into a training-set consisting of 75% of the data and a complementary test-set with 25% of the data. The resulting rule bases (compare next subsection) are derived based on the algorithmically identified reducts based on the underlying training-sets. The predictive performance of generated rule bases is then validated against the complementary test-set. The goodness of classification is evaluated through a cross-validation according to an accuracy parameter on a 0–1 scale, that is, the rule base generated with 75% of the K is used to predict the outcome D of the remaining test data (25% of K). The accuracy of the model is defined as a percentage based on the ratio between the number of correct implications $(CxD \rightarrow V)_{corr}$ and the total number of implications$(CxD \rightarrow V)_{tot}$. The validation procedure entails repeating this process for 16 times, always using different cuts on the data set. This entails, first, a partition of the data set in four parts (e.g., A, B, C, and D). Each of these parts is used once as a test-set and three times as one of three of the training set. In a next step, the partition point of the data in four parts is randomly changed and the same algorithm is repeated. Applying this procedure 4 times leads to a total number of 16 evaluations on 4 different cuts on the entire data set and, hence, considering an overall number of 4 different test data sets (25% of K, compare Shuib, Bakar, & Othman, 2009). It has to be remarked that, in spite of the significant efforts put maximising the prediction accuracies, they may still partially depend on alternative selections (or proportions) of the training data.

3.2.2 | Results

In Tables 4–7 respectively, excerpts of the entire rule bases are shown, corresponding to the set of previously outlined models in form of 10 significant rules in decreasing order with respect to the specified goodness criterion. The structure of the Tables is as follows: column 1—rule number, column 2—rules of the rule base ($CxD \rightarrow V$), column 3—LHS Support (i.e., number of objects in the training set, matching the Left Hand Side or IF-part of the rule), column 4—RHS Support (i.e., number of objects in the training set, matching the Right Hand Side or THEN-part of the rule), column 5—RHS Accuracy (i.e., ratio between the RHS Support and the LHS Support.), column 6—LHS Coverage (i.e., ratio between the LHS Support and the number of objects in the training set), column 7—RHS Coverage (i.e., ratio between the RHS Support and the number of objects in the test set, in other words the percentage expressing the extent of outcomes covered by the rule). In each of the reported Tables, Variable ([X,Y]) means that the value of the variable lies in the closed interval [X,Y]. The symbol * in the closed interval substituting X (or Y) indicates the smallest (largest) measured value of the variable with respect to the entire data set. Equivalently, it indicates the smallest (largest) value of the lowest (uppermost or highest) discretisation interval of the considered variable. The definitions and descriptions of all variables are provided in Table 2.

Satisfactory prediction accuracies could be obtained for all models. The prediction accuracy for classified post event functionality, *FPost*, was 68% for the general model, whereas for the material category concrete alone the accuracy raised to a value of 79% and for the material category masonry the value decreased to 60%. The reasons for the lower prediction accuracy of the model designed for the material category masonry in comparison to prediction accuracy obtained through the model designed for the material category concrete reside in less predictable damaging mechanisms. Whereas for the material category concrete the damage patterns tend to progress form the structural parts directly exposed to the flow process toward the central part of the

TABLE 4 Excerpt of the rule base of the general model comprising 10 rules with the largest coverage

Nr.	Rule	LHS Support	RHS Support	RHS Accuracy	LHS Coverage	RHS Coverage
1	h_0 ([*, 1.78]) AND Geometry ([3.85, 4.03]) AND Inclination ([0.35, 0.42]) AND Diameter ([0.87,1.39]) = > FPost (2) OR FPost (1)	3	1, 2	0.333, 0.667	0.01	0.022, 0.014
2	Flowtype (2) AND h_0 ([3.55, *]) AND Material (2) = > FPost (4)	6	6	1	0.03	0.29
3	h_0 ([3.23, 3.55]) AND History (2) AND Inclination ([0.14, 0.35]) = > FPost (3)	3	3	1	0.01	0.14
4	w_0 ([12.90, 22.50]) AND Inclination ([0.42, 0.6)] AND Diameter ([1.39, *]) = > FPost (4)	3	3	1	0.01	0.14
5	Flowtype (2) AND h_0 ([3.55, *]) AND w_0 ([22.50,*]) = > FPost (4)	3	3	1	0.01	0.14
6	h_0 ([*, 1.78]) AND Material (1) AND DamPre ([0.04,0.26]) = > FPost (1)	17	17	1	0.07	0.12
7	Flowtype (3) AND h_0 ([*, 1.78]) AND Inclination ([*,0.14]) = > FPost (1)	14	14	1	0.06	0.10
8	Flowtype (3) AND h_0 ([3.55, *]) AND DamPre ([0.04,0.26]) = > FPost (3)	2	2	1	0.01	0.10
9	Flowtype (2) AND Diameter ([0.64, 0.87])= > FPost (4)	2	2	1	0.01	0.10
10	h_0 ([2.28, 3.23]) AND Geometry ([2.85, 3.45]) AND Diameter ([*, 0.53]) = > FPost (3)	2	2	1	0.01	0.10

LHS = left hand side; RHS = right hand side.

TABLE 5 Excerpt of the rule base of the model (material type—concrete) comprising 10 rules with the largest coverage

Nr.	Rule	LHS support	RHS support	RHS accuracy	LHS coverage	RHS coverage
4	w_0 ([*, 14.75]) AND Geometry ([3.55, *]) AND Inclination ([0.3450, 0.4150]) AND Diameter ([0.8150,*]) = > FPost (2) OR FPost (1)	3	1, 2	0.333, 0.667	0.03	0.038, 0.023
5	h_0 ([2.45, 3.125]) AND w_0 ([14.75, *]) AND Geometry ([2.8500, 3.5500]) AND Inclination ([*,0.1450]) = > FPost (3)	2	2	1	0.02	0.25
18	History (1) AND Geometry ([2.85, 3.55)]) AND Inclination ([*,0.1450]) = > FPost (1)	15	15	1	0.13	0.18
25	h_0 ([2.45, 3.125]) AND w_0 ([*, 14.75]) = > FPost (1)	15	15	1	0.13	0.18
21	h_0 ([*, 2.35]) AND Geometry ([3.55, *]) AND DamPre ([0.0748, 0.2657]) = > FPost (1)	14	14	1	0.12	0.16
13	h_0 ([*, 2.35]) AND w_0 ([14.75, *]) AND Diameter ([*,0.815]) = > FPost (2)	4	4	1	0.03	0.15
17	h_0 ([*, 2.35]) AND DamPre ([0.2657, *]) = > FPost (2)	4	4	1	0.03	0.15
30	h_0 ([*, 2.35]) AND Geometry ([3.55, *]) AND Inclination ([*, 0.145]) = > FPost (1)	13	13	1	0.11	0.15
1	h_0 ([3.125, 3.60]) AND Age (1) = > FPost (1)	11	11	1	0.09	0.13
11	h_0 ([3.125, 3.60]) AND Geometry ([*, 2.85]) AND Inclination ([0.18, 0.345]) = > FPost (3)	1	1	1	0.01	0.13

LHS = left hand side; RHS = right hand side.

TABLE 6 Excerpt of the rule base of the model (material type—masonry) comprising 10 rules with the largest coverage

Nr.	Rule	LHS support	RHS support	RHS accuracy	LHS coverage	RHS coverage
4	h_0 ([*, 0.95]) AND Geometry ([4.90, *]) AND DamPre ([0.0164, *]) = > FPost (2) OR FPost (1)	4	2, 2	0.5, 0.5	0.04	0.111, 0.035
5	h_0 ([1.125, 2.225]) AND Age (4) AND Geometry ([4.90,*]) AND Inclination ([0.241, *]) = > FPost (2) OR FPost (4) OR FPost (1)	4	1, 1, 2	0.25, 0.25, 0,5	0.02	0.055, 0.047, 0.035
18	h_0 ([1.125, 2.225]) AND Inclination ([*, 0.241]) AND Diameter (0.78) = > FPost (2) OR FPost (1)	2	1, 1	0.5, 0.5	0.02	0.055, 0.017
25	h_0 ([*, 0.95]) AND Geometry ([4.90, *]) AND Inclination ([*, 0.241]) = > FPost (2) OR FPost (1)	2	1, 1	0.5, 0.5	0.02	0.055, 0.017
21	Geometry ([3.35, 4.90]) AND Diameter (0.78) => FPost (4) OR FPost (1)	2	1, 1	0.5, 0.5	0.02	0.047, 0.017
13	Flowtype (1) AND Age (3) AND Geometry ([2.10, 3.35]) = > FPost (1) OR FPost (2)	3	2, 1	0.667, 0.333	0.03	0.035, 0.055
14	h_0 ([2.225, 3.05]) AND Age (4) AND Diameter (0.78) = > FPost (1) OR FPost (4)	3	2, 1	0.667, 0.333	0.03	0.035, 0.045
30	Flowtype (1) AND Age (4) AND Geometry ([3.35,4.90]) = > FPost (1) OR FPost (2)	2	1, 1	0.5, 0.5	0.02	0.017, 0.055
1	Flowtype (2) AND h_0 ([3.05, *]) = > FPost (4)	8	8	1	0.07	0.38
11	Diameter (2.00) AND DamPre ([0.0164, *]) = > FPost (3)	3	3	1	0.03	0.23

LHS = left hand side; RHS = right hand side.

check dam, for the masonry case the less homogeneous damage patterns can be explained by a larger scatter of the material property values and by the presence of inbuilt weak shear planes within the stone structure which may fail under collisional loadings (Trenkwalder-Platzer, 2014). Moreover, masonry check dams built in staircase like fashion in steep debris flow channels may be locally subjected to significant impulsive horizontal earth pressures in the central parts of the check dam, which are caused by large boulder collisions in the short channel tracts between two adjacent structures (Valentini, 2012). An accuracy of 86% could be obtained by simplifying the decision attribute to a binary form, discerning only whether the post event functionality was given or not. This approach leads to high accuracy values, at the expense of a lower explanatory power of model, since the number of outcomes of the decision attribute *FPost* has been reduced to two outcomes only.

4 | DISCUSSION AND CONCLUSIONS

By means of descriptive statistics, in this work the relationships between observed characteristics of three torrential hazard events occurred in South Tyrol, Italy, were explored and the process-structure interaction was quantified through the damage characteristics *DamPre*, *DamCh*, and *DamPost.* The survey of the damage characteristics was carried out through field inspections, measurements and subjective state

TABLE 7 Excerpt of the rule base of the simplified general model with a binary decision variable functionality (yes/no) comprising 10 rules with the largest coverage

Nr.	Rule	LHS support	RHS support	RHS accuracy	LHS coverage	RHS coverage
1	h_0 ([*, 2.23]) AND Material(1) = > Functionality (YES)	52	52	1	0.23	0.23
2	Age(1) = > Functionality (YES)	26	26	1	0.11	0.14
28	h_0 ([2.95, 3.05]) AND Inclination ([0.38, *]) AND Diameter ([1.39, *)) = > Functionality (NO)	5	5	1	0.02	0.12
3	h_0 ([2.28, 2.95]) AND DamPre ([*, 0.04]) = > Functionality (YES)	22	22	1	0.10	0.12
4	Flowtype (3) AND w_0 ([5.90, 6.50]) = > Functionality (YES)	21	21	1	0.09	0.11
5	w_0 ([8.85, 9.25]) AND Material (1) = > Functionality (YES)	20	20	1	0.09	0.11
7	h_0 ([2.28, 2.95]) AND History (1) = > Functionality (YES)	18	18	1	0.08	0.10
7	h_0 ([2.28, 2.95]) AND History (1) = > Functionality (YES)	16	16	1	0.07	0.09
8	Flowtype (3) AND h_0 ([3.05, 3.55]) AND Geometry ([2.85, 4.90]) = > Functionality (YES)	16	16	1	0.07	0.09
9	w_0 ([13.30, 14.10]) = > Functionality (YES)	14	14	1	0.06	0.08

LHS = left hand side; RHS = right hand side.

assessments. Specifically, an increased damage impact on the central structural parts of the structures could be retraced thereby confirming previous results obtained by Von Maravic (2010) and Dell'Agnese et al. (2013). Based on the distribution of the damage characteristics, namely the Pre-Event Damage conditions (*DamPre*), the Post-Event Damage conditions (*DamPost*), and the associated condition Change (*DamCh*), it can be stated that check dams built of concrete show a better performance with respect to resistance and exhibit also longer durability in comparison to masonry check dams. These results are consistent with previous findings reported in literature (Hübl & Fiebiger, 2005). Based on the surveyed damage characteristics and depending on the construction material, a remarkable difference in damage behaviour could be detected. Check dams built of concrete exhibit more uniform damage patterns and total structural failure could not be observed. Fore masonry check dams, instead, (a) a larger variability with respect to the observed damage patterns, (b) single cases of total structural failure, and (c) a faster damage progression with respect to the material type concrete (i.e., particularly if the considered structure was pre-damaged) could be retraced. Previous attempts to provide statistical tools for the prediction of event based damage susceptibility by applying multiple linear and non-linear regression models failed to provide conclusive insights. As also Dell'Agnese et al. (2013) pointed out, the collected data set did not allow reliable relationships between event characteristics and expected damage to check dams to be established. In fact, many variables other than event intensity may exert some influence on the physical vulnerability of these structures, that is, type, age, geometry, pre-event conditions, and event type, thus resulting in an extremely high number of possible combinations to be assessed in order to determine regression equations for each combination.

To overcome such difficulties and with the aim to provide a decision support tool for mountain stream managers, Rough Set Analysis was applied to derive compact rule bases out of structured empirical data sets. As outlined in the previous section both in theory and practice, four Rough Set models were set up and the associated rule bases were derived as a support for an enhanced prediction of damage susceptibility. The accuracies of the damage susceptibility predictions (compare previous section) accuracies obtained by the application of the Rough Set Data Analysis model provide a reliable basis to improve check dam maintenance strategies.

It can be stated on a more general level that being able to predict the performance decay of check dams in case of debris flows in mountain streams allows to enhance the design of such structures by properly take into consideration the effects of process-structure interaction and, as a consequence, to identify more suitable construction sites, to better quantify both the overall number and necessary dimensions of planned check dams to achieve the desired hazard mitigation result. From a check dam management perspective, the developed Rough Set Data Analysis model may contribute to accomplish the shift from a reactive maintenance approach based on capillary and expensive post-event onsite inspections, to a proactive maintenance strategy, which, based on predicted performance decays, may allow preventive maintenance and reinstatement works for key functional elements of the protection system.

To further improve the prediction results of the setup Rough Set Data Analysis model and to open new research possibilities in the field, it is essential to refine data collection protocols to accurately collect field data for both pre and post event check dam conditions and reliably assess local process intensities. A larger data set may improve the prediction accuracies of the obtained rule bases, or, alternatively, lead to their slight modification based on the support, accuracy and coverage measures of the rule set. In this respect, Trenkwalder-Platzer (2014) points out that the consideration of further condition attributes C may lead to an overall improvement of the prediction performance of the developed models. As already mentioned, the evaluation of

pre- and post-event damage conditions still requires the use of expert judgement. This partial gap, however, indicates that agencies in charge of the life-cycle management of check dams should invest appropriate amounts of resources to continuously monitor through proper measuring instruments the structural resistance parameters to systematically reduce subjective evaluation elements.

In parallel, it is essential to systematically collect data about ongoing scouring process downstream of the considered check dams (see Comiti et al., 2010), to monitor seepage filtration processes (Suda, 2012) and to use such data for broader prediction purposes.

Rough Set Data Analysis might be applied with success also to other damage susceptibility problems in flood risk management, in particular if statistical techniques fail to completely make sense of data related to pre- and post-damage conditions of structures exposed to flood impacts (e.g., vulnerability of residential buildings, damage susceptibility of bridges clogged by large wood).

ACKNOWLEDGEMENTS

Bruno Mazzorana has been supported by the Iniciativa Científica Milenio (ICM) through grant NC160025 “Millennium Nucleus CYCLO - The Seismic Cycle Along Subduction Zones”.

ORCID

B. Mazzorana http://orcid.org/0000-0003-1218-4495

REFERENCES

Armanini, A., & Larcher, M. (2000). Rational criterion for designing opening of slit-check dam. *Journal of Hydraulic Engineering*, *127*, 94–104.

Bergmeister, K., Suda, J., Hübl, J., & Rudolf-Miklau, F. (2009). *Schutzbauwerke gegen wildbachgefahren: Grundlagen, entwurf und bemessung, beispiele*. Berlin: Ernst & Sohn.

Comiti, F., Lenzi, M. A., & Mao, L. (2010). Local scouring at check dams in mountain rivers. In M. Conesa-Garcia & M. A. Lenzi (Eds.), *Check dams, morphological adjustments and erosion control in torrential streams* (pp. 263–282). Hauppauge, NY: Nova Publ.

Dell’Agnese, A., Mazzorana, B., Comiti, F., Von Maravic, P., & D’Agostino, V. (2013). Assessing the physical vulnerabilità of check dams through an empirical damage index. *Journal of Agricultural Engineering*, *XLIV*, e2.

Düntsch, I., & Gediga, G. (2000). Rough set data analysis. *Encyclopedia of Computer Science and Technology*, *43*(28), 281–301.

Fuchs, S., Keiler, M., Sokratov, S., & Shnyparkov, A. (2013). Spatiotemporal dynamics: the need for an innovative approach in mountain hazard risk management. *Natural Hazards*, *68*, 1217–1241.

Hübl, J., & Fiebiger, G. (2005). Debris-flow mitigation measures. In M. Jakob & O. Hungr (Eds.), *Debris-flow hazards and related phenomena*. Berlin: Springer.

Johnson, D. S. (1974). Approximation algorithms for combinatorial problems. *Journal of Computer and System Sciences*, *9*, 256–278.

Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis*. New Jersey: Prentice-Hall.

Mazzorana B. (2008). *Das EF30Forward Konzept: ein Hinweisinstrument zur Ermittlung der Zuverlässigkeit und Funktionseffizienz von Wasserschutzbauten.* International Symposium Interpraevent, Klagenfurt, Austria, *1*:415–424.

Mazzorana B., & Fuchs S. (2010). *A conceptual planning tool for hazard and risk management.* International Symposium Interpraevent, Taipei, Taiwan.

Mazzorana, B., Simoni, S., Scherer, C., Gems, B., Fuchs, S., & Keiler, M. (2014). A physical approach on flood risk vulnerability of buildings. *Hydrology and Earth System Sciences*, *18*, 3817–3836.

Mazzorana, B., Trenkwalder-Platzer, H., Fuchs, S., & Hübl, J. (2014). The susceptibility of consolidation check dams as a key factor for maintenance planning. *Österreichische Wasser- und Abfallwirtschaft*, *66*(5–6), 214–216.

Munakata, T. (2008). *Fundamentals of the new artificial intelligence: Neural, evolutionary, fuzzy and more*. London: Springer.

Olson, D., & Delen, D. (2008). *Advanced data mining techniques*. Berlin: Springer.

Pawlak, Z. (1997). Rough set approach to knowledge-based decision support. *European Journal of Operational Research*, *99*(1), 48–57.

Rimböck, A., Höhne R., Rudolf-Miklau, F., Pichler, A., Suda, J., Mazzorana, B., & Papež, J. (2014). *Persistence of Alpine natural hazard protection.* Platform on Natural Hazards of the Alpine Convention. Retrieved from http://www.naturgefahren.at/eu-internationales/planalplcm.html

Rinaldi, M., Surian, N., Comiti, F., & Bussettini, M. (2011). The morphological quality index (IQM) for stream evaluation and hydromorphological classification. *Italian Journal of Engineering Geology and Environment*, *11*(1), 17–36.

Rutkowski, L. (2008). *Computational intelligence: Methods and techniques*. Berlin: Springer.

Shuib, N. L. M., Bakar, A. A., & Othman, Z. A. (2009). *Performance study on data discretization techniques using nutrition dataset.* International Symposium on Computing, Communication, and Control (ISCCC 2009), Singapore. 304–308.

Skowron, A., & Rauszer, C. (1992). The discernibility matrices and functions in information systems. In R. Słowinski (Ed.), *Intelligent decision support: Handbook of applications and advances of rough set theory, volume 11 of System Theory, Knowledge Engineering and Problem Solving* (pp. 331–362). Dordrecht: Kluwer.

Suda, J. (2012, 2012). *Instandhaltung von Schutzbauwerken gegen alpine Naturgefahren [Maintenance strategies for protection works]*. Wien: Guthmann-Peterson.

Trenkwalder-Platzer H.J. (2014). *Vulnerabilitätsanalyse zur Entwicklung einer nachhaltigen Instandsetzungsstrategie von Wildbachsperren angewandt am Gadriabach und am Hölderlebach, Südtirol.* (Unpublished master’s thesis). BOKU—Universität für Bodenkultur, Vienna.

Üblagger G. (1972). *Retendieren, Dosieren und Sortieren (Retaining, dosing and sizing).* Mitteilungen der forstl. Bundesversuchsanstalt, Wien, Austria, *102*:335–372.

Valentini, R. (2012). *Analisi modellistica dell'efficienza delle opere idrauliche in alveo.* (Unpublished master’s thesis). Universitá degli Studi di Trento, Trento.

Vinterbo, S., & Øhrn, A. (2000). Minimal approximate hitting sets and rule templates. *International Journal of Approximate Reasoning*, *25*(2), 123–143.

Von Maravic P. (2010). *Evaluation of the physical vulnerability of check dams exposed to the impacts of torrential processes through experimental analysis.* (Unpublished degree dissertation). University of Padova, Padova.

Vorogushyn, S., Merz, B., & Apel, H. (2009). Development of dike fragility curves for piping and micro-instability breach mechanisms. *Natural Hazards and Earth System Science*, *9*, 1383–1401.

How to cite this article: Mazzorana B, Trenkwalder-Platzer H, Heiser M, Hübl J. Quantifying the damage susceptibility to extreme events of mountain stream check dams using Rough Set Analysis. *J Flood Risk Management*. 2018;e12333. https://doi.org/10.1111/jfr3.12333