ELSEVIER

Research article

Standardized disaster and climate resilience grading: A global scale empirical analysis of community flood resilience

Stefan Hochrainer-Stigler [*], Laurien Finn, Stefan Velev, Adriana Keating, Reinhard Mechler

IIASA- International Institute for Applied Systems Analysis, Laxenburg, Austria

ARTICLE INFO

ABSTRACT

Suitable and standardized indicators to track progress in disaster and climate resilience are increasingly considered a key requirement for successfully informing efforts towards effective disaster risk reduction and climate adaptation. Standardized measures of resilience which can be used across different geographical and socioeconomic contexts are however sparse. We present and analyze a standardized community resilience measurement framework for flooding. The corresponding measurement tool is modelled based on and adapted from a so-called 'technical risk grading' approach as used in the insurance sector. The grading approach of indicators is based on a two-step process: (i) raw data is collected, and (ii) experts grade the indicators, called sources of resilience, based on this data. We test this approach using approximately 1.25 million datapoints collected across more than 118 communities in nine countries. The quantitative analysis is complemented by content analysis to validate the results from a qualitative perspective. We find that some indicators can more easily be graded by looking at raw data alone, while others require a stronger application of expert judgement. We summarize the reasons for this through six key messages. One major finding is that resilience grades related to subjective characteristics such as ability, feel, and trust are far more dependent on expert judgment than on the actual raw data collected. Additionally, the need for expert judgement further increases if graders must extrapolate the whole community picture from limited raw data. Our findings regarding the role of data and grade specifications can inform ways forward for better, more efficient and increasingly robust standardized assessment of resilience. This should help to build global standardized and comparable, yet locally contextualized, baseline estimates of the many facets of resilience in order to track progress over time on disaster and climate resilience and inform the implementation of the Paris Agreement, Sendai Framework, and the Sustainable Development Goals.

1. Introduction

The number of disaster events as well as the magnitude of disaster losses have been increasing over time (MunichRe, 2018; SwissRe, 2018). Floods are especially devastating and were the most frequent type of disaster globally (43 percent of all recorded events in EM-DAT) and also affected the largest number of people (more than two billion) between 1998 and 2017 (CRED and UNISDR, 2018). The current understanding is that these increases have been largely driven by growth in vulnerability and exposure of humans and assets to natural hazards (Meyer et al., 2013). While the frequency and severity of natural hazards are already being influenced by climate change (IPCC, 2018), observed impacts and projected risks are also strongly determined by non-climatic factors

(Bouwer, 2019; Mechler and Bouwer, 2015). Addressing increasing disaster risk to build disaster and climate resilience therefore requires a deeper understanding of the factors underlying and causing natural hazard-induced disasters (Birkmann et al., 2015; Chang et al., 2018). The consideration of resilience as a multidimensional concept has been identified as having potential for contributing to this understanding, as well as for identifying effective and efficient options for reducing and managing risk today and in the future (Cai et al., 2018; Hochrainer-Stigler et al., 2016; Keating et al., 2016; Keating and Hanger-Kopp, 2020; Chang et al., 2015).

The development of suitable indicators to track progress in disaster and climate resilience is currently seen as a key requirement for informing successful efforts towards disaster risk reduction and climate

adaptation (see Asadzadeh et al., 2017 for a systematic overview). For example, resilience is one of the key terms used in the Sendai Framework for Risk Reduction and is linked to various quantitative and qualitative targets and priorities (UN, 2015a). Similarly, the Sustainable Development Goals mention resilience prominently across numerous dimensions and targets (UN, 2015b; see also Hák et al., 2016). Finally, the Paris Agreement is framed around an ambition of climate-resilient development (UNFCCC, 2015).

While a myriad of definitions and conceptualizations of disaster resilience have been put forward by researchers, multilateral organizations, development agencies and non-governmental organizations (NGOs) (Mochizuki et al., 2018; Cerè et al., 2017; Pasteur and McQuistan, 2016; Keating et al., 2016; Constas and Barrett, 2013; Folke et al., 2010; UNDRR, 2009), standardized measures of resilience that can be used across different geographical and socioeconomic contexts are sparse (Cai et al., 2018; Cutter et al., 2014; Sherrieb et al., 2010). However, standardized measures are required in order to compare resilience levels between risk owners (e.g. households, communities or countries), track progress over time and establish best practices. This paper analyses one such approach (see Gibbons et al., 2020 for another example in a different context): the Flood Resilience Measurement for Communities (FRMC) - a standardized community flood resilience measurement framework modelled on a so-called 'technical risk grading standard' (TRGS) approach (Keating et al., 2017). The TRGS was developed and is used by Zurich Insurance Group and was adapted to be used in the Zurich Flood Resilience Alliance (ZFRA).

The FRMC adapted the TRGS approach to the context of community resilience to flooding. The approach brings together quantitative and qualitative data about the attributes, resources, and capacities that contribute to community flood resilience, allowing trained assessors to "grade" these factors based on the TRGS approach. The central feature of this approach is that both quantitative and qualitative indicators of resilience are graded on the same ordinal scale (A, B, C, D). This feature sets this approach apart from other current efforts to measure resilience that most often use different scales for different dimensions (e.g. percent of population, dollar values, etc.).

The data used for our analysis come from a large-scale application of the FRMC approach in 118 communities over 2016 to 2018 across nine countries, generating over 1.25 million datapoints. The core objective of this paper is to analyze if the FRMC approach may operate as a TRGS for resilience or in other words a "Technical Resilience Grading Standard" (TResGS), and what operational lessons can be learned in that regard. In doing so we constructed several classification and performance indicators in order to explore specific aspects of the operationalization effort, including confidence of grading as well as the robustness of the grading over time.

Based on the findings of our analysis we identify recommendations for future iterations of the FRMC and similar efforts for standardized resilience measurement frameworks and indices. Critically, we find that the TResGS approach is especially useful for assessing qualitative, subjective resilience dimensions such as ability, perception, feel, trust, and so on. Assessing these types of subjective indicators is highly dependent on expert judgment, whereas grading of more objective, quantifiable resilience dimensions such as poverty rates or financial savings are more directly attributable to the raw data gathered and therefore need less involvement of experts for grading. Furthermore, we find that the specific way in which the grade definitions are specified, e.g. through intervals or based on extrapolation from household samples, significantly influences the grading. The findings have important implications for building standardized resilience indicators in real-world settings and can be used as guidelines for the construction of new ones for other types of risks.

Our paper is organized as follows: the next section gives a short overview of the FRMC, including background information, the dataset, and how the raw data was collected, and grades assigned by experts. Section 3 then describes the methodology applied including a discussion of the categorization and performance indicators used. Section 4 presents the results and section 5 discusses them comprehensively, focusing on six key messages. Finally, section 6 concludes and provides an outlook.

2. Research area and data

2.1. Measuring resilience: community flood resilience measurement framework and tool

While there is some general agreement that resilience can be understood as a multidimensional capacity (Keating et al., 2016; Campbell et al., 2019; Laurien et al., 2020) having a clear and agreed conceptualization of resilience is important for collaborations based on the concept. The ZFRA reviewed existing definitions and built on these to conceptualize disaster resilience as: "The ability of a system, community or society to pursue its social, ecological and economic development objectives, while managing its disaster risk over time in a mutually reinforcing way" (Keating et al., 2016). This conceptualization is centered on the capacity of a community to achieve its wellbeing goals (holistically defined) in the face of disaster risks. In the context of community flood resilience, this conceptualization is refined to: "The ability of a community to pursue its social, ecological and economic development objectives, while managing its flood risk over time in a mutually reinforcing way." In other words, to thrive in the face of flood risk. This conceptualization underpins the FRMC.

When the Zurich Flood Resilience Alliance took on the task of designing a community flood resilience measurement framework (see Keating et al., 2017) based on the conceptualization described above, they found that the multiple, diverse factors contributing to community flood resilience and clear need for multiple data sources is an environment similar to that faced by 'risk engineers'. Zurich Insurance risk engineers assess risk to a facility via a TRGS: a standardized measure that assesses the facility's characteristics, and risk-management interventions and processes. The TRGS is used to organize and make sense of the data gathered about the facility under assessment and to provide a consistent benchmark against which to quantify risk. For each peril/hazard, it offers a tool that takes into account the different factors that make up the risk associated with that peril. Each peril-specific TRGS includes a number of risk categories, and each of the categories is made up of several risk factors. The risk factors are graded according to pre-defined evidence/data that the risk engineer collects. Grading is on a scale from A-D with A meaning 'best practice for managing the risk' and D meaning 'significantly below good standard, potential for imminent loss'. Risk engineers compare data gathered from their site visit with the grade definitions in the TRGS to allow them to make a judgment about the level of risk and then use the results to conduct conversations with the customer about how to manage the risks they are facing.

The TRGS approach was adapted to community flood resilience because it helps to make sense of diverse data gathered about the situation being assessed (Keating et al., 2017). The Gen 1 (generation 1; there is currently a second generation developed based partially on the results reported here) FRMC framework and associated tool explored here was designed to comprehensively measure community level resilience to flooding in the form of a TResGS. It consists of 88 indicators, called "sources of resilience" (henceforth 'sources') which are based around the five capitals of the Sustainable Livelihoods Framework (DfID, 1999), i.e., the 88 sources are split across the 5 capitals, namely human, social, physical, natural, and financial (see Supplementary I and for a full discussion Keating et al., 2017). Sources were identified for each of the five capitals (5C) based on literature and expert input (see Keating et al., 2017 for a detailed description of framework development). A necessary criteria for a source of resilience to be included was that it needed to provide one (or more) of the 4 properties of a resilient system (4R): robustness, redundancy, resourcefulness, and rapidity (Bruneau, 2006; Cimellaro et al., 2010).

This conceptual framework was then operationalized via the (Gen 1) Flood Resilience Measurement Tool (FRMT) - an integrated, web-based and mobile device platform that implementation teams use to collect data on the 88 sources of resilience through one or more of five data collection methods selected by the users: household surveys, community focus group discussions, key informant interviews, interest group discussions, and third party data. The FRMC approach includes pre-defined questions for each data collection method for each source of resilience. For the purposes of this paper we refer to the answers to these questions as the 'raw data'.

Raw data was collected by field teams in a collaborative fashion involving both community stakeholders and NGO/humanitarian organization partners. The raw data collected was then used by trained expert assessors from within the user organization to assign a grade from A to D (A being the best and D being the worst) for each of the 88 sources of resilience, according to specific definitions for each grade for each source (from now on called expert judgement or simply the grades). Furthermore, it was possible to include comments in regard to user's confidence of the grading and any specific problems they wanted to note in regard to the grading process. As indicated, this approach is similar to the TRGS approach used by risk engineers at Zurich Insurance Group. Grade results can then be displayed in various ways, including according to the 5Cs framework, to inform and enable a discussion on how to identify potential measures for building resilience in the respective community. Fig. 1 summarizes the FRMC process.

Grading was undertaken by in-country community program managers. They were development practitioners working in the flood risk management and development space, and had strong links to the communities where the FRMC was being applied. Because of the significant role that expert judgement plays in the grading process, it was essential that graders be trained to use the FRMC. This training was also essential to ensure standardization of the grades across graders and communities. All expert graders were trained in a week long workshop in Zurich, Switzerland by the ZFRA. The training included evaluations to ensure raw data were graded in a standardized way across graders. Expert graders were provided with extensive written material on the underlying concepts and grading process, as well as ongoing support from the FRMC design team. Expert graders were in all cases supported in the grading task by a team of local colleagues including data collectors and those with intimate knowledge of the community being graded.

Summarizing the FRMC process: for each of the 88 sources of resilience (middle of Fig. 1), the FRMC platform requires users to select data collection methods for each source (left hand side of Fig. 1), and assign the data collection work to individual field team members (middle of Fig. 1). Data collected in the field via the mobile app is automatically updated in the online platform where experts use it to assign grades. During the grading process assessors are asked to provide information about the grading confidence and relevance. Finally, the tool generates tables and graphs to help visualize and analyze the results (right hand side of Fig. 1). All collected data is stored in a secure and password protected database.

We pre-processed the raw data as well as corresponding source grades to make them manageable for our analysis. Given the large amount of data – raw data and grades from 118 communities - we wrote a script in the JSON (JavaScript Object Notation) coding language, an advantage of this software being that it is easily readable, intuitive and can be relatively easily implemented in an internet webpage environment. The entire data management process was built according to the Google database guidelines (Google Inc, 2016, 2012; Sato, 2012). The final outcome of the data management process has been a large multi-layer table, containing all raw data questions and answers, and grades, in one place.

In the next sub-section, we provide a specific example of how users selected and collected the raw data and graded it using the TResGS approach, in order to facilitate a better understanding of the actual grading process.

2.2. Raw data collection and grading example

As an illustrative example of how the raw data collection and grading process was implemented, we present one of the 88 sources of resilience fully specified – 'Access to school facilities'. The source is explained to the assessor with the following description and instructions:

"This aspect of the education theme considers the adequacy of the infrastructure to support provision of education and how it stands up in flood situations - Schooling is an important aspect of daily life. Both the interruption itself and the lost education time lead to problems (children at home instead of daily rhythm at school). Schooling during floods should obviously be conducted only where and when it is safe to do so depending on the flood scenario. For flash flood situations, rapidity and robustness is key and schooling should resume as soon as possible. For long-standing, large-scale standing water flood situations, it is important that schooling can continue, such as in alternate locations or safe locations."

Data for this source must be collected via at least one of the five data collection methods as appropriate to context as determined by the user. Data collection questions for each method are shown in Table 1 below.[1] Once a data collection method is selected for a source, all questions specified for that method must be asked.

The TresGS for the 'Access to school facilities' source of resilience is graded A to D with the following guidance:

- **A:** School facility (or location where formal school setting takes place) is built robust, located away from flood zone and accessible through safe and protected ways even during and after floods – schooling continues to take place.
- **B:** School facility is impacted by flooding but maintains sufficient basic staffing and equipment to provide care. OR school may be impacted but informal schooling is planned to go on in a safe place during and after floods.
- **C:** School facility is impacted and cannot avoid significant lost school curriculum. OR while informal schooling may be available, it is unplanned or inconvenient and leads to significant lost school curriculum.
- **D:** No schooling facility. OR school prone to damage rendering it inoperational during flood. OR school not accessible during flood for either teachers or students.

2.3. Data used

Our data came from the application of the FRMC by five organizations working in twelve country programs across nine countries, with a total of 118 communities. The FRMC was implemented in two time periods in each community (from now on called baseline and endline), between 2015 and 2018. The selection of communities was based on a set of criteria including need for external support, history of past flood events (high flood risk), location of communities in the broader river basin (and representativeness for their region), and willingness to be part of the program. In total more than 350,000 households or approximately 1 million people are located in communities reached by the FRMC in phase I of the Alliance (Table 2). Our dataset includes more than 10,000 data points of source grades (88 grades in 118 communities) and more than 1.25 million data points from raw data collection.

While the criteria for selecting communities were similar across the country programs, the selected 118 communities vary with regard to several key community characteristics that likely impact community flood resilience. For example, the communities ranged in terms of settlement type between urban (20%), peri-urban (30%) and rural (50%)

---

[1] It should be noted that while this example source allows mostly dichotomous answers for all data source options, allowed answers vary from yes/no to other response lists, and free form entry.
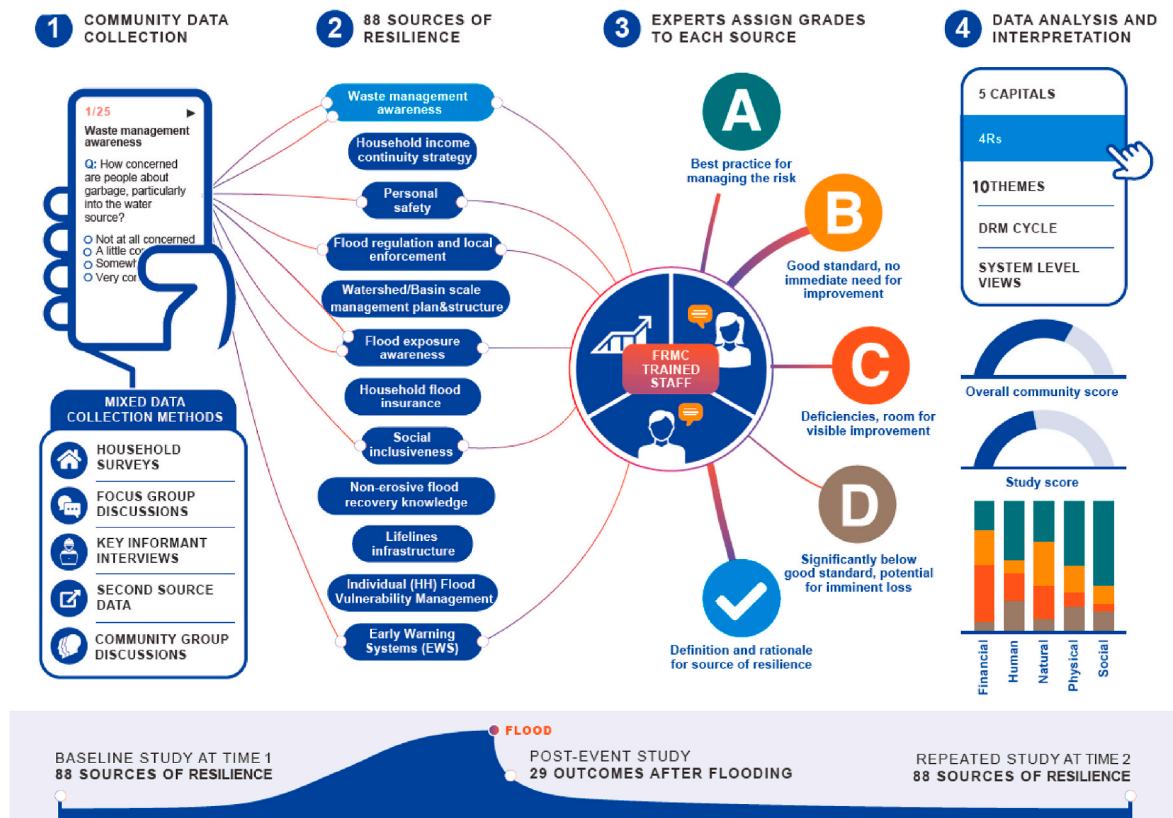
Fig. 1. FRMC data implementation process. Source (Laurien et al., 2020).

settings. The majority of rural communities are in Afghanistan, Mexico, Nepal, Timor-Leste and Bangladesh, while the majority of urban communities are in Indonesia (together with Peru, Haiti and USA) (see the discussion in Laurien et al., 2020 for more information on the community characteristics).

3. Methodological approach

As shown in section 2, the process of translating raw data to resilience grades is standardized via a clear set of guidelines about raw data collection and corresponding questions, and how to interpret raw data according to the grade definitions. These guidelines, questions and expert grading process make up the TResGS, which is analogous to the TRGS approach. It must be emphasized that for each of the 88 sources of resilience across the five capitals, very different question types are needed to inform grading, e.g. social capital sources of resilience need quite different raw data questions in terms of tangibility, compared to financial capital ones. Furthermore, feedback from users indicated that grading must be undertaken by experts who are very familiar with the community as the raw data itself needs to be interpreted.

We explored the relationship between raw data and resultant grades to shed light on 1) which data collection methods are most appropriate for which sources of resilience, 2) what type of questions are most closely linked to specific grading outcomes, and 3) what type of questions under which circumstances are most suited to TResGSs. To undertake this analysis, we utilize multiple lines of evidence, i.e. both advanced quantitative approaches (e.g. multivariate generali linear modelling) and more qualitative ones (e.g. bayes classifier content analysis) . The results of both the quantitative and qualitative analysis are used to build a classification scheme and performance index, the results of which will be summarized into six key messages at the end of the paper.

As indicated, the FRMC database is very large, multifaceted and complex and therefore implementing traditional data analysis techniques, e.g. manually testing a set of models, proved to be challenging and thus automated big data analysis techniques were employed, as discussed next.

3.1. General linear modelling approach

How the raw data is related to actual grades is one of the main research questions addressed here. There are various ways to approach this question, a major one being statistical analysis. Due to the mixed-method approach of the FRMC the raw data variables have various scales. We therefore first applied a general linear modelling (GLM) approach because it can simultaneously manage continuous, ordinal, nominal and binary variables to estimate model parameters. The (Likert-like) grades were treated as continuous variables as usually done in social science research. Due to the size of the dataset we implemented an algorithm to select a suitable regression model automatically. The model selection process first identified a set of best models from all possible models (the candidate set) based on a genetic algorithm. Models in the candidate set were then ranked according to two well-accepted information criteria, namely the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). As our dataset contained two time periods (baseline and endline for each community), we further distinguish between a baseline and endline GLM for each source of resilience. Additionally, we performed diagnostic checks of the final models to test whether they met the GLM assumptions. The most important information used for the classification scheme and performance index discussed below are the final model equation itself and the variance explained by the model (i.e. adjusted R-square).

3.2. Content analysis and expert grading

We also conducted a semantic text analysis using the comments

Table 1

Possible methods and corresponding questions for gathering raw data for the grading of 'Access to school facilities'. Source: (Zurich Flood Resilience Alliance internal report, 2017).

Household survey questions	HH Answers
Does school take place during and after flood events? *(this may be due to damage to the school or the way to get to school, but also because the school is needed for emergency shelter)*	1 – Yes/2 – No
Has the school facility been damaged during the last floods so it could not operate anymore?	1 – Yes/2 – No
Can schools be reached during and after floods safely by staff and students?	1 – Yes/2 – No
Community questions	**Community Answers**
Does school take place during and after flood events? *(this may be due to damage to the school or the way to get to school, but also because the school is needed for emergency shelter)*	1 – Yes/2 – No
Has the school facility been damaged during the last floods so it could not operate anymore?	1 – Yes/2 – No
Can all reach the school facility during flooding?	1 – All/2 – Some/3 – None
Key informant questions	**Key informant Answers**
Ask e.g. the principal: Locate school facility or where schooling/teaching takes place on a map – Do schools get affected during floods? Do schools get used as emergency shelter and thus schooling is interrupted?	1 – Yes/2 – No
Has the school facility been damaged during the last floods so it could not operate anymore?	1 – Yes/2 – No
Can all reach the school facility during flooding?	1 – All/2 – Some/3 – None
Interest group questions	**Interest group Answers**
Ask the teachers group: Locate school facility or where schooling/teaching takes place on a map – Do schools get affected during floods? Do schools get used as emergency shelter and thus schooling is interrupted?	1 – Yes/2 – No
Has the school facility been damaged during the last floods so it could not operate anymore?	1 – Yes/2 – No
Can all reach the school facility during flooding?	1 – All/2 – Some/3 – None
Third party source questions	**Third party source Answers**
Locate school facility or where schooling/teaching takes place on a map – Do schools get affected during floods? Do schools get used as emergency shelter and thus schooling is interrupted?	1 – Yes, 2 – No
Has the school facility been damaged during the last floods so it could not operate anymore?	1 – Yes,2 – No
Can all reach the school facility during flooding?	1 – All,2 – Some, 3 – None

inserted into the FRMC tool by the expert graders. This content analysis was based on a simple natural language processing and applied three semantic analysis methods - a dictionary-based sentiment analysis, a naïve Bayes classifier, and content specific hard coded rules. The reason for using a multi-method approach here is due to the drawbacks for each method. For example, the shelf dictionary-based method does not account for content specific words; the word "disaster" has a negative connotation, however because floods are a main topic within the comments in our context the word "disaster" would have a neutral meaning.

Each comment went through three steps. It was first checked against a hard-coded set of customized rules as well as the sentiment analysis. If two conflicting results were found (e.g. a positive rating in the sentiment analysis and a negative rating in the hard-coded rule) then the comment was flagged, manually checked and rated. If there was agreement between the two it was rated accordingly. If there was not a clear result (e. g. a positive rating in the sentiment and a neutral in the hard-code rules) then the comment was rated based on the naïve Bayes algorithm.

If the content analysis revealed that the source was unclear, difficult to understand or grade, or not enough information was available from the raw data, then the comment receives a rating of minus one. If the comment was neutral or not well specified regarding difficulties, then it received a grade of zero, if the source was easy to be graded and/or understandable it would receive a positive grade (plus one). Based on the results for the comments on each source, we calculated a Grader Confidence Rating (GCR) between zero and one for each source. The GCR was calculated as the number of total comments minus the number of comments rated negative one divided by the number of total comments. Hence the larger the GCR the more confident the experts were with the grading.

3.3. Classification schemes and performance index

To interpret our results we built a five-category classification scheme and classified each source of resilience as either 1) Stable and Predictable, 2) Unstable but Predictable, 3) Stable and Unpredictable, 4) Unstable and Unpredictable or 5) Not Applicable. Categories 1–4 are based on the stability/instability and predictability/unpredictability of the relationship between raw data and grade using the results of the GLM. Category 5 refers to our GLM approach, specifically that no final model

found, or model assumptions not fulfilled, or because of collinearity issues. However, we still looked at these sources but in a qualitative manner. It should be noted that being classified in category 2–5 does not mean that gathering raw data for that source of resilience is not useful; it simply means that graders may have to interpret the raw data in each individual case to make the actual grading.

The classification scheme and results aided further analysis that embedded these results into a performance index. The goal was to create an index that comprehensively assessed the need for expert grader judgement in assigning grades, given raw data. This was based on the assumption that some source grades can be reliably predicted based on the raw data, while others require interpretation of the raw data by expert graders. The performance index includes four indicators:

- the adjusted R-square for the baseline for predictability (in a statistical sense), which ranges from 0 to 1 (no predictability to full predictability);
- the adjusted R-square for the endline, as above;
- a simple matching coefficient (SMC) assesses to what extent the same significant independent variables are found in both the baseline and endline GLMs (ranging between zero and one). This is calculated according to the number of questions (i.e. the independent and significant raw data questions for the best model) appearing in both the baseline and endline models divided by the total number of unique questions; and;

Table 2

Summary of countries and communities where FRMC was applied, 2015–2018.

Country	Number of communities	Estimate of total population (000′)
Afghanistan	12	13
Bangladesh	9	39
Haiti	4	36
Indonesia	40	258
Mexico	19	7
Nepal	21	19
Peru	5	40
Timor-Leste	6	4
USA	2	640
Total	**118**	**1056**

- the Grader Confidence Rating (GCR) indicating how confident the grader is about the grading of the specific source of resilience, ranging between 0 and 1 (not confident to fully confident).

The performance index score is a simple average of the four indicators. A performance index score between 0 and 1 allows for easy interpretation of the results, where 1 means the grade can be assigned on the raw data only, and 0 means full expert judgement is required. We give some concrete examples of the related calculations in the results section. We then summarize the results in six key messages presented in the discussion section.

4. Results

Due to the large data output produced we focus here on main results and refer to the Supplementaries I–V for further details. Additionally, as part of the main results we present specific examples to increase readability and understanding of our findings.

4.1. Classification of sources of resilience

We start our discussion with a presentation of the adjusted R-squares found for each source of resilience and best GLM models for the baseline and endline cases. This goodness of fit measure indicates how well the estimated model can explain the variance observed in the empirical data. This is particularly pertinent as high values (i.e. close to one) indicate that the raw data can be used directly to grade a source (through the GLM model equation found). The baseline and endline regression models for all sources can be found in Supplementary II. Supplementary III also includes the estimated parameters for each source, and Supplementary IV includes the selected variables in their question form as well as the data collection method and Supplementary VI the corresponding adjusted R squares for each source. Overall, we find considerable fluctuations for each source of resilience for both the baseline and endline. Fig. 2 shows the R-squares from the best GLM models for each source of resilience, sorted by the baseline case. Based on this Figure we give one specific classification example for different sources below.

As an example, the "Educational attainment" source of resilience (named as H16, see Supplementary I for the abbreviations) showed large R-squares for both the baseline and the endline, and additionally showed the identified variables to be significant in the final GLM models (SMC score of one) (Supplementary II and Supplementary III). It is therefore categorized as "Stable and Predictable". This result can be partly explained by the fact that the grade definitions refer to percentage bands of community members that have completed primary school education. There is no qualitative interpretation required to grade this source, so the relationship between the raw data and grade is stable and predictable. Furthermore, the input method here was consistent across all communities, namely third-party sources which are arguably the most

appropriate method, given that this information can be found in official documents or from previous community studies.

Contrary to the example above, the source 'Mitigation financing' (F13) shows considerable fluctuations in the R-square for the baseline and endline GLMs, ranging from 0.85 to 0.23 respectively (see Fig. 2). It is therefore classified as unpredictable (in a statistical sense, this does not mean that it is actually unpredictable, as the expert grader would have used local knowledge to make sense of the raw data). However, the linear models (Supplementary II and Supplementary III) showed some significant similarities; specifically, the endline model was a reduced form of the baseline model. For this source of resilience, the best baseline model is a linear function of 4 raw data variables, all of them regarding whether enough money from the government is available to protect a given percentage of total homes from flooding. In functional form one may write this with our coding (see Supplementary I, II and IV) as F13 ~ F13K014DY + F13C017DY + F13I018DY + F13T019DY. In the best endline model only one raw data variable (whether or not any money is available to protect homes from flooding) was found, in functional form: F13 ~ F13K014DY. Consequently, the SMC score for 'Mitigation financing' (F13) was calculated to be 0.6. For this source all data collection methods were utilized to gather data. We therefore categorize it as Unstable but Predictable.

Moving to another example for our classification scheme, the already introduced source 'Access to School Facilities' (P05) showed an R-square of 1, indicating some collinearity issues. The linear models found were also quite different for the baseline and endline case (e.g. in our functional form P05 ~ P05T078MM + P05K075MM + P05K073DZ + P05K074DZ + P05D078DZ + P05D079DZ for the baseline and P05 ~ P05C080DZ + P05C081MM + P05I073MM + P05K075MM for the endline). Hence, a SMC score of only 0.4 was calculated. We included it in the "Not applicable" category.

For sources like "Income and Affordability" (F02) which have a similar structure as H16 (percentage band grade definitions resulting in a "stable and predictable" classification) some detailed data is needed in terms of variables but the best-fit for baseline and endline are near identical in terms of raw data inputs (baseline F02 ~ F02T004DZ + F02T005DZ, endline F02 ~ F02D005SS + F02D003SS + F02D004DZ + F02T004DZ + F02T005DZ). The R-square is however very small, and we therefore categorize it as stable but unpredictable. Finally, some sources such as "Habitat connectivity" (N02), did not show any stability or predictability and as such are classified as unstable and unpredictable.

For further interpretation of the results Fig. 3 shows spider diagrams for our categories. They indicate quite a diverse picture with the most surprising finding of this analysis being that the sources do not cluster into different capitals: each classification group contains sources from across the five capitals. Such behavior is not typically found in similar resilience indicator measurement research (Cutter, 2016). We can also make some observations in regard to data collection methods and the classification scheme. For example, the number of data collection methods used as well as the specific methods applied (e.g. HH or
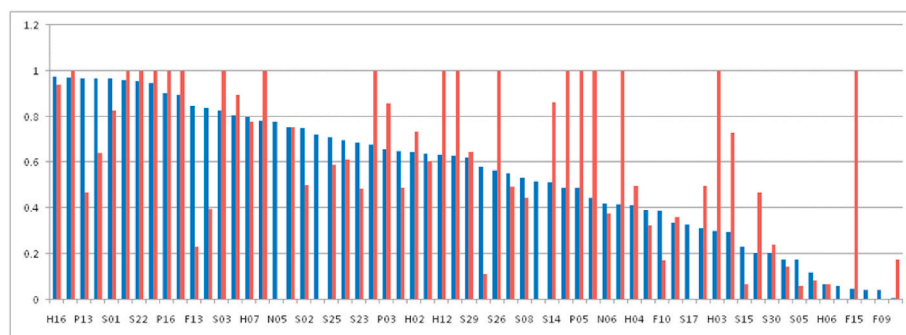

Fig. 2. Adjusted R square (y-axis) for the 20 best models found and corresponding sources of resilience (x-axis) for baseline (blue) and endline (red). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
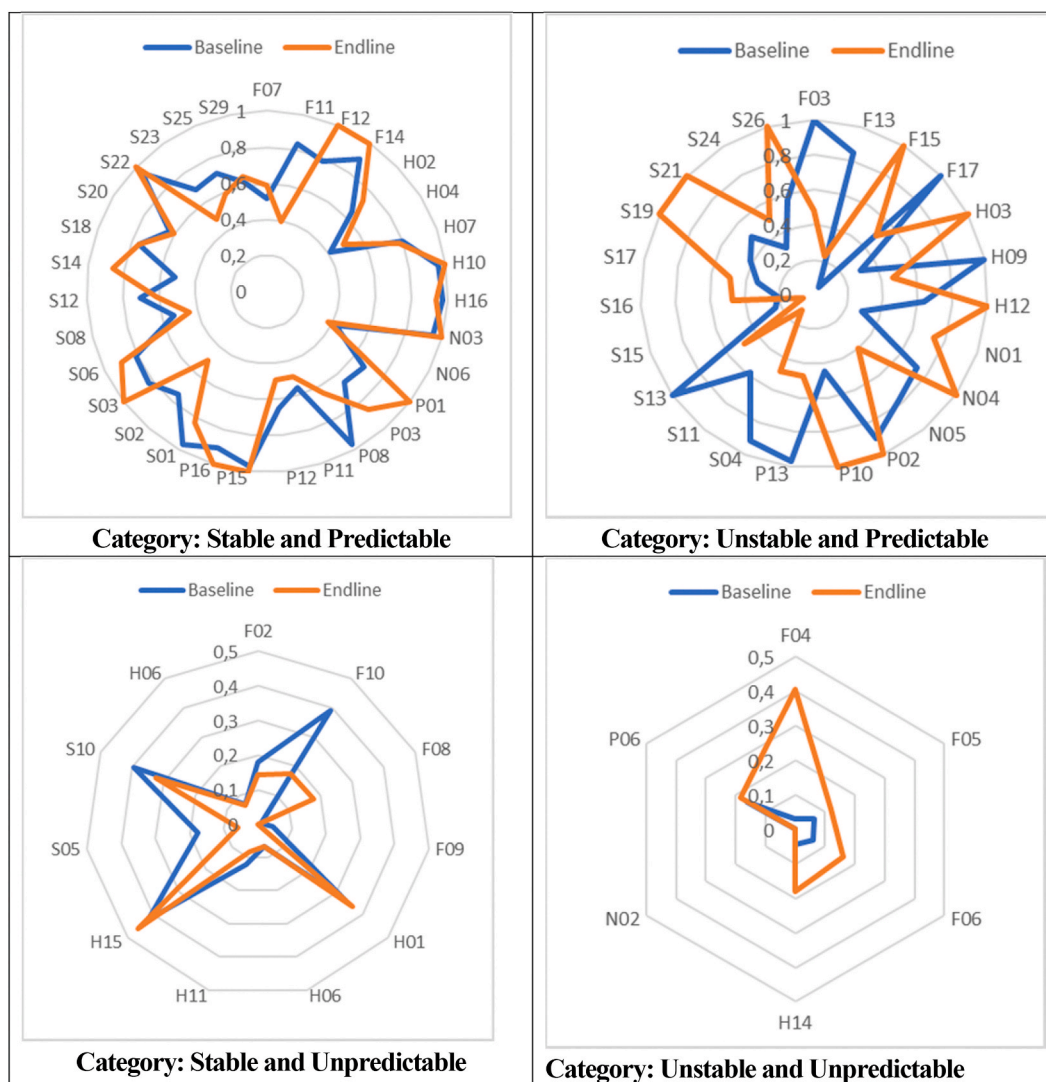
Fig. 3. Spider diagrams for endline and baseline sources of resilience according to the four categories.

third-party sources) fluctuates across the sources of resilience (Supplementary IV). Interestingly, we find that data collected via household surveys is often related to the unpredictability and instability category. On the other hand, predictability and stability increases if more than two data collection methods are used for the grading (see also the color grading in Supplementary IV). To shed more light on this complex picture, we next look at the more comprehensive performance index.

4.2. Performance index results

One additional dimension not explicitly taken into account yet is the confidence of the expert in regard to the grade assigned for the specific source of resilience. Furthermore, the analysis so far has looked at the GLM model results (e.g. R-square, model equations) individually. As discussed above, we designed a performance index to provide a more comprehensive assessment of the relationship between raw data and source grades. The performance index is comprised of four indicators: 1) adjusted R-square for the baseline, 2) adjusted R-square for the endline, 3) the simple matching coefficient (SMC), and 4) the Grader Confidence Rating (GCR). All of these indicators were standardized to be between 0 and 1 and the average taken and hence we assume equal weights for each. The closer the performance index is to 1, the more the raw data can predict the grade assigned.

In the categorization analysis above we identified the importance of

data collection method and number of methods used. We also find a complex interrelationship between these with no single variable - such as capital group, data collection method used or the number of methods used - that sufficiently explains the distribution of the sources across our categories. The performance index shows similar findings. For example, if we separate the index into four categories (i.e. 1 if the index is below 0.25, 2 if is between 0.26 and 0.5, 3 if is between 0.51 and 0.75 and 4 if it is above 0.75) and relate them to the percentage of data collection method used across the sample, a non-linear relationship can be found. For indications, Fig. 4 shows this for the baseline case and two methods. Similar patterns emerged for the other methods used, and when using endline data.

As Fig. 4 shows, the percentage of the 118 communities which used HH survey questions for a source in the baseline is related to a decrease in the performance index for that source. The opposite relationship exists for the percentage that used key informants, which is positively related to the performance index. An inverse U-shaped relationship with the performance index was found for the other data collection methods.

To explore whether these results may be related to the fact that our performance index assigns equal weights to the four indicators, we performed hierarchical cluster analysis using the Wald criterion for the four indicators as well. Three sub-groups with approximately the same number of observations were detected (see the dendrogram in Supplementary V). Regarding the four indicators in these sub-groups, sub-
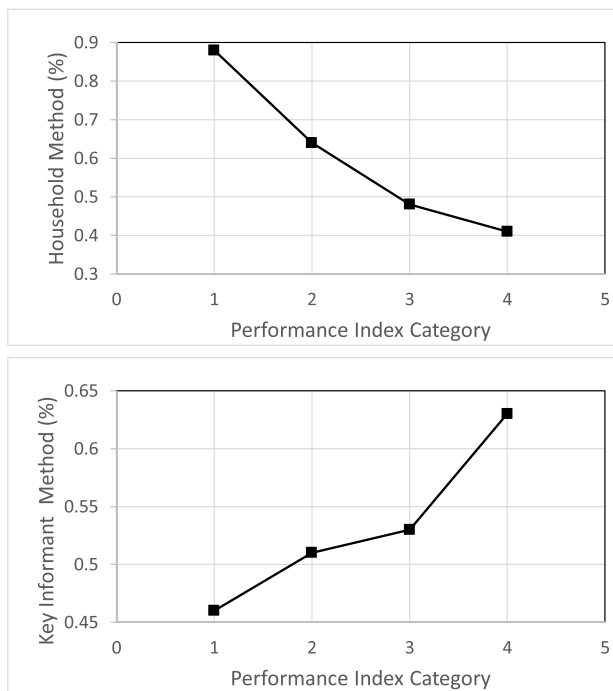
Fig. 4. Percentage (of total) of given input method used and performance index categories.

group 1 had generally high indices, sub-group 2 had low indices overall, and sub-group 3 was between the two extremes. Using this classification based on the four indicators we again looked at the percentage of data collection method used for a given source of resilience as done in Fig. 4. We again found the HH method to be related to low performance levels, especially if it was the only method used. We performed some additional forward and backward regression analysis similar to the GLM approach using the performance index as the dependent variable and the data collection method and number of methods as independent variables. The results of this further supported the findings outlined above.

Table 3 summarizes our findings visually (all details necessary to calculate the performance index for each source can be found in Supplementary VI). It shows the results for all 88 sources of resilience using green to indicate an overall strong relationship between raw data and grading (performance index above 0.6), and red to indicate a poor relationship (performance index at or below 0.6). As can be observed in Table 3, expert judgement is not related to any specific capital type or method used, confirming our findings from the category classification scheme and performance index calculation.

Summarizing, the results did not reveal an obvious structure between the raw data and grades for specific types of capitals or methods used; instead quite a complex picture emerged. In the next section we present six key findings from our analysis using the classification scheme and performance index results within the context of the specific types (discussed below) of raw data questions asked for a given source of resilience.

5. Discussion: six key messages

Our analysis revealed several important points which we summarize and discuss in the form of six key messages. Firstly, we find that the way in which the grade definitions are specified significantly influences the classification and performance index of the source. For example, sources with grade definitions defined as quantifiable categories, such as the aforementioned "Educational Attainment" (H16), do not need expert judgment but rather the right quantitative data, which in many cases is already available through third parties, to assign the grade.

However, secondly, we find that the first finding does not hold in all cases, especially where the raw data is more difficult to obtain and verify. Source "Income and Affordability" (F02) is a good example: the grade definitions relate to the percent of households that are able to afford their health, education and nutrition needs on a daily basis. This type of information is not readily available, and as such household surveys are required. Household surveys about expenditure are notoriously unreliable, hence more expert judgment is needed in cases such as this.

Thirdly, grade specifications based on subjective estimates and/or concepts usually result in low performance. For example, source of resilience "Social norms and personal security" (S05) contains the following (truncated) grade specifications: A: all people feel safe, B: most people feel safe, C: only some people feel safe, and D: people generally do not feel safe. The meaning of 'most' and 'only' are highly subjective and hence open to interpretation by the expert graders. Furthermore, the concept of 'feeling safe' is also highly subjective. Because judgement varies considerably between expert graders, sources that include subjective interpretations are less related to the raw data.

Fourthly, we find that the capital a source is assigned to does not influence the categorization and performance (see also Table 3). For example, the median of the performance index for the financial capital sources is 0.61, for human capital it is 0.71, for natural capital 0.62, for physical capital 0.73, and for social capital 0.73. In other words, it is not the type of capital and corresponding sources that have an impact on the performance index, but rather the way the grade definitions are specified and the ability to gather specific and reliable data to grade the source (key findings 1–3).

Fifthly, we find that the more quantitative the grade definitions are, the higher the performance index. For example, source of resilience "Functioning financial market" (F14) grading is based on the number of formal or informal institutions that households can access for savings and loans. It is a relatively straightforward process to gather data about the number of institutions, which then makes it relatively easier to assign the grade. In comparison, sources such as "Business credit access" (F05) are much harder to grade because data collection is more difficult. Furthermore, sources of resilience like F05 require extrapolation from a limited and questionably representative sample to the whole community. Hence, expert judgment plays a more significant role in such grading.

Sixthly and finally, relatively long and complicated grade definitions are related to poorer performance. For example, source "Conservation management plan" (N05) is related to biodiversity action plans and strategies, which is difficult to grade even when raw data is available (see also sources from the physical capital, such as P09). We want to emphasize again that the relationship between raw data and grades measured through the performance index is not an assessment of the value of the various sources of resilience. Indeed, many of the sources that scored poorly on the performance index are nonetheless essential for community flood resilience and must not be neglected. We discuss this issue in more detail in our last section.

6. Conclusions

Technical Risk Grading Standards (TRGS) are designed to help risk engineers make sense of the data they gather about the site they are assessing via a consistent benchmark against which to quantify risk. For resilience measurement such an approach has much potential because resilience is a multi-dimensional concept with many difficult-to-measure latent factors. The first generation FRMC as described here brings together quantitative and qualitative data about the attributes, resources and capacities that contribute to resilience (equivalent to risk factors in the TRGS approach). The FRMC grading approach is a two-step process, were first raw data is collected, and then experts grade the sources of resilience based on this data. The FRMC embodies a Technical Resilience Grading Standard (TResGS) that allows not only for

Table 3

Performance index and related sources of resilience questions. F indicates Financial Capital, H Human capital, P Physical capital, S Social capital and N Natural capital. Abbreviations for each source of resilience can be found in Supplementary I.

Source	INDEX	Source	INDEX	Source	INDEX	Source	INDEX	Source	INDEX
F01	0.82	H01	0.54	P01	0.80	S01	0.84	S21	0.73
F02	0.35	H02	0.71	P02	0.96	S02	0.73	S22	0.30
F03	0.73	H03	0.64	P03	0.79	S03	0.85	S23	0.23
F04	0.17	H04	0.58	P04	0.84	S04	0.65	S24	0.19
F05	0.28	H05	0.90	P05	0.64	S05	0.26	S25	0.59
F06	0.19	H06	0.25	P06	0.30	S06	0.89	S26	0.74
F07	0.61	H07	0.81	P07	0.30	S07	0.95	S27	0.95
F08	0.20	H08	0.91	P08	0.75	S08	0.61	S28	0.95
F09	0.34	H09	0.70	P09	0.44	S09	0.96	S29	0.75
F10	0.33	H10	0.91	P10	0.68	S10	0.47	S30	0.44
F11	0.72	H11	0.29	P11	0.63	S11	0.42	S31	0.90
F12	0.85	H12	0.77	P12	0.73	S12	0.79	S32	0.28
F13	0.58	H13	0.96	P13	0.74	S13	0.67	S33	0.78
F14	0.95	H14	0.16	P14	0.59	S14	0.73		
F15	0.44	H15	0.64	P15	0.95	S15	0.34		
F16	0.80	H16	0.93	P16	0.95	S16	0.46		
F17	0.81			N01	0.62	S17	0.61		
				N02	NA	S18	0.84		
				N03	0.86	S19	0.75		
				N04	0.87	S20	0.69		
				N05	0.54				
				N06	0.60				

the multidimensional factors contributing to community flood resilience to be assessed, but also helps identify actions for enhancing resilience.

Analysis of the 1.25 million datapoints collected across more than 118 communities in nine countries were complemented by content analysis to validate the results from a more qualitative perspective. Our findings regarding the role of data and grade specifications can inform ways forward for better, more efficient and robust standardized assessment of resilience. This also enables an analysis of the dynamics of resilience dimensions over time due to changes in the underlying capitals (Laurien et al., 2020). It therefore should help to build global standardized and comparable yet locally contextualized baseline estimates of the many facets of resilience, in order to track progress over time for a sustainable future as laid out in the Sendai Framework and the SDGs.

Overall, we find that for the FRMC – the TResGS – some sources of resilience are more easily graded by looking at raw data alone, and others require more expert judgement. We find that the relationship between the raw data and grade stability and predictability is not related to the capital group the source belongs to, as it is usually assumed in the literature. Instead, we find the most significant driver of stability and predictability, as well as performance, to be whether the grade definitions are quantified categories or are more qualitative and subjective. There are indications that the availability of high-quality raw data is also important for the stability and predictability of the grading. Furthermore, expert judgment plays a more significant role in instances where

there is a need to extrapolate the whole-of-community situation from a limited household survey sample, particularly when there are questions related to the representativeness of the sample. Finally, relatively long and complicated grade definitions are related to a smaller performance index, meaning a larger role for expert judgement.

It is critical to note that our assessment of the relationship between raw data and grades is not an assessment of the value of the various sources of resilience. Many of the sources that scored poorly on the performance index are nonetheless essential for community flood resilience. Indeed, the role of expert judgment is a fundamental principle of the TRGS and TResGS approaches. Risk engineering acknowledges that many important aspects cannot be assessed without expert judgment. This point reinforces the importance of comprehensive training and consistent guidelines for experts undertaking resilience assessments. In cases where the relationship between raw data and grades is straightforward, the role of expert judgement is less important, however, it is still a valuable check to ensure accurate grading. Our results and analysis identify important issues to be considered in the construction of resilience measurement frameworks. In particular regarding the role raw data plays in the assessment process, there is a need to carefully design indicators (sources of resilience in this case) that a) are clearly related to their raw data, and b) require raw data that is as accessible and reliable as possible.

The issue of raw data and grade definitions was also an important consideration raised in user feedback during the first testing phase of the

FRMC. In response to this feedback much attention was paid to the relationship between raw data and grade definitions, and the objectivity of grade definitions, in the design of the FRMC Next Gen 2 (with 44 sources of resilience instead of 88) which is currently being implemented across the globe.

CRediT authorship contribution statement

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.

Acknowledgement

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jenvman.2020.111332.

References

Asadzadeh, A., Kötter, T., Salehi, P., Birkmann, J., 2017. Operationalizing a concept: the systematic review of composite indicator building for measuring community disaster resilience. Int. J. Disaster Risk Reduct. 25, 147–162.

Birkmann, J., Cutter, S.L., Rothman, D.S., Welle, T., Garschagen, M., Van Ruijven, B., O'Neill, B., Preston, B.L., Kienberger, S., Cardona, O.D., Siagian, T., 2015. Scenarios for vulnerability: opportunities and constraints in the context of climate change and disaster risk. Climatic Change 133 (1), 53–68.

Bouwer, L.M., 2019. In: Mechler, R., Bouwer, L.M., Schinko, T., Surminski, S., Linnerooth-Bayer, J. (Eds.), Observed and Projected Impacts from Extreme Weather Events: Implications for Loss and Damage BT - Loss and Damage from Climate Change: Concepts, Methods and Policy Options. Springer International Publishing, Cham, pp. 63–82. https://doi.org/10.1007/978-3-319-72026-5_3.

Bruneau, M., 2006. Enhancing the resilience of communities against extreme events from an earthquake engineering perspective. J. Secur. Educ. 1, 159–167. https://doi.org/10.1300/J460v01n04_14.

Cai, H., Lam, N.S.N., Qiang, Y., Zou, L., Correll, R.M., Mihunov, V., 2018. A synthesis of disaster resilience measurement methods and indices. Int. J. Disaster Risk Reduct. 31, 844–855. https://doi.org/10.1016/J.IJDRR.2018.07.015.

Campbell, K.A., Laurien, F., Czajkowski, J., Keating, A., Hochrainer-Stigler, S., Montgomery, M., 2019. First insights from the Flood Resilience Measurement Tool: a large-scale community flood resilience analysis. Int. J. Disaster Risk Reduct. 40, 101257. https://doi.org/10.1016/j.ijdrr.2019.101257.

Cerè, G., Rezgui, Y., Zhao, W., 2017. Critical review of existing built environment resilience frameworks: directions for future research. Int. J. Disaster Risk Reduct. 25, 173–189. https://doi.org/10.1016/j.ijdrr.2017.09.018.

Chang, S.E., Yip, J.Z., de Jong, S.L.V.Z., Chaster, R., Lowcock, A., 2015. Using vulnerability indicators to develop resilience networks: a similarity approach. Nat. Hazards 78 (3), 1827–1841.

Chang, S.E., Yip, J.Z., Conger, T., Oulahen, G., Marteleira, M., 2018. Community vulnerability to coastal hazards: developing a typology for disaster risk reduction. Appl. Geogr. 91, 81–88.

Cimellaro, G.P., Reinhorn, A.M., Bruneau, M., 2010. Framework for analytical quantification of disaster resilience. Eng. Struct. 32, 3639–3649. https://doi.org/10.1016/j.engstruct.2010.08.008.

Constas, M., Barrett, C., 2013. Principles of resilience measurement for food insecurity: metrics, mechanisms, and implementation plans (No. 1). In: Expert Consultation on Resilience Measurement Related to Food Security.

CRED, UNISDR, 2018. Economic Losses, Poverty and Disasters 1998-2017, UNISDR (United Nations Office for Disaster Risk Reduction) Centre for Research on Epidemiology of Disasters (CRED).

Cutter, S.L., 2016. The landscape of disaster resilience indicators in the USA. Nat. Hazards 80, 741–758. https://doi.org/10.1007/s11069-015-1993-2.

Cutter, S.L., Ash, K.D., Emrich, C.T., 2014. The geographies of community disaster resilience. Global Environ. Change 29, 65–77. https://doi.org/10.1016/J.GLOENVCHA.2014.08.005.

DfID, 1999. Sustainable Livelihoods Guidance Sheets Introduction: Overview, Sustainable Livelihoods Guidance Sheets. https://doi.org/10.1002/smj.

Folke, C., Carpenter, S.R., Walker, B., Scheffer, M., Chapin, T., Rockström, J., 2010. Resilience thinking: integrating resilience, adaptability and transformability. Ecol. Soc. 15 (4) https://doi.org/10.5751/ES-03610-150420.

Gibbons, L.V., Pearthree, G., Cloutier, S.A., Ehlenz, M.M., 2020. The development, application, and refinement of a Regenerative Development Evaluation Tool and indicators. Ecol. Indicat. 108, 105698.

Google Inc, 2016. Storing and Analyzing Your Data in Google's Cloud.

Google Inc, 2012. The Google Cloud Adoption Framework.

Hák, T., Janoušková, S., Moldan, B., 2016. Sustainable Development Goals: a need for relevant indicators. Ecol. Indicat. 60, 565–573.

Hochrainer-Stigler, S., Mochizuki, J., Pflug, G., 2016. Impacts of global and climate change uncertainties for disaster risk projections: a case study on rainfall-induced flood risk in Bangladesh, 0 J. Extrem. Events 3, 1650004. https://doi.org/10.1142/S2345737616500044.

IPCC, 2018. IPCC special report 1.5 - summary for policymakers. In: Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P.R., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., Connors, S., Matthews, J.B.R., Chen, Y., Zhou, X., Gomis, M.I., Lonnoy, E., Maycock, T., Tignor, M., Waterfield, T. (Eds.), Global Warming of 1.5°C. An IPCC Special Report on the Impacts of Global Warming of 1.5°C above Pre-industrial Levels and Related Global Greenhouse Gas Emission Pathways, The Context of Strengthening the Global Response to the Threat of Climate Change. https://doi.org/10.1017/CBO9781107415324.

Keating, A., Campbell, K., Mechler, R., Magnuszewski, P., Mochizuki, J., Liu, W., Szoenyi, M., McQuistan, C., 2016. Disaster resilience: what it is and how it can engender a meaningful change in development policy. Dev. Pol. Rev. 35, 65–91. https://pure.iiasa.ac.at/id/eprint/11897/.

Keating, A., Campbell, K., Szoenyi, M., McQuistan, C., Nash, D., Burer, M., 2017. Development and testing of a community flood resilience measurement tool. Nat. Hazards Earth Syst. Sci. 17, 77–101. https://doi.org/10.5194/nhess-17-77-2017.

Keating, A., Hanger-Kopp, S., 2020. Practitioner perspectives of disaster resilience in international development. Int. J. Disaster Risk Reduct. 42 https://doi.org/10.1016/j.ijdrr.2019.101355.

Laurien, F., Hochrainer-Stigler, S., Keating, A., Campbell, K., Mechler, R., 2020. A typology for community flood resilience. Reg. Environ. Change 20 (1), 1–14.

Mechler, R., Bouwer, L.M., 2015. Understanding trends and projections of disaster losses and climate change: is vulnerability the missing link? Clim. Change 133 (1), 23–35. https://doi.org/10.1007/s10584-014-1141-0.

Meyer, V., Becker, N., Markantonis, V., Schwarze, R., Van Den Bergh, J.C.J.M., Bouwer, L.M., Bubeck, P., Ciavola, P., Genovese, E., Green, C., Hallegatte, S., Kreibich, H., Lequeux, Q., Logar, I., Papyrakis, E., Pfurtscheller, C., Poussin, J., Przyluski, V., Thieken, A.H., Viavattene, C., 2013. Review article: assessing the costs of natural hazards-state of the art and knowledge gaps. Nat. Hazards Earth Syst. Sci. 13 (5), 1351–1373. https://doi.org/10.5194/nhess-13-1351-2013.

Mochizuki, J., Keating, A., Liu, W., Hochrainer-Stigler, S., Mechler, R., 2018. An overdue alignment of risk and resilience? A conceptual contribution to community resilience. Disasters 42, 361–391. https://doi.org/10.1111/disa.12239.

MunichRe, 2018. NatCatSERVICE Loss Events Worldwide 1980–2017 [WWW Document]. Munich Reinsur. URL. https://natcatservice.munichre.com/. (Accessed 9 August 2018).

Pasteur, K., McQuistan, C., 2016. From Risk to Resilience: A Systems Approach to Building Long-Term, Adaptive Wellbeing for the Most Vulnerable. RPBS1. Rugby, UK.

Sato, K., 2012. An inside Look at Google BigQuery, White Paper. Google Inc.

Sherrieb, K., Norris, F.H., Galea, S., 2010. Measuring capacities for community resilience. Soc. Indicat. Res. 99, 227–247. https://doi.org/10.1007/s11205-010-9576-9.

SwissRe, 2018. Sigma explorer - catastrophe and insurance market data | Swiss Re Institute [WWW Document]. URL. https://www.sigma-explorer.com/. (Accessed 14 November 2019).

UN, 2015a. Sendai Framework for Disaster Risk Reduction 2015 - 2030 (United Nations).

UN, 2015b. Transforming Our World: the 2030 Agenda for Sustainable Development.

UNDRR, 2009 UNISDR terminology on disaster risk reduction. Int. Strat. Disaster Reduct. 1–30, 978-600-6937-11-3.

UNFCCC, 2015. Adoption of the Paris agreement. Report No. FCCC/CP. /2015/L.9/Rev.1.

Zurich Flood Resilience Alliance internal report, 2017. First Insights from the Flood Resilience Measurement Tool Baseline Data. Laxenburg, Austria.