Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

Natural Hazards and Earth System Sciences
Discussions

Vulnerability of bridges to scour: insights from an international expert elicitation workshop

Rob Lamb[1,2], Willy Aspinall[3,4], Henry Odbert[4], Thorsten Wagener[5,6]

[1]JBA Trust, South Barn, Skipton, BD23 3AE, UK
[2]Lancaster Environment Centre, Lancaster University, LA1 4YQ, UK
[3]Aspinall and Associates, UK
[4]School of Earth Sciences, University of Bristol, Bristol, UK
[5]Civil Engineering, University of Bristol, UK
[6]Cabot Institute, University of Bristol, UK

Correspondence to: Rob Lamb (rob.lamb@jbatrust.org)

Abstract. Scour (localised erosion) during flood events is one of the most important threats to bridges over rivers and estuaries, and has been the cause of numerous bridge failures, with damaging consequences. Mitigation of the risk of bridges being damaged by scour is therefore important for many infrastructure owners, and is supported by industry guidance. Even after mitigation, some residual risk remains, though its extent is difficult to quantify because of the uncertainties inherent in the prediction of scour and the assessment of the scour risk. This paper summarises findings of an international expert workshop on bridge scour risk assessment exploring uncertainties about the vulnerability of bridges to scour. Two specialised structured elicitation methods were applied to explore the factors that experts in the field consider important in assessing scour risk, and to derive pooled expert judgements of bridge failure probabilities conditional on a range of assumed scenarios describing flood event severity, bridge and watercourse types and risk mitigation protocols. The experts' judgements broadly align with industry good practice, but indicate significant uncertainty about quantitative estimates of bridge failure probabilities, reflecting the difficulty in assessing the residual risk of failure. The data and findings presented here could provide useful context for the development of generic scour fragility models, and their associated uncertainties.

1. Introduction

This paper summarises the outcomes of an international expert elicitation workshop on bridge scour risk assessment held in London in 2015. The workshop brought together 17 experts from the UK, USA, New Zealand and Canada, including representatives from industry, academic researchers, and public agencies. Our motivation was to explore, in quantitative terms, uncertainties about the vulnerability of bridges to scour, with the ultimate aim being to inform the development of fragility functions that may be applied within a broad scale risk modelling framework (where "broad scale" indicates modelling over an extensive network of assets rather than detailed, site-specific risk assessment, for example see Hall et al., 2016).

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

Natural Hazards and Earth System Sciences Discussions

Scour refers to localised erosion that can undermine the foundations of bridges where they cross water. It is associated with high flows around the bridge piers, abutments and surrounding channel reaches, especially during flood events. The loss of support and consequent foundation movement caused by scour can result in costly damage to the structure, service restrictions, and perhaps more importantly compromised safety for users of a bridge. In extreme cases the bridge structure may collapse. A critical threat to infrastructure around the world, scour is cited as the most common cause of bridge failure (Kirby et al., 2015); its importance is discussed further in Sect. 2 below.

Scour risk is managed through the application of assessment, monitoring and maintenance protocols, which are reviewed in Sect. 2. These protocols are undoubtedly effective in reducing risk by prioritising scour protection works, helping to spot incipient problems and triggering maintenance or other mitigation actions when needed. Even so, the evidence of occasional scour-related bridge failures indicates that some residual risk remains. This residual risk is difficult to manage, representing as it does a combination of rare events and uncertainties about the actual (as opposed to designed) response of assets to flooding. A generic framework for assessing the risk in terms of uncertain failure probabilities is outlined in Sect. 3.

The combination of infrequent natural drivers in the form of flood events, complex physical processes, and the difficulties, costs and uncertainties associated with measurements mean that it is difficult to quantify scour risk with confidence and, in particular, to extrapolate from historical or experimental evidence to more extreme situations. In these circumstances, the knowledge and judgement of experts constitutes an especially valuable source of information that can be harnessed to augment data from other sources. A formal process of elicitation was applied to develop a synthesis of current knowledge from expert judgements.

The elicitation methodology, described in Sect. 4, was a two-stage process. In the first stage, a categorical approach was used to examine which factors determine the likelihood of scour at a bridge, and how experts think those factors should be ranked in importance. The second stage involved a quantitative assessment of bridge failure probabilities for a range of plausible scenarios under stated conditions and assumptions. The elicitation techniques included methods to weight information from the group of experts so as to promote the most accurate and unbiased judgement of uncertainty, using control questions to 'calibrate', jointly, the statistical accuracy and informativeness of the experts' uncertainty judgements. These are traits that can differ – sometimes substantially – from one expert to another, and can be adjusted for by empirical scoring rules to generate an optimal group decision. Results of the elicitation are presented in Sect. 5 and discussed in Sect. 6, highlighting both implications for scour risk management and methodological conclusions relating to the process of expert elicitation.

2. Motivation

Scour is well-known to be an important hazard. A survey of notable bridge failures around the world by Smith (1976) found that almost half were associated with "flood and foundation movement", including collapses at many different types of bridges. In the US, scour is thought to be the most common cause of highway bridge failures (Kattell and Eriksson, 1998).

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

Natural Hazards and Earth System Sciences Discussions

Using the US National Bridge Inventory, Cook (2014) also found the most likely cause of bridge collapses to be "hydraulic in nature", mostly scour, and determined that collapses caused by hydraulic factors were not related to the age of the bridge.

In the UK, on the rail network alone, more than 100 bridge collapses since 1843 have been attributed to scour in rivers and estuaries, causing 15 fatalities (Rail Safety and Standards Board, 2005, van Leuwen and Lamb, 2014). Recent cases include the collapse at Glanrhyd, Wales, in 1987, which led to the deaths of four people when part of a passenger train fell into the River Towy, and the failure of the Lower Ashenbottom viaduct in Lancashire, in June 2002. During the 2009 floods in Cumbria, UK, seven road and foot bridges failed due to a combination of scour and hydrodynamic loading, with the collapse of the Northside road bridge in Workington causing one fatality and significant disruption to communities. More recently 131 bridges were damaged during flooding in the same region, many because of scour (Cumbia County Council, 2016, Zurich Insurance Group and JBA Trust, 2016).

For UK rail bridges, the known bridge failures evince issues and uncertainties associated with assessment of scour risk, for example suggesting that some historical failures occurred after relatively minor flood events rather than extreme floods, perhaps because, prior to the introduction of modern scour management practices, there is more likely to have been undetected scour damage during a sequence of events that ultimately led to failure. Some uncertainties relate to data errors or missing information. However, the complexity of physical scour processes also leads to uncertainty in scour models. This complexity includes some inherently unpredictable factors such as the occurrence and severity of flood flows, and the accumulation of debris, which can amplify scour through additional turbulence and enhanced local flow velocities.

2.1. Risk-based management and industry practice

In the UK and many other countries, bridges are designed, inspected and maintained so as to withstand damage during events that are "reasonably foreseeable" over their intended service life (UK Roads Liaison Group, 2009). As with many infrastructure assets, there is a balance to be struck between the costs of reducing the risk of scour, likely damage, and expectations of public safety. Design guides, monitoring, inspections and detailed modelling all help to establish the level of resources needed to achieve an appropriate balance, noting that the question of what is "appropriate" is ultimately a matter of judgement and policy.

Risk-based asset management concepts are widely applied to help inform these judgements. A risk assessment involves considering the outcomes that could result from a combination of drivers, such as extreme weather events, and the performance of assets when subjected to those events. Kirby et al. (2015) and Arneson et al. (2012) give comprehensive guidance for scour risk management, including references to numerous industry and government agency sour management protocols, including the UK Design Manual for Roads and Bridges (Highways England, 2016), US National Bridge Inspection Standards (FHWA, 2016), and US Forest Service scour assessment process (Kattell and Eriksson, 1998).

Scour risk management guidance typically deals with uncertainty through a combination of quantitative and qualitative analysis within a tiered structure, where relatively inexpensive, rapid "high level" screening is used to prioritise further investment of resources for more detailed assessments at bridges where scour may be more likely to occur, or where its

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

Natural Hazards and Earth System Sciences Discussions

consequences may be worst. Multiple factors are typically considered at each level within a tiered assessment, including physical characteristics of the bridge structures, the watercourses that they cross, their wider flow and sediment regimes and historical observations or recent changes relating to scour.

Most guidance involves some probabilistic analysis, which is usually introduced through the estimation of potential scour depth for an assumed "design flood", specified in terms of an annual exceedance probability (AEP, or equivalently a return period, which, when expressed in units of years, is numerically equal to 1/AEP). The design flood scour estimate can be compared with an estimated or known foundation depth to calculate a risk score. Recommended design flood conditions for UK railway and road bridge scour assessments are 1/200 AEP. In the USA the design condition is typically 1/100 AEP, but with a margin of safety that the structure should not fail in a 1/500 AEP event (Kirby et al., 2015). The probabilistic analysis is one part of the broader scour risk assessment protocols set out in industry guidance.

In this study the objective is to focus on uncertainties and their role in the probabilistic analysis of scour. In contrast to a design event analysis, we seek to explore how uncertainty about scour risk could be captured through a generic fragility model for bridge failure probability, reflecting a range of loading conditions, and including possible increases in vulnerability following exposure to flooding.

We ask explicitly how a general probabilistic failure model of this type could be formulated. The underlying motivation is an interest in generalising from detailed understanding of scour at specific bridges to consider the risks aggregated over a network or portfolio of assets, to support analysis either for a "generic bridge", or in a distributed, network-scale model of risk. The former case represents situations in which there may be inadequate information to carry out a detailed risk assessment. The latter is important in the context of strategic decisions about future planning, investment and operations for various infrastructure systems (e.g. Hall et al., 2016). This type of generalisation may not be appropriate for application to engineering decisions at individual assets, but is relevant as part of the higher-level risk screening that forms one tier in scour management approaches applied in practice.

3. Scour risk analysis framework

In the case of scour risk, the underlying hazard events are flood flows to which bridges and their foundations may be vulnerable. The drivers are uncertain because of the apparently stochastic nature of flood events, meaning that it is not known for certain whether a flood of some given level of severity or extremeness will be encountered during the design life of the bridge, or indeed in any specified period of time. Compounding this, it is not certain that an asset will perform as intended in response to any particular event or sequence of events, especially when conditions exceed design specifications. Indeed, for assets of unknown age and origin there may be no applicable specification for the design, although retrospective assessments and structural improvements can be, and often are, made.

To assess the risk associated with scour thus requires an understanding of the type of events that could plausibly occur and how an asset might respond to them. Although there could be many ways to do this, we argue that a powerful and general

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

Natural Hazards and Earth System Sciences Discussions

approach is, if possible, to treat the flood hazard and the asset performance in terms of probabilities, which allows the risk assessment to be framed ultimately in terms of a probability distribution of outcomes.

A high-level conceptual risk model for bridge failure from scour is outlined in Fig. 1, where the processes that create the flood hazard are described in terms of the probability distribution of some relevant load variable, and the response of the bridge is described by a fragility function, representing the probability of a failure occurring conditional on an assumed load level. Figure 1 maps directly onto well-established, generic risk modelling frameworks, including the source – pathway – receptor concept widely used in environmental risk assessment (Defra, 2011), the loading and fragility concepts of reliability analysis (Ellingwood, 2008, USACE, 2010) and the hazard – vulnerability – loss concepts often applied in natural hazard risk assessments for insurance.

The scour risk can be expressed in generic terms via the distribution function $F[Y(L,S)]$ of possible outcomes Y when a bridge is subjected to some load representing the source of the scour hazard, where L is a random variable describing the relevant loading condition(s) and S is a state variable that is used to describe the uncertain response of a bridge under a given load (e.g. $S = 1$ if the bridge "fails" due to scour and $S = 0$ otherwise). The distribution function $G(l) = \Pr[S = 1 \mid L \leq l]$ is the probability of failure conditional on a load event $L = l$. At this point no precise definition of loading condition or failure is offered. Failure could legitimately be defined as catastrophic collapse of the bridge, or in terms of a failure to continue providing some specified level of service (e.g. safe passage for traffic). The function $G(l)$ can be called a fragility function or vulnerability function, and is central to this analysis.

Our aim here is to inform the development of suitable descriptions of scour vulnerability by investigating two questions:

1) What variables should be chosen to describe the loading conditions relevant to scour risk?
2) What are the failure probabilities associated with a range of possible loading conditions, and how uncertain are they?

For an asset-specific model there may be an obvious loading condition, such as flood water level at the bridge, together with detailed data or models to help predict the performance of the structure. In a more general analysis, the definition of the relevant load condition is not necessarily clear because the factors that matter most may vary from asset to asset. Whilst this study does not progress to a full description of fragility functions, the results may help to inform their development by informing the choice of relevant loading conditions and providing a pooled expert assessment of failure probabilities and associated uncertainties.

4. The role of expert elicitation

Both of the questions posed above could be tackled through empirical analysis or modelling of data for specific bridges. Deterministic models exist to predict the scour depths at structures for prescribed conditions including equilibrium scour,

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

Natural Hazards and Earth System Sciences Discussions

(Melville, 1997) and time-varying scour (Melville and Chiew, 1999, Coleman et al., 2003). In general, scour prediction models are based either on physical hydraulic formulae with coefficients calibrated form laboratory or field data, or may have a statistical basis (e.g. Hong et al., 2012). Model coefficients inferred from empirical observations inevitably carry some uncertainty, which can be expected to increase and may be difficult to quantify, when generalising beyond the sample or type of structure used in the original inference.

In assessing the risk of scour failure over a broad network of assets and over an arbitrary time period, deterministic models for scour also need to be combined with analysis of the frequency or probability of hazardous flood events, (see, for example Decò and Frangopol 2011), introducing further uncertainty inherent in the assessment of extremes. For a broad scale analysis some significant sources of uncertainty therefore remain that reflect the unpredictability of any given asset's actual performance under a range of conditions, and the generalisation from specific cases to generic classes of structure for use in broader-scale risk analysis.

Inevitably, uncertainty has a major influence on a risk assessment and on any associated decisions in circumstances such as this where rare events are being considered. In these situations, there may be a need to appeal to the judgement and advice of experts, and some subjectivity is inevitable in the interpretation of terminology and data.

Soliciting expert advice for decision support is not new. Often it has been pursued on an informal basis. In this study, a structured approach has been taken to elicit expert judgements from a range of opinions such that a rational consensus emerges about appropriate levels of uncertainty to be used in risk analysis. The formalised elicitation methodologies we adopted are designed to tie results into stated and transparent methodological rules, with the goal of treating expert judgements in the same way as other scientific data in a formal decision process. Various methods for assessing and combining expert uncertainty have been described in the literature. Until recently, the most familiar approach has been one that advocates a group decision-conferencing framework for eliciting opinions, but other approaches now exist for carrying out this process more objectively. Two methods were selected for this study, corresponding to the two motivating questions discussed above:

1) Expert judgement on choice of variables to describe the loading conditions in scour vulnerability analysis:

   A specialised variant of the survey method of paired comparison was selected to assess judgements about the relative importance of factors that control vulnerability to scour. Initially, the method involves presenting a list of items and asking each expert to express a preference or importance ranking for every pairwise combination of the items. Then, a unique probabilistic inversion technique is used to reveal the overall preference ordering of the items, both for each expert and for the group, along with a numerical assessment of the logical coherence of the responses in terms of 'circular triads' in the experts' responses (i.e. if item A is ranked above item B by an expert, and B is ranked above C, then C should not be ranked above A). The software tool UNIBALANCE (Macutkiewicz and Cooke, 2006) was used to process experts' preferences as individuals and as a group, to construct a formal

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

Natural Hazards and Earth System Sciences Discussions

probabilistic group representation of the alternative views expressed through the paired comparison elicitation. The UNIBALANCE analysis output provides objective measures of confidence about the extent to which the experts believe it is possible to discriminate between alternative factors.

2) Failure probabilities associated with a range of possible loading conditions, and associated uncertainties:

For uncertainty quantification, a structured expert judgment procedure formulated by Cooke (1991), known as the "Classical Model", was adopted in this study. This approach is supported by a software package called EXCALIBUR (Cooke and Solomatine, 1992). This is a quantitative elicitation method used to assess numerical estimates of uncertain parameters or variables, in this case scour failure probabilities conditional on various stated assumptions.

The Classical Model approach has been extensively used elsewhere in natural hazards risk assessments (e.g. Bamber and Aspinall, 2013, Aspinall and Cooke, 2013) and in many other uncertainty-related problem areas (e.g. Cooke and Goossens, 2008). The method was evaluated in detail by Aspinall et al. (2016) in the context of a global mega-elicitation for the World Health Organization.

In similar vein to the present study, an elicitation using paired comparison probabilistic inversion jointly with uncertainties elicited with the Classical Model was reported by Tyshenko et al. (2011) for an elicitation for prion disease risk, but the combination of these methods has not to our knowledge previously been documented in the natural hazards or civil engineering literature.

5. Results

5.1 Question (1): Vulnerability factors that should be considered in assessing risk of scour

In the first stage of the elicitation, and following some discussion of issues and available information, the experts in the group were asked to complete a series of paired comparisons structured around the following question: "What are the most important factors that should be considered in assessing scour risk to bridges?". In each case the probabilistic inversion technique was used to calculate a group score and associated uncertainty.

The factors to be ranked were proposed by the project team and amended following an initial group discussion at the workshop (Table 1). The initial discussion raised concerns that the risk assessment priorities for piers and abutments may differ. The analysis was therefore carried out twice, treating scour at bridge piers and scour at abutments as separate issues. Results are shown in Table 2 for each of the potential assessment factors, and in Fig. 2, where the scores for scour risk to abutments and piers can be compared.

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

Natural Hazards and Earth System Sciences Discussions

In the expert group's view, the most important factors in assessing vulnerability to scour (though with only weak discrimination between the factors) were as follows:

1) Scour history, i.e. whether or not scour has been a problem in the past
2) The morphological regime in the watercourse, including removal of sediments and morphological instability
3) Characteristics of bridge structure, including foundation type and depth, and the degree to which the flow is constricted at the bridge
4) The existence of inspection and scour assessment policy, and existence of prior scour protection
5) Watercourse characteristics or changes that may be unpredictable (e.g. debris accumulation) or cause progressive change in vulnerability (e.g. weir removal), but may be detectable in time to intervene during or between flood events
6) Uncertainty in knowledge about the foundations
7) Attributes of the bridge structure other than the foundations and constriction of the flow (e.g. bridge type, bridge span, construction date)
8) Recent flood history

Generally, factors ranked as important in determining the risk of abutment scour were also ranked as similarly important for scour at piers. The presence of an oblique approach flow was considered markedly more important for scour at piers than at abutments, although of less importance than other factors considered, in both cases.

5.1.1 Definition of loading conditions for fragility functions

Further discussion led to a refined set of factors that might be proposed to define relevant loading conditions for a scour fragility function. The experts were asked to rank this list in order of relevance. Overall, the ranking scores (Table 3) are quite compressed, ranging from 0.31 for the existence of a 'Scour assessment procedure', implying this was judged to be of relatively low importance amongst the list of factors proposed for determining bridge vulnerability, to 0.65 for 'Frequency and amount of debris', which was of greatest concern. The factors appear subjectively to separate into three clusters of differing importance, comprising two, three and five factors, respectively, labelled A, B and C in Table 3. The uncertainty about the rank order is broadly consistent for all factors.

Five factors appear to emerge as a preferred group from which the load variable in a fragility function might be defined. One is related to debris load. The others relate to hydraulic conditions during a flood event, including flood flow, flood flow return period, flow velocity and also duration of high flow.

Flood flow, velocity and flood return period may be intrinsically linked. However, the return period, or, alternatively, exceedance probability is a more abstract measure of a load event's intensity that is open to interpretation with respect to the choice of methods applied to define a "flood event" and estimate its probability. In contrast with the physical parameters more usually considered for asset-specific scour assessments, a probabilistic definition of loading may be viewed as a standardised measure of the load intensity defined on a common scale (e.g. the annual exceedance probability or return

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

period in years) regardless of the actual physical scale of the system (e.g. channel width and depth, typical flow rates, or upstream catchment area). The results also suggest there is value in further investigation of the role of event durations within scour fragility analysis and the possibility that sequences or clusters of high flow events may also be important, although it may be more complicated to incorporate these temporal factors within a fragility function.

5.1.2 Potential changes in scour vulnerability

Finally, the expert group was asked to consider factors judged to be important in determining how the risk of failure may change under different circumstances. The factors discussed, and the group's ranking of them, are in Table 4. In this case, there is greater spread in factor rankings, suggesting that the expert group was clearer about discriminating between factors that could be used to determine how scour risk may change. Change in inspection regime was identified as the most important factor.

Climate change did not emerge as an important consideration in the ranking scores. Post-hoc discussion with some members of the expert panel showed that the factor labelled "Climate Change affects frequent extreme rainfall" was interpreted variously as meaning "the impacts of climate change on failure risk in the next few years" or "the impacts on risk in the long term". In either case, detailed feedback suggests that there may be important contextual differences in relation to this question. In the USA, a typical bridge design standard may be based on a 1/100 annual probability storm, but with an expectation of withstanding a more extreme storm of 1/500 annual probability. Hence even if climate change projections point to an increase in storm severity, the factor of safety allows for some confidence that the bridge scour risk is not unacceptably increased. This remark was made in the context of a typical service life of 75 years, with a re-evaluation of the required design being planned at that point, in effect allowing for a degree of planned adaptation. One of the US experts observed that the UK experts may not be able to assume a specified design standard for older bridges, especially if their foundation depths are not known precisely, and therefore may be more sensitive to the risk of increased flooding in a changing climate.

The discussion above brings out some ambiguities within the group's pooled responses owing to different assumptions made by participants from different countries about terminology and design standards.

5.2 Question (2): Quantitative elicitation of failure probabilities, with uncertainties

In the quantitative elicitation, the expert group was asked to estimate bridge failure probabilities, associated with scour caused by flooding under a range of conditions. In each case, the experts were asked for lower, central and upper values, corresponding to their judgements about the 5^{th}, 50^{th} and 95^{th} percentiles of the range within which the true failure probability lies. The individual responses were pooled, with and without weighting, using the Classical Model (Cooke, 1991).

The failure probabilities were requested under various different conditions relating to: flood return period; type of watercourse; type of bridge foundations; type of monitoring/inspection and maintenance policy in force ("maintenance").

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

The definitions of generic types of watercourse, foundation and "maintenance" regime generated lengthy debate, primarily reflecting geographical differences in emphasis between the UK and North American experts. The following definitions were eventually adopted as a working compromise with the general assent of the group. The group agreed to have to in mind physiographic and climatic conditions typical of the UK context, i.e. predominantly a humid temperate climate and a mixture of upland and lowland rivers, and to exclude more extreme (by UK standards) environments such as large continental scale rivers, Alpine rivers or rivers flowing in arid regions.

Two generic types of watercourse were specified: 1) Unmanaged watercourse – no channel or upstream measures specifically designed to reduce scour risk (such as active vegetation management to reduce risk of debris or promote sediment stability); 2) Managed watercourse – actively managed to control or reduce scour risk (or for other primary purposes also serving to reduce scour risk).

Two generic foundation types were specified: 1) Shallow foundations – a class including some historical masonry structures in the UK, particularly in lowland rivers, where foundations may be shallow pads or piles; 2) Deep/bedrock – a class that would include modern deep piles and also historical structures build directly onto solid bedrock, for example some UK bridges over upland rivers.

Three potential asset management regimes were specified, one of which relates to current practice: 1) None – a counterfactual assumption (at least for UK, North America and regions with rigorous engineering codes) of no investment of resources in monitoring, inspection or maintenance of scour protection maintenance works; 2) Routine – an investment of resources roughly similar to present-day good practice in the UK, US, Canada or New Zealand; 3) Premium – a counterfactual and significantly enhanced level of investment in inspection, monitoring and maintenance, featuring pro-active, highly precautionary investments in maintenance and scour protection.

After much discussion, the workshop group settled on a definition of "failure" as damage caused by the flood event to the structure, foundations or approaches, probably due to scour, sufficient to: cause a threat to safety; disrupt service and require repair action; cause collapse or would cause collapse if left unattended. (Note that this is a less restrictive definition of failure than one in which only a catastrophic collapse of the structure would be considered.)

5.2.1 Guide to interpreting the results

Results of the elicitation are plotted in Fig. 3-5. In each case, the bars represent the range of the 5th to 95th percentile estimates pooled from the expert group. The bold lines and symbols are the result of pooling the experts' estimates with weightings applied based on the performance of each individual in assessing uncertainty through the calibration questions. The lighter grey lines and symbols are the equivalent estimates, but this time combined with equal weight afforded to each expert. Results have been plotted on a logarithmic scale because in some cases the estimated probability ranges cover several orders of magnitude.

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

Natural Hazards and Earth System Sciences Discussions

5.2.2 Event failure probabilities (fragility estimates)

The pooled estimates of failure probabilities (Fig. 3) tend, as expected, to increase as the intensity of the flood event increases. The failure probabilities also appear to decrease with improving maintenance regime.

Differences in the central estimates of failure probability with respect to flood event return period, maintenance assumption or watercourse/foundation type are generally rather smaller than the uncertainty ranges associated with the estimates. Note that the ranges are quantile estimates and not associated with any prescribed error distribution. Clearly the expert group's assessment of uncertainty is to place wide margins on any fragility estimate. Indeed, it would be surprising if this were not the case, given the nature of the problem as posed.

Although set against a wide range of uncertainty, the estimates of failure probability appear to increase systematically as flood event return period increases, and in line with expectations if comparing an obviously more resilient scenario (e.g. bridge with deep/bedrock foundations and "premium" maintenance) with a more vulnerable one (e.g. a bridge with shallow foundations and no maintenance).

Different assumptions about the foundation/watercourse type seem to cause large variation in the estimates of the upper uncertainty bounds under no maintenance or routine maintenance, particularly for the more extreme flood events (100-year and 500-year return period).

In comparison with an equally-weighted group estimate, the performance-weighted estimates display more constrained uncertainty. In particular, this is marked for the 100-year flood event results, where the application of weighting conditioned on the calibration questions results in a much lower pooled estimate of the upper quantile (95^{th} percentile) on failure probability. Other than for the managed, deep/bedrock case, this "calibration" of the upper failure probability bounds is not accompanied by a downward shift in the lower bounds. For the more extreme, 500-year return period flood, the weighting against performance on calibration questions makes little difference; this would suggest that although accounting for individual experts' skill in assessing uncertainty may help to refine group judgements about moderate failure probabilities, it does not constrain the very wide uncertainty in judgements about failure probability under very extreme flood conditions.

5.2.3 Annual failure probabilities

The experts were also asked to give ranges for their estimates of the annual probability of failure, again considering the three notional "maintenance" regimes and the four foundation and watercourse types.

The results (Fig. 4) follow expected patterns in that larger failure probabilities were estimated for the shallow foundation cases than for deep foundations, estimated failure probabilities were higher for an unmanaged watercourse than a managed watercourse, and estimated failure probabilities decrease as the assumed maintenance regime improves

The overall effect of applying performance weighting, based on calibration questions, has been to constrain the ranges of uncertainty without causing marked changes in the central estimates of failure probability for most cases. It is interesting to note that this performance-weighted modulation of elicited ranges is much more pronounced for the cases that describe

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

Natural Hazards and Earth System Sciences Discussions

inherently more resilient bridges (i.e. deep/bedrock foundations). An implication is that pooled estimates based on performance-weighted judgements appears to have resulted in a rather less precautionary judgement about uncertainty for the most resilient asset types.

Clearly the question, as it was posed, required the experts to make some general assumptions, either implicitly or explicitly, about the probability distribution of flood flows at a bridge and actual or inferred design standards. This lack of specific context for the calculation may account for some of the uncertainty expressed by the experts. Some discussion was held about whether the annual failure probability is in fact determined completely by design standards and the statistical distribution of floods, although in the UK this position would not correspond with evidence of railway bridge failures that have occurred under a wide range of conditions (van Leuwen and Lamb, 2014), suggesting that it may not be appropriate to treat vulnerability as a deterministic function.

5.2.4 Conditional event failure probabilities

The experts were asked to consider conditional failure probabilities for a generic bridge (defined below), when subjected to flood conditions of different levels of severity, conditional on the assumption that a preceding 100-year return period flood had already occurred, and with no intervening maintenance. The term "generic bridge" was taken to mean that variations in foundation, river characteristics or maintenance protocols were to be included as part of the uncertainty in the estimates. Pooled responses are shown in Fig. 5.

The pooled central estimates correlate with the severity of the flood event, as expected. For an extreme 1000-year event, the central estimate of the group is that there is more than a 50% chance of failure. However, the ranges express what is essentially a position of complete uncertainty about the most pessimistic (i.e. upper bound) judgement about the failure probability uncertainty, with the performance weighted group estimates differing little from the equally weighted estimates.

It can be seen that in the judgement of the group, the likelihood of a failure under extreme conditions of a sequence of 100-year flood followed by 1000-year flood is at least 1%. This is about 10,000 times more likely than the most optimistic pooled judgement made about failure probability for a minor, 5-year flood following after the 100-year event.

5.2.5 Triggers for asset inspection

As a supplementary question, experts were asked to make a judgement about a threshold flood return period that should trigger a new inspection. The pooled responses, shown in Table 5, indicate that the experts envisage a long upper tail in their judgement of uncertainty about a trigger threshold defined in this way. All experts express some belief within the elicited uncertainty (5^{th} to 95^{th} percentile estimates) that an inspection trigger based on flood rarity could possibly be encountered with a probability of close to 1.0 in any given year (return period ≈ 1 year). When pooled with equal weights the group median response was to suggest an inspection threshold at a flood of 1-in-26 annual exceedance probability, and that the inspection threshold might (at an upper, 95^{th} percentile, limit of uncertainty) be set as high as once in 318 years. This upper limit would indicate a considerably more relaxed inspection criterion than scour assessment protocols in use today. However,

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

when the pooled response is weighted according to the experts' judgement of uncertainties during the calibration exercise, the assessments become much more precautionary, with a median response that inspections be triggered by any flood of 1/5.6 annual exceedance probability, and, interpreting the 95th percentile estimate in terms of a long-run average frequency, with an annual probability of inspection being triggered of at least 1-in-48.

6. Discussion and conclusions

The elicitation workshop has provided to the authors' knowledge the first formal, pooled assessment of experts' judgements and uncertainties about scour risk. It has helped to provide a rational ordering of factors that could be considered in designing scour vulnerability assessment protocols and risk analysis models.

The heterogeneity of river environments, bridge types and engineering approaches makes it very difficult to specify a generic fragility model. However, despite these challenges, the group succeeded in reaching workable compromises about generic descriptions of bridges, maintenance regimes and risk factors that could be used, for the purposes set out in Sect. 2, in a quantitative fragility model.

After carefully debating the definition of terms, the group's input to a structured elicitation process enabled pooled estimates of scour fragility to be derived, expressed as the probability of a bridge failure conditional on flood events of varying severity, where this severity was also expressed in probabilistic terms. From this information, it may be possible to construct scour fragility functions.

Our conclusions are summarised below in four parts, relating to: the identification of factors considered important in determining the vulnerability of bridges to scour (Sect. 6.1); failure probabilities and associated uncertainties (Sect. 6.2); methodological considerations regarding the elicitation process (Sect. 6.3); and how the findings relate to current industry guidance on scour management (Sect. 6.4).

6.1 Choice of factors for scour vulnerability assessments

The findings of the workshop were well-aligned with current industry guidance on scour assessment, highlighting the importance of foundation depth, scour depth (either measured or predicted from modelling), river typology (i.e. whether a steep channel or lowland watercourse) and foundation material (e.g. clay, rock or of unknown type), which are all taken into consideration.

Additionally, the expert group identified other factors that are potentially important in assessing scour risk and that might be given greater emphasis in risk assessment guidance. These factors highlight the potential influence of changes to a watercourse at and around a bridge: dredging or sand/gravel extraction; removal of weirs near bridge; and influence of flood defences.

The group also highlighted the importance of inspection and assessment regimes (i.e. the level of resources committed to scour monitoring and assessment, or changes in that commitment) in controlling the risk posed by bridge scour.

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

Risk factors relating to hydraulic conditions during flood events (flood flow magnitude, duration, and flow velocities around the structure) and morphological regime (dredging) were consistently ranked by the group as important in determining scour vulnerability, although there was considerable ambiguity about the relative importance of many other factors, supporting the application of multi-factorial approaches to risk assessment.

In addition to variables expressed on physical scales, the return period (or exceedance probability) of a flood event was identified as a possible approach to define a generic loading condition for the development of bridge scour fragility functions. Fragility functions are not incorporated into routine scour management guidance. The data presented here could be used to give some context to functions of this type should there be future work to develop reliability analysis models based on fragility concepts.

6.2 Expert views on scour failure probabilities and associated uncertainties

Experts' estimates of failure probability appear to increase systematically as the assumed loading, i.e. flood event severity, increases. Their failure probability estimates also differ, as might be expected, with respect to assumed differences in vulnerability relating to bridge foundation type, watercourse characteristics and the amount of resource committed to inspection and maintenance

Expert judgements about fragility for any given bridge during a relatively modest flood event of 25-year return period indicated failure probabilities of around 1% or smaller, with uncertainties ranging from around 0.01% up to a few percent.

For an extreme flood with a 500-year return period, experts' central estimates suggest that a well-maintained bridge in a morphologically stable channel with modern or bedrock foundations has less than a 20% chance of failing due to scour, rising to nearer 50% for a poorly maintained bridge, or a bridge in an unstable channel on weak foundations; however uncertainty about these estimates is very wide, with experts judging that the true chance of failure could conceivably be less than 1% or nearly 95%.

Different assumptions about the foundations and watercourse type led to large variations in estimates of the uncertainty about failure probabilities under assumptions of no maintenance or routine (i.e. "business-as-usual") maintenance, particularly for the more extreme flood events (100-year and 500-year return periods)

Subjectively wide uncertainties were indicated in the group fragility estimates, reflecting a combination of differences in interpretation and, as revealed through calibration questions, differences between experts in their inherent assessments of uncertainties.

These results are not replacements for modelled or empirically-derived estimates of vulnerability. Rather, they add a view of broader uncertainties that are not easily captured in models or engineering formulae, and include subjective interpretations and judgements. In this sense the results help to paint a more complete picture of uncertainty about scour risk and to highlight the continuing need for monitoring, and research, to constrain uncertainties about scour risk.

Increasing assumed levels of resourcing for monitoring and scour assessment translated into reductions in the experts' estimates of annual or flood-event failure probabilities, but these reductions were small relative to the experts' overall

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

Natural Hazards and Earth System Sciences Discussions

judgements of uncertainty, which were affected very little by those different assumptions. This finding appears to indicate some tension between qualitative statements, which stressed the importance of monitoring and assessment as a vital plank in scour risk management, "best" estimates of failure probabilities which reflect these statements to some extent, and judgements of uncertainty, which appear to remain very conservative under the three assumed levels of resourcing that we tested.

6.3 Methodological findings

The workshop demonstrated that specialised elicitation methods often previously applied for very extreme natural and anthropogenic hazards could be used successfully to investigate infrastructure failure risks that are relatively infrequent, although not extremely rare compared with some other hazards, and subject to uncertainties of measurement and modelling. The workshop format stimulated strong debate about the problem definition, and the different assumptions relevant in different countries, in particular relating to the age profile and physical scale of bridges and rivers when comparing, say, the UK with North America.

The expert group had time to debate and modify the elicitation questions during the workshop, however the time available was necessarily constrained. A number of the expert panel members commented during the workshop, and in subsequent feedback, that it would have been useful to define the context for each elicitation question in more detail. For instance, assumptions made about inspection and maintenance protocols may have led to differences in how individual experts interpreted those questions. If experts assumed that bridges are routinely inspected after any flood event, then the occurrence of sequences of events might be viewed as less important than other vulnerability factors because any problems found in the inspection would be addressed in a manner commensurate with the nature and extent of the problem. Under these circumstances, past flooding experience may not have been regarded as an important primary indicator of increased vulnerability. Feedback after the workshop indicated also that there could be differences of interpretation relating to the physical and engineering context for a particular structure. For example, the questions did not specifically distinguish between channels with cohesive versus non-cohesive sediments, or tidal versus non-tidal flows.

Following informal feedback and discussions with some of the group, we conclude that there would be merit in holding some form of initial consultation, prior to an elicitation workshop of this type, to establish whether an expert group feels the intended target questions are defined precisely enough, and with sufficient supporting contextual information to be interpreted unambiguously. Bearing in mind that the aim of an elicitation is to gather evidence of experts' judgements about uncertainties, rather than their capacity to access information from the literature or other resources, there then would be a further challenge to provide sufficient but not excessive context material without inducing pre-judgement influences, such as availability bias.

When individual experts' estimates of failure probabilities were combined according to their uncertainty judgement weights, validated against a set of control questions in the Classical Model analysis, the pooled uncertainty bounds became narrower relative to those produced by unweighted averaging, particularly for situations where a bridge is inherently resilient

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

Natural Hazards and Earth System Sciences Discussions

(i.e. lower failure probability cases); this appears to reflect a less tentative, less precautionary judgement about uncertainty for the most resilient asset types when compared with a naïve, uncritical appraisal of all experts' responses.

There are intangible benefits to be gained from fostering communication and discussion between internationally diverse groups of experts from various different sectors, and the workshop, with its structured elicitation process, provided a constructive – and stimulating – forum for such exchanges.

6.4 Comparison with industry scour risk assessment guidance

In this study, the factors identified as important in assessment of scour risk are broadly consistent with industry guidance (summarised with a UK focus, but including reference to international good practice, by Kirby et al., 2015). Factors considered by the expert group that do not have obvious counterparts within industry guidance, for either screening or detailed assessments, related to: sequences of events, expressed here in terms of the number of floods in recent years; construction date of a bridge; angle of the approach flow, and removal of weirs in the vicinity of a bridge (although the latter is considered in various contexts by Kirby et al., 2015 and Arneson at al., 2012). None of the above factors was ranked within the nine most important by the expert group.

This study was informed by a framework for risk analysis predicated on a probabilistic treatment of hazards and fragility, extending further than the "design event" concept adopted within most industry guidance. In UK scour management guidance, a detailed scour assessment involves estimating potential scour depth for a design event and comparing this with foundation depth. Starting from the perspective that failure probability is conditional on loading, which could be defined in many different ways, the study has explored formulations for a more general, probabilistic failure function and the associated uncertainties about estimates of failure probabilities over a wide spectrum of load events. In assessing possible definitions of the load condition, the duration of a flood event and the possibility of sequences of events increasing the chance of a failure are regarded as important considerations, in addition to measures of peak hydraulic load. Flood return period, or exceedance probability, was considered as a standardised, probabilistic expression for the load condition in a fragility function.

Knowledge and data uncertainties are considered within industry guidance through a combination of qualitative and quantitative measures. Here, a more explicit quantification of expert judgements about uncertainty was possible through the application of structured elicitation methods. Pooled judgements about uncertainty in scour failure probabilities are more tightly constrained by taking account of the empirical calibration of individual experts' accuracy in assessing uncertainties, although this effect diminished as more extreme, and therefore rarer, flood events were considered.

Pooled expert estimates of failure probabilities reduced when considering scenarios involving increasing assumed levels of resources invested in scour assessment and maintenance. This can be seen as coherent in relation to the use of tiered risk management approaches involving generalised, high level screening followed by selective detailed assessments to enhance confidence in the mitigation of scour risk on a prioritised basis. Indeed, the essential role of investment in scour assessment and maintenance was stressed repeatedly by the expert group. However, the experts' weighted and pooled judgements about uncertainty remained wide regardless of whether assessment and maintenance was assumed to be more or less intensive than

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

Natural Hazards and Earth System Sciences Discussions

the *status quo*, suggesting that residual uncertainties remain, even after mitigation of the risk of scour, and that the residual risk of bridge failures should not be ignored.

Author contribution

RL prepared the manuscript with contributions from all co-authors, WPA and HO facilitated the elicitation exercise and carried out the data analysis.

Acknowledgments

We are grateful to Lisa Hill (University of Bristol) for project management support and to each of the participants in the expert group, listed below in alphabetical order, who gave their time to contribute so enthusiastically to the workshop: Michael Beer (Professor of Uncertainty in Engineering, University of Liverpool); Jeremy Benn, (Executive Chairman, JBA Group): Kevin Dentith (Chief Engineer Bridges & Structures, Devon County Council); Rob Ettema (Professor of Civil and Architectural Engineering, University of Wyoming); Kevin Giles (Senior Project Engineer, Network Rail); Peggy Johnson (Professor of Civil Engineering, Penn State University); Andy Kosicki (Chief, Structure Hydrology and Hydraulics Division, MD State Highway Administration); John Lane (Structures Engineer, Rail Safety Standards Board); Caroline Lowe (Principal Engineer, Network Rail); John McRobert (Highway Structures Unit, Department for Regional Development, Northern Ireland); Bruce Melville (Professor of Civil and Environmental Engineering, University of Auckland); Chris Perkins (Senior Programme Manager, Asset Management, Network Rail); John Phillips (Environment Agency); Marta Roca-Collell (Principal Engineer, HR Wallingford); Max Sheppard (Principal, INTERA Incorporated); Thorsten Wagener (Professor of Civil Engineering, Bristol University); Bruce Walsh (Principal, Northwest Hydraulic Consultants); Lyle Zevenbergen (Hydraulic Engineer, Tetra Tech Surface Water Group).

The elicitation was anonymised throughout and the analysis processed by a neutral facilitator. By the nature of the elicitation methodology, the findings presented here cannot be attributable to any individual. The subsequent interpretations presented in this paper are those of the authors.

This study was funded by the UK Natural Environment Research Council (NERC) under the Environmental Risks to Infrastructure Innovation programme, grant number NE/M008746/1. RL was funded by the JBA Trust (www.jbatrust.org). Jim Hall (Oxford University) is thanked for proposing a bridge scour elicitation study following earlier work within the CREDIBLE consortium (NERC Grant NE/J017299/1), which also supported WPA and TW.

References

Arneson, L. A., Zevenbergen, L. W., Lagasse, P. F. and Clopper, P. E.: Evaluating Scour at Bridges, 5th Edition, Federal Highway Administration Hydraulic Engineering Circular No. 18, Publication No. FHWA-HIF-12-003, U.S. Department of Transportation, Washington D.C., 340 pp., 2012.

Aspinall, W. P., Cooke, R. M.: Expert Elicitation and Judgement, in: Risk and Uncertainty Assessment in Natural Hazards, Rougier, J. C., Sparks R. S. J., Hill, L. (Eds), Cambridge University Press, Chapter 4, 64-99, 2013.

Aspinall, W. P., Cooke, R. M., Havelaar, A. H., Hoffmann, S., Hald, T.: Evaluation of a Performance-Based Expert Elicitation: WHO Global Attribution of Foodborne Diseases, PLoS ONE 11(3): e0149817, doi:10.1371/journal.pone.0149817, 2016.

Bamber, J. and Aspinall, W. P.: An expert judgement assessment of future sea level rise from the ice sheets, Nature Climate Change, 3, 424-427 doi:10.1038/nclimate1778, 2013.

Coleman, S. E., Lauchlan, C. S. and Melville, B. W.: Clear-water scour development at bridge abutments, Journal of Hydraulic Research, 41:5, 521-531, DOI: 10.1080/00221680309499997, 2003.

Cooke, R.M.; Experts in Uncertainty, Oxford University Press, 1991.

Cooke, R. M. and Goossens, L. L. H. J.: TU Delft expert judgment data base, Reliability Engineering and System Safety – Expert Judgement 93(5): 657-674, 2008.

Cooke, R. M. and Solomatine, D.: EXCALIBR User's Manual, Delft, Delft University of Technology/SoLogic Delft: 33 pp, 1992.

Cook, W.: Bridge Failure Rates, Consequences, and Predictive Trends, Utah State University, All Graduate Theses and Dissertations, Paper 2163. http://digitalcommons.usu.edu/etd/2163 (accessed December 2014), 2014.

Cumbria County Council: Cumbria Floods Update on bridges and highways, http://www.cumbria.gov.uk/news/2015/december/09_12_2015-114140.asp, Accessed September 2016.

Decò, A. and Frangopol, D. M.: Risk assessment of highway bridges under multiple hazards, Journal of Risk Research, 14:9, 1057-1089, DOI: 10.1080/13669877.2011.571789, 2011.

Defra: Green Leaves II: Guidelines for Environmental Risk Assessment and Management, Department for Environment, Food and Rural Affairs Report PB13670, London, 84 pp., https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/69450/pb13670-green-leaves-iii-1111071.pdf (Accessed online September 2016), 2011.

Ellingwood, B. R.:. Structural reliability and performance-based engineering, Structures and Buildings 161(SB4):199–207, 2008.

FHWA: National Bridge Inspection Standards, U.S. Department of Transportation Federal Highways Administration, available online at https://www.fhwa.dot.gov/bridge/nbis.cfm (Accessed September 2016), 2016.

Hall, J. W., Nicholls, R. J., Tran, M. and Hickford, A. J. (Eds.): The Future of National Infrastructure: A System-of-Systems Approach, Cambridge University Press, ISBN: 9781107066021, 338 pp., 2016.

Highways England: Standards for Highways online resources, https://www.gov.uk/guidance/standards-for-highways-online-resources, 2016.

Hong, J-H., Goyal, M. K., Chiew, Y-M., Chua, L. H. C.: Predicting time-dependent pier scour depth with support vector regression, Journal of Hydrology 468–469, doi.org/10.1016/j.jhydrol.2012.08.038, 2012.

Kattell, J. and Eriksson, M.: Bridge Scour Evaluation: Screening, Analysis, and Countermeasures, United States Department of Agriculture Forest Service, Technology and Development Program, 7700 –Transportation Systems, Report 9877 1207 – SDTDC, 12 pp., available online at http://www.fs.fed.us/eng/structures/98771207.pdf (accessed 15 December 2014), 1998.

Kirby, A. M., Roca, M., Kitchen, A., Escarameia, M., Chesterton, O. J.: Manual on scour at bridges and other hydraulic structures, second edition, CIRIA Report C742, ISBN: 978-0-86017-747-0, London, 2015.

Macutkiewicz, M. and Cooke, R. M.: UNIBALANCE Users Manual, TUDelft, 31pp, 2006.

Melville, B. W.: Pier and abutment scour: Integrated approach, Journal of Hydraulic Engineering – ASCE, Vol.123(2), pp.125-136, 1997.

Melville, B.W. and Chiew, Y. M.: Time Scale for Local Scour at Bridge Piers, Journal of Hydraulic Engineering – ASCE, 25(1), 59-65, 1999.

Rail Safety and Standards Board: Safe Management of Railway Structures, Flooding and Scour Risk, Report No. T554, London, 104 pp., Available online at http://www.rssb.co.uk/research-development-and-innovation (accessed 17 December 2014), 2005.

Smith, D. W.: Bridge failures, Proc. Inst. Civ. Engrs, Part 1, Vol 60. p. 367-382, 10.1680/iicep.1976.3389, 1976.

TSO (2010) UK Design Manual for Roads and Bridges

UK Roads Liaison Group: Background Briefing on Highway Bridges, http://www.ukroadsliaisongroup.org/en/utilities/document-summary.cfm?docid=59ABF16C-03DE-4864-B31C64A86371A89F (Accessed June 2015), Chartered Institution of Highways & Transportation, London, 2009.

Tyshenko, M. G., El-Saadany, S., Oraby, T., Darshan, S., Aspinall, W., Cooke, R., Catford, A. and Krewski, D.: Expert elicitation for the judgment of prion disease risk uncertainties, J. Toxicol. Environ. Health A., 74(2-4): 261-285, 2011.

USACE: Beyond the Factor of Safety: Developing Fragility Curves to Characterize System Reliability, Water Resources Infrastructure Program Report ERDC SR-10-1, July 2010, U.S. Army Corps of Engineers, Washington, DC, 59 pp: http://www.dtic.mil/dtic/tr/fulltext/u2/a525580.pdf (Accessed online September 2016), 2010.

van Leuwen, Z. and Lamb, R.: Flood and scour related failure incidents at railway assets between 1846 and 2013, JBA Trust Report No. W13-4224, 34 pp., Available online at http://www.jbatrust.org/how-we-help/publications-resources/risk-analysis/flood-and-scour-related-failure-incidents-at-railway-assets/ (Accessed October 2016), 2014

Zurich Insurance Group and JBA Trust: Flooding after Storm Desmond, http://www.jbatrust.org/how-we-help/publications-resources/review-of-flood-response-in-cumbria-following-storm-desmond/, Accessed September 2016, 2016.

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

Natural Hazards and Earth System Sciences Discussions

Table 1: Proposed vulnerability factors

Group	*Proposed factors*	*Comments*
Characteristics of the bridge structure	▪ Foundation depth ▪ Foundation type ▪ Structure span ▪ Construction date ▪ Existence of scour protection ▪ Flow constriction at the bridge ▪ Bridge type	Relate to static characteristics of the structure.
Characteristics of the watercourse	▪ Bed material ▪ Unstable watercourse	Factors relating to hydro-morphological situation in the river
Hydraulic conditions	▪ Flow velocity ▪ Location on a river bend or confluence ▪ Oblique approach flow	Location on bend/confluence and oblique approach were included in view of their potential effects on velocity distributions and turbulence.
History and uncertainty about information	▪ Application of scour assessment and monitoring procedures ▪ Whether there is a history of scour problems ▪ Whether or not foundation depth is known ▪ Whether or not foundation type is known ▪ Number of floods in the last 5 years ▪ History of debris accumulation	Broad group of factors reflecting how much is known about scour vulnerability at a bridge, including evidence from past events (especially previous occurrence of scour) and also whether the bridge characteristics are well known.
Change factors	▪ Sand/gravel extraction in the reach near the bridge ▪ Weir has been removed near bridge	Changes at the bridge or elsewhere in the watercourse that could lead to changes in susceptibility to scour.

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

Natural Hazards and Earth System Sciences Discussions

Table 2: Ranking scores for the importance of factors that should be considered in assessing scour risk to bridges (Question 1). Higher score indicates greater importance. "St. dev." is the score standard deviation derived from probabilistic inversion of experts' collective responses.

		Piers		*Abutments*	
Item	*Factor description*	*Score*	*St. dev.*	*Score*	*St. dev.*
1	Foundation depth	0.61	0.26	0.59	0.28
2	Foundation type	0.63	0.32	0.53	0.28
3	Whether foundation depth is known or not	0.51	0.35	0.51	0.33
4	Whether foundation type known is known or not	0.43	0.32	0.43	0.29
5	Bed material	0.47	0.23	0.45	0.29
6	Structure span	0.25	0.24	0.39	0.31
7	Scour history	0.71	0.24	0.69	0.23
8	Application of scour assessment and monitoring procedures (labelled "assmt/procedure")	0.58	0.29	0.51	0.29
9	Construction date	0.33	0.16	0.30	0.24
10	Flow velocity	0.59	0.19	0.67	0.23
11	Number of floods in the last 5 years	0.32	0.23	0.39	0.27
12	Existence of scour protection	0.64	0.20	0.53	0.29
13	Location on a river bend or confluence	0.36	0.20	0.35	0.20
14	Oblique approach flow	0.48	0.26	0.34	0.24
15	Constriction at bridge	0.56	0.23	0.57	0.27
16	Bridge type	0.18	0.17	0.24	0.20
17	History of debris accumulation	0.57	0.26	0.50	0.26
18	Unstable watercourse	0.68	0.23	0.63	0.25
19	Sand/gravel extraction in the reach near the bridge	0.71	0.24	0.67	0.23
20	Weir has been removed near bridge	0.55	0.21	0.48	0.24

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

Natural Hazards and Earth System Sciences
Discussions

Table 3: Ranking scores for factors according relevance to defining the loading condition for a scour fragility function

Item	*Name*	*Score*	*St. dev.*	*Cluster*
9	Frequency and amount of debris	0.65	0.25	A
1	Peak flow	0.63	0.25	A
6	Flow return period	0.61	0.30	A
7	Flow velocity relative to sediment critical flow	0.59	0.26	A
3	Time during which flow is greater than a critical threshold for scour initiation ("Time flow > threshold")	0.59	0.26	A
2	Peak water level	0.45	0.26	B
4	Time during which level is greater than a critical threshold for scour initiation ("Time level > threshold")	0.45	0.26	B
5	Number of "high flows" (capable of causing scour) in last year	0.41	0.28	B
10	Sediment concentration reaching the bridge at high flows ("High flow sediment concentration")	0.34	0.25	C
8	Application of scour assessment and monitoring procedures ("Assessment/procedure")	0.31	0.23	C

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

Natural Hazards and Earth System Sciences
Discussions

Table 4: Ranking scores for factors affecting change in scour vulnerability

Item	*Name*	*Score*	*St. dev.*
4	Inspection regime changes	0.69	0.26
5	Maintenance regime changes	0.62	0.25
7	Dredging up/downstream	0.61	0.25
9	Watercourse changes	0.58	0.27
8	Weir/dam removal	0.54	0.25
6	Flood defence construction	0.52	0.24
2	Catchment land manage changes	0.47	0.27
1	Climate change affects frequency of extreme rain	0.22	0.20
3	Bridge use demands	0.22	0.19

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

Natural Hazards and Earth System Sciences Discussions

Table 5: Judgements about flood relative magnitude (in return period, years) appropriate to trigger asset inspection.

	Lower value (5th percentile)	*Median (50th percentile)*	*Mean*	*Upper value (95th percentile)*
Group estimate pooled with experts' weighted according to calibration questions	1.0	5.6	15	48
Group estimate pooled with experts' weighted equally	1.2	26	94	318

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

Natural Hazards and Earth System Sciences
Discussions

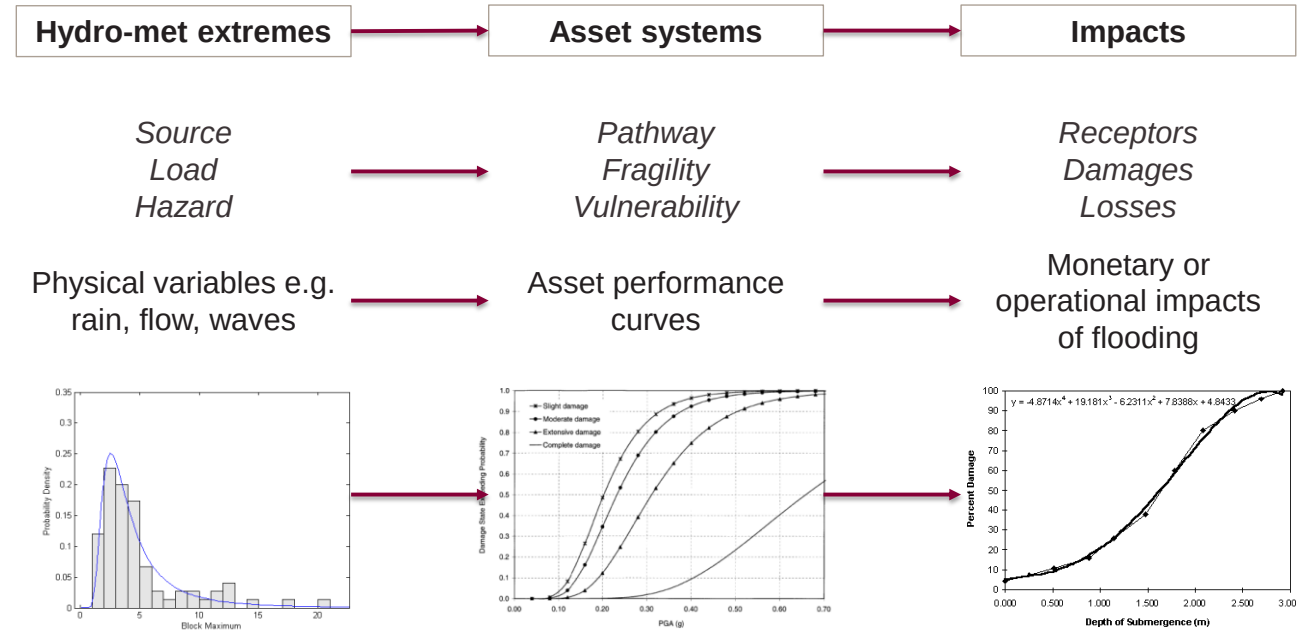

Figure 1. High-level conceptual risk model

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

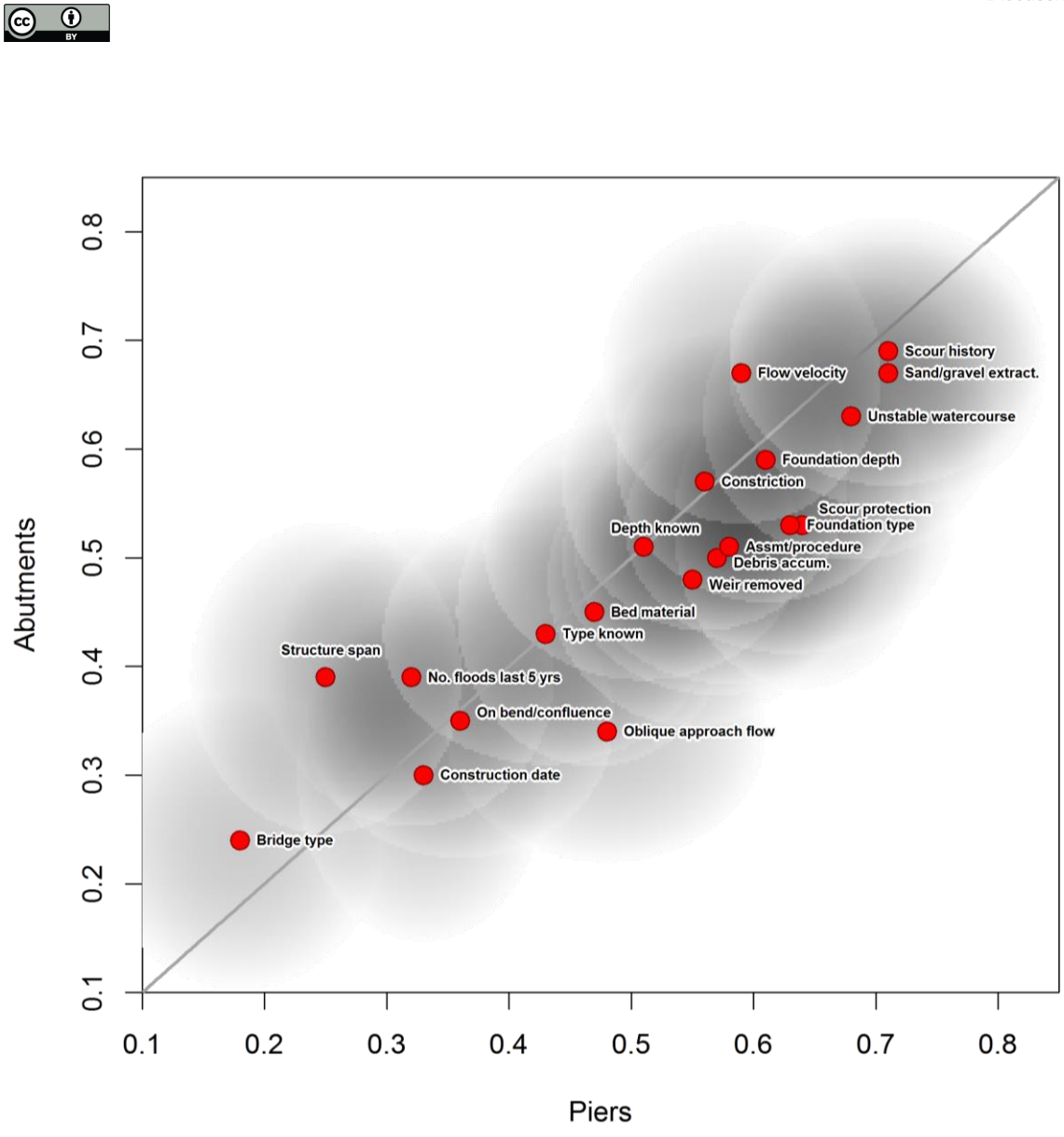

Figure 2. Ranking scores for the importance of factors that should be considered in assessing scour risk to bridges (Question 1), comparing the ranking and confidence of scores when considering scour at bridge piers (horizontal axis) and at abutments (vertical axis). Higher score indicates greater importance. Ellipses depict 95% confidence areas for factor ranking scores from probabilistic inversion of experts' collective responses. Horizontally extended ellipses indicate greater variance in ranking factors for Piers relative to rankings for Abutments; vertically extended ellipses indicate greater variance for abutments.

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

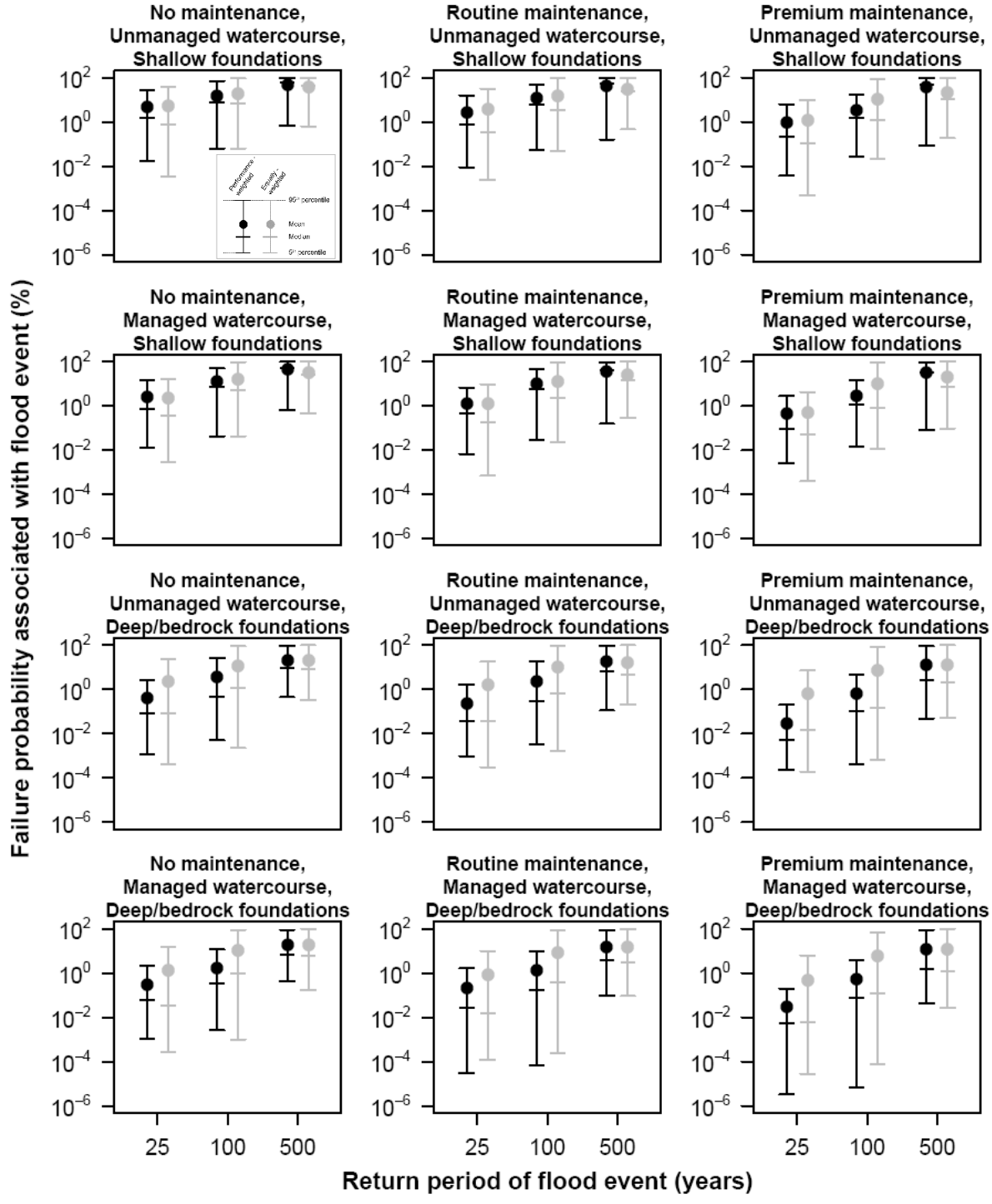

Figure 3. Fragility estimates for bridge failure probability as a function of flood event rarity.

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

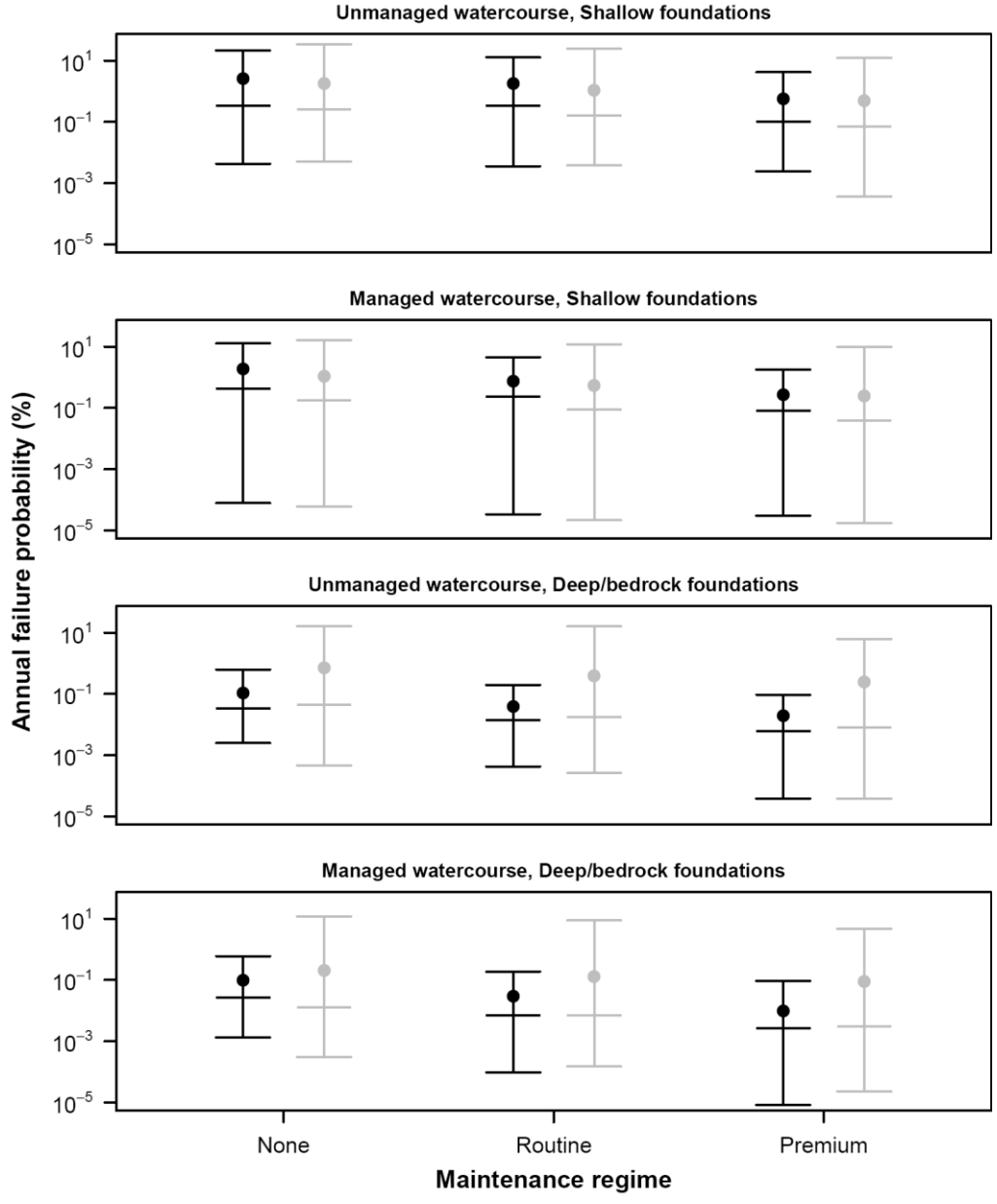

Figure 4. Fragility estimates for annual unconditional bridge failure probability under three assumed monitoring and maintenance ("maintenance") regimes.

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-350, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 25 November 2016

Natural Hazards and Earth System Sciences
Discussions

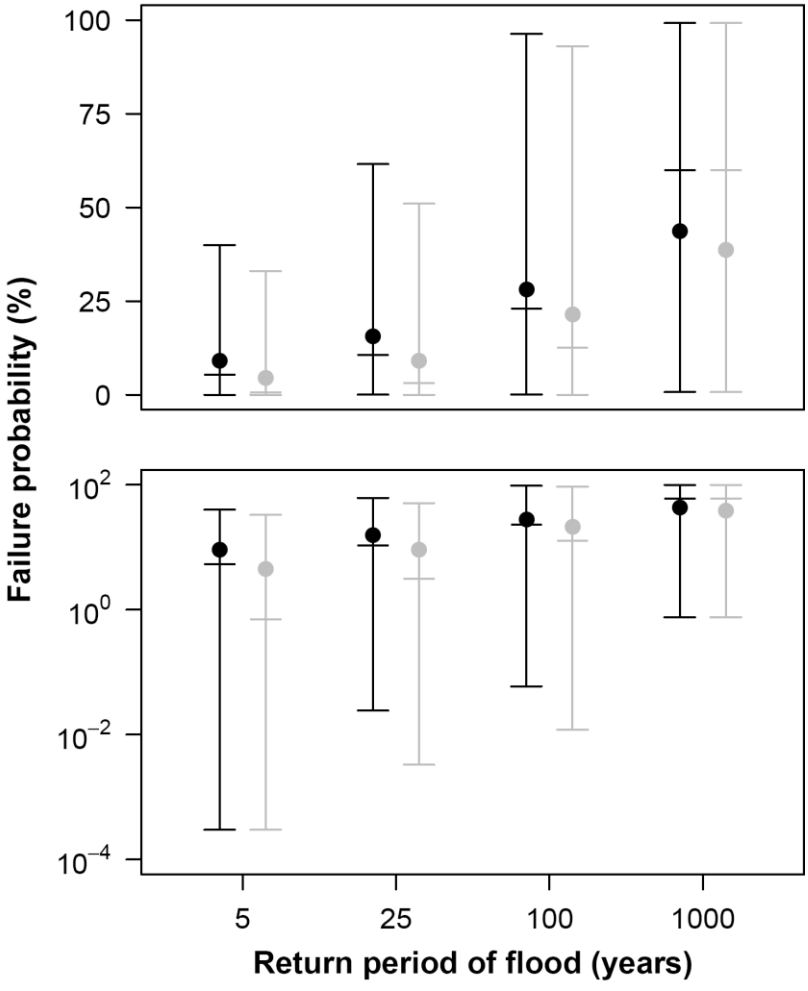

Figure 5. Estimated bridge failure probabilities as a function of flood event rarity, conditional on a preceding flood event of 100-year return period having occurred with no intervening maintenance action. Upper and lower panel show the same data, plotted on different scales.