Case study on expert elicitation of slope failure probabilities for quantitative risk estimation – Evaluating the spread of elicited values

Renato Macciotta

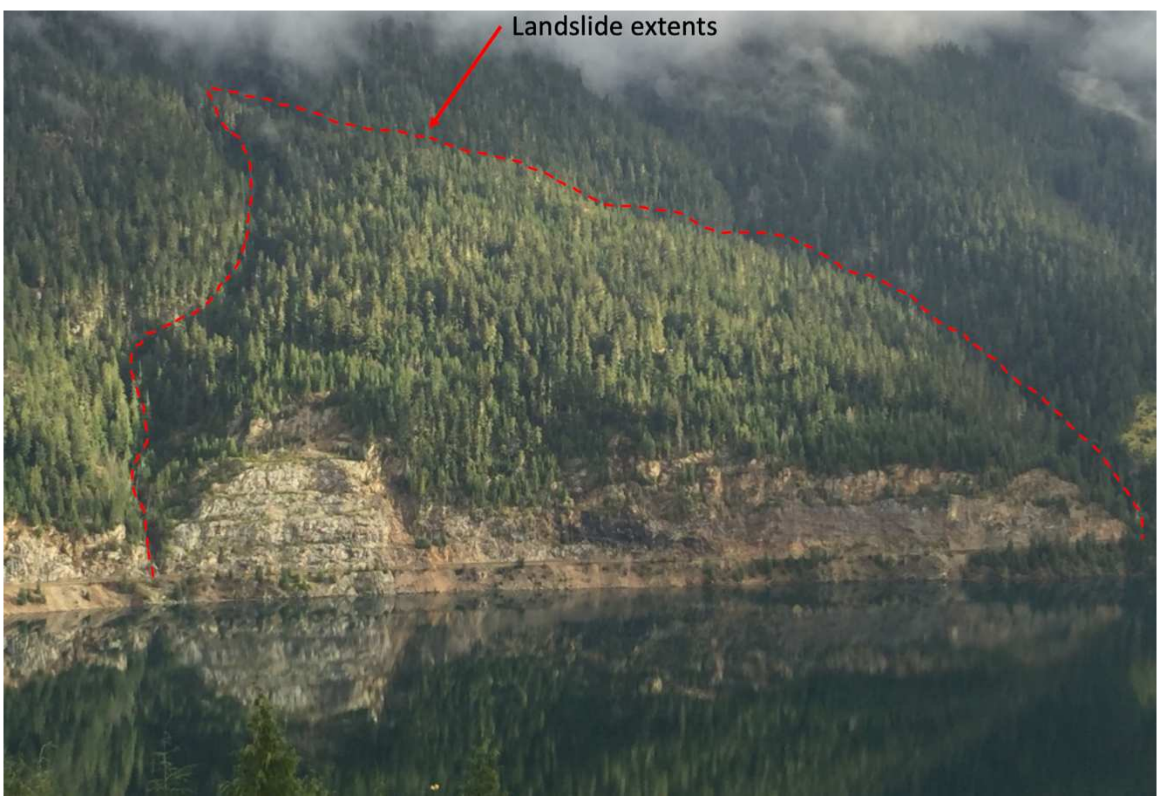
Department of Civil and Environmental Engineering, School of Engineering safety and Risk Management University of Alberta, Edmonton, AB, Canada

macciott@ualberta.ca

Abstract

The benefits of quantitative risk assessments for landslide management are well known and have been illustrated in several publications. Challenges remain regarding its application for low-probability, high-magnitude events. These challenges are associated with the difficulties in populating our models for risk calculations, which largely require the input of expert opinion. This paper presents a case study of the elicitation process and spread obtained in the elicited probabilities.

The Checkerboard Creek slope is a continuously deforming rock slope located within the Revelstoke Dam reservoir in British Columbia, Canada. Estimating the risks associated with the Checkerboard Creek slope requires assessing its probability of failure. The nature of the slope forming materials and its deformation mechanisms do not allow for the calculation of the probability of failure based on limit equilibrium models, and probabilistic numerical simulations for this slope were technically challenging and time prohibitive. A workshop was developed to elicit failure probabilities, under different potential scenarios, in light of the available information and previous analysis of the slope behaviour.



1 CHECKERBOARD CREEK ROCK SLOPE

The Checkerboard Creek rock slope is located approximately 1.5 km upstream (north) of the Revelstoke Dam, along the eastern slopes of the Columbia River Valley, BC, Canada. It has a height of approximately 260 m with an overall slope angle of about 30 degrees, being steeper at the toe (45 degrees) and softer in the upper area (25 degrees) (Stewart and Moore, 2002; Watson et al., 2004 and 2006; Lorig et al., 2009; Macciotta et al., 2010 and 2016).

As part of the Revelstoke Dam Project, relocation of Highway 23 to a higher elevation (about 70 m higher) was required to accommodate the designed reservoir level. A steep design of the road cuts was adopted (15 m high faces at 1H:4V inclination and with 3 m wide benches) and no rock support specified. Considerable backbreak of the cut benches occurred during excavation, leading to a change in the rock slope design and the initiation of more detailed investigations in the area. Slope inspections carried out during the reservoir infilling in 1984 identified a series of tension cracks north of Checkerboard Creek, in the slope upper zone.

1.1 Landslide extents and geological setting

The extent of the deforming rock mass has been interpreted from geological studies and deformation monitoring. Figure 1 shows the location and extents of the landslide. The upper boundary is well defined by the alignment of the uppermost exposed tension cracks. The lateral boundaries, as well as the toe boundary, are not as clear and have been interpreted from the site geology, slope topography and deformation patterns. The active zone has an average slope angle of approximately 45 degrees, being steeper at the toe (road cut) with a slope angle of 50 to 60 degrees (Stewart and Moore, 2002; Watson et al., 2004 and 2006; Lorig et al., 2009; Macciotta et al., 2010 and 2016).

The Checkerboard Creek rock slope comprises massive to weakly foliated granodiorite overlying the easterly dipping Columbia River Fault, which has developed a broad zone of altered and mechanically deformed rock. Shears and joints in the area dip steeply into and out-of-slope at angles of 60 to 90 degrees from horizontal. The rock mass quality ranges from very strong, fresh, undisturbed and blocky rock to highly weathered and altered, weak and disturbed rock. Sheared and crushed zones are commonly found. The poor-quality rock mass is typically found within 60 m from the slope surface, where the active deformations have been observed. The rock mass beneath this area is generally fair to good in quality, with localized zones of poor-quality rock along shear zones and sub-vertical joints.

The shear zones in general are undulated, contain 1 to 5 mm gouge or breccia, have an associated zone of crushed and brecciated rock 10 to 50 cm thick, and often display a thin layer 1 to 2 mm thick of graphitic or altered rock along a discrete, discontinuous surface. Thicker shear zones, 5 to 15 m thick, are found dipping into the slope at 20 to 35 degrees. Wider shears appear to be continuous along strike and dip, and have developed up to 15 m of cataclastic rock, breccia and clay gouge. Two major joint sets can be observed: 1) a primary set sub-parallel to the slope contours, dipping 60 to 80 degrees, and 2) a secondary set roughly orthogonal to the primary set, dipping 85 to 90 degrees. It is believed this joint pattern yielding a blocky rock

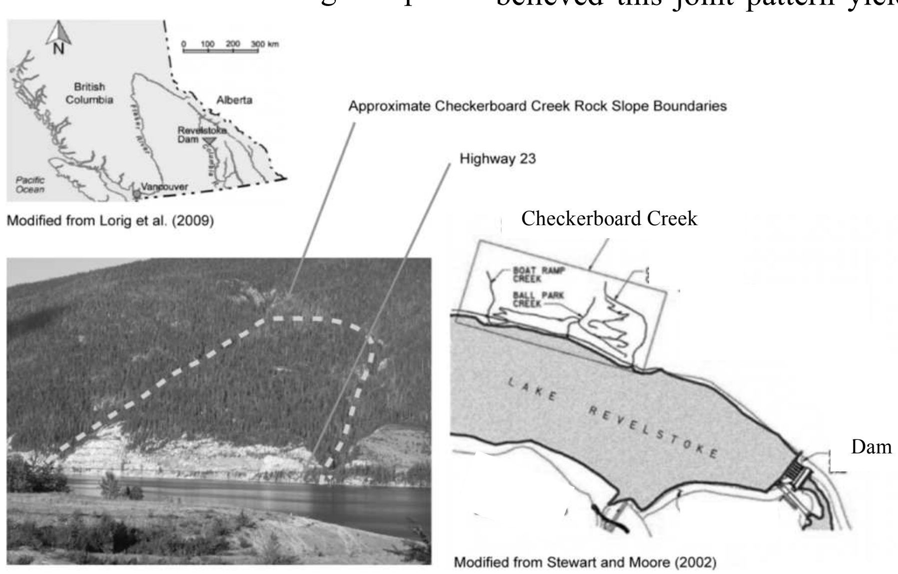


Figure 1. Location and front view of the Checkerboard Creek slope

mass is responsible for the moderate- to high-weathered rock reaching depths of 50 to 75 m below the surface.

1.2 Ground water

Numerous discrete pore pressure differences of less than 40 m across short distances have been measured, which is indicative of a compartmentalized groundwater regime. Continuously saturated conditions have been observed 50 to 80 m below the surface. These depths are deeper than the observed extent of the displacing rock mass. Seasonal variations in piezometric levels of up to 20 m have been detected just below the deforming zone; however, some measurements indicate pressures can fluctuate higher than the base of the deforming mass. Piezometric probes within the active displacement zone generally have shown zero pore pressure; however, transient groundwater flows and pressures are interpreted to develop in the upper compartmentalized zones. In this respect, some instrumentation installed in the unsaturated zone have shown transient water pressures of 1 to 3 m within the deforming mass (Stewart and Moore, 2002; Watson et al., 2004).

1.3 Landslide displacement patterns

Displacement monitoring has revealed an annual displacement cycle dominated by an active period from early October to April (i.e., early autumn throughout late winter) and a relatively quiet period from May to September (i.e., spring and summer) (Stewart and Moore, 2002; Watson et al., 2004). The displacement rate of the deforming rock mass is 0.5 to 13 mm/y, the greatest being at the surface and decreasing progressively with depth up to a point where no deformation is detected (up to about 60 m below surface). The deformations are generally widely distributed within the deforming mass; however, there are zones where these are more concentrated or absent. These patterns indicate that deformations are occurring along individual features distributed within the entire rock mass rather than sliding as a block through a continuous failure plane. Geomorphic evidence, mainly from open tension cracks, indicates a long history of deformation with an estimated surface displacement of about 10 m or more (Stewart and Moore, 2002; Watson et al., 2004, Woods et al., 2019).

The annual displacement rates and the annual displacement cycle are remarkably repeatable and follow a linear trend regardless of the amount of precipitation. The onset or acceleration of displacements are sometimes triggered during periods of increasing water pressures; however, any correlation with recorded piezometric levels or their rates of change is not consistent. The annual cycle shown by the deformation measurements is more strongly correlated to seasonal temperature variations in the bedrock near the surface than to groundwater pressures. At the onset or acceleration of movement, and during the active displacement period, the near surface bedrock temperature is decreasing. During the inactive months, the near surface bedrock temperature is increasing. It has been postulated that cooling of the near surface bedrock induces a reduction in the effective normal stress on sub-vertical discontinuities sub-parallel to the slope contours. This results in outward and downward displacement of the slope. During warming periods, the normal stresses increase and prevent further slipping. Detailed numerical analysis simulating the seasonal temperature fluctuations support this idea, however studies continue as ground water pressures have also been identified as causes for some acceleration episodes.

1.4 Previous studies on landslide behaviour

A discontinuum numerical model using UDEC (Universal Distinct Element Code) software from Itasca Consulting Group was developed (Watson et al. 2006; Lorig et al. 2009). The objective was to better understand the mechanisms leading to the slope behaviour and to use a calibrated model to predict the rock mass behaviour under several probable conditions. The results can be summarized as follows:

- Both annual temperature fluctuations and piezometric elevation fluctuations are capable of producing displacements consistent with the observed ones.
- Displacements appear to be controlled mainly by yielding on steep discontinuities.
- Deformations resulting from piezometric fluctuations near the base of the deforming rock mass were very sensitive to the interpreted extent of the applied pressures, but comparable to observations.
- Thermal fluctuations of the near surface bedrock induce displacements which are greatest at the surface, but extend deeper than 50 m into the slope. These were very sensitive to input parameters such as thermal expansion coefficient, applied temperature profile, discontinuity

strengths, elastic modulus and location of the phreatic surface.

- Under static conditions, slope failure is only initiated when rock mass strengths are considerably reduced. These strength reductions are believed to require large deformations, and even in this case, failure is limited to the toe of the slope (less than 0.5 Mm3, face of the slope cut).
- Seismic loading under several conditions did not result in large-scale collapses; the maximum residual slope displacements were less than about 700 mm for the 10 000-year-return period seismic event (Peak Ground Acceleration of 0.2 times the gravity).
- Even under unrealistic rock mass and discontinuity strength reductions, the models did not predict failure of the total deforming mass. Only local collapse of the highway cut was observed for some seismic scenarios.
- It is considered the model did not have enough geological detail to achieve good reliability regarding small-scale failure predictions of the highway cut.

2 EXPERT ELICITATION

An expert opinion can be defined as a formal judgment, or belief, of an expert on a particular matter, which is subjective, and typically based on uncertain information or knowledge (Ayyub, 2001). Expert opinion is part of every engineering and risk management decision. However, as the uncertainties become greater than the knowledge of the system analyzed, structured methods for expert elicitation become a necessity.

One of the most common methods for expert elicitation is the Delphi method (Helmer, 1968). it consists of the following steps (after Ayyub 2001):

- Definition of the information required from the elicitation process,
- Development of questionnaires to gather the information,
- Selection of experts,
- Familiarization of experts about the subject matter to be elicited,
- Elicitation of experts - gather responses to the questionnaires,
- Aggregation and presentation of results,
- Review of results by the experts and revision of initial responses. Responses outside the 25% and 75% percentile values should be accompanied by proper justification,
- Repetition of previous three steps until consensus is achieved or results are considered precise enough, and;
- Summary of results and justification.

Variations of this method typically consist of the type of questionnaire (direct, indirect or parametric), the level of interaction between experts, and how results from different experts are aggregated.

The questionnaires can be developed such that the elicitation process is direct, indirect or a parametric estimation that deals with uncertainty (Ayyub 2001). The direct method elicits a direct estimate of the belief of an expert in the information required from the process. Indirect elicitation consists in eliciting answers on metrics more familiar to the experts, and later deriving the required metrics for analysis. An example of this would be eliciting the time to failure of a system and then deriving the failure probabilities from it. Another approach is to build questionnaires comparing the unknown required information with familiar situations with known answers. Answers are in the form of relative measures with respect to the familiar situations for which actual statistical or measured values exist.

A parametric estimation obtains a measure of the information required and a measure of the confidence interval. An initial step consists in eliciting the information (such as the median estimate of a probability), followed by a second step where a measure of dispersion is elicited. These elicited values can then be used to compute confidence bounds for the elicited information (Preyssl and Cooke 1989, Ayyub 2001).

There can be different levels of interaction among experts. The nominal group technique discussed in Morgan and Henrion (1992) includes discussion among experts after the initial elicitation responses are presented.

2.1 Aggregation of Expert Elicitation Responses

Consensus can be reached through discussion among the elicited panel of experts. This can include measures of uncertainty such as confidence intervals or outer quartile values. The shortcomings rely in the potential for conformity within the group, dominance of strong-minded participants and biases due to common background (Ayyub 2001).

Aggregation can also be done mathematically. Responses can be treated as a statistical sample and point estimates derived from them (average, median, percentiles). There will be situations where different experts might provide responses with different degrees of reliability on their answers. A scoring of experts can then be applied to weight the responses during aggregation. This process can consist in self-scoring or each participant scoring the other members of the panel.

Methods for weighted combinations of opinions are summarized in French (1985), Genest and Zidek (1986) and Ayyub (2001). These include weighted arithmetic, geometric and harmonic averages. The statistical sample of responses can also be increased according to each expert score and percentiles be derived from the modified sample. Ayyub (2001) also describes three methods for expert elicitation aggregation based on uncertainty measures. These can deal with various uncertainty types, offering a strengthened approach. On the other hand, they are of an increased complexity and computational demand.

The minimum uncertainty criterion focuses on finding the tendency of the responses, in order to maximize information retention, thus minimizes uncertainty. This can lead to a reliance on the response tendency as representative of the true value. The maximum uncertainty criterion focuses on utilizing all the information gathered, both the tendency of responses and the spread (or uncertainty). Likelihood distribution of responses can be built given a set of constraints based on the elicited responses. The maximum uncertainty criterion would then rely on probability distributions showing maximum entropy (uniform distribution if maximum and minimum values are selected as constraints, normal distribution if an expected value and variance are selected as constraints). Aggregation can also be done by defining intervals for the elicited responses. These can then be treated as fuzzy sets, where the intervals are a special case. Fuzzy arithmetic and calculus can then be applied to combine expert opinions in this approach (Ayyub 2001).

3 METHODOLOGY ADOPTED AND WORKSHOP PROCESS

In this paper the maximum uncertainty was adopted, through eliciting expected values and their plausible ranges for each probability elicited.

The workshop started presenting the details of the case study, followed by a participant calibration session. The uniqueness of the case study prevents calibration through statistical analyses of recorded failures under similar conditions, therefore calibration focused on achieving a consistent sense of frequency magnitudes of rare events. It was also intended to make the participants feel more comfortable assessing numbers within several orders of magnitude.

3.1 Calibration process

The calibration consisted on presenting participants with situations for which their frequencies / magnitudes are reported in the literature, but not readily known by the public (i.e., probability of being in a car accident each year). The participants are then asked to assess these situations. The reported statistics / magnitude for each situation is then given before the next calibration question. The questions used for the calibration process were the following (answers in parenthesis):

Based on your travel experience and daily commute:

1. How many seats can a light rail / metro car accommodate? (60 to 80)
2. How many people fit in a light rail / metro car at full capacity (seating and standing)? Think about peak hours. (about 2×10^2)
3. In terms of seating capacity, how many times more seats does a commercial airplane (i.e., as used for intercontinental flights) accommodate than a light rail / metro car? (5 to 8 times)

For the average Canadian resident:

4. What is the annual probability of being killed in a motor vehicle accident? (about 10^{-4})
5. How many times more / less likely is it to be killed during a commercial airline flight that same year? (10^{-3})
6. What is the annual probability of being killed by lightning? (about 5×10^{-7})
7. How many times more / less likely is it to be murdered that same year when compared to being killed by lightning? (about 10^2)

Based on your geotechnical experience and daily routine:

8. How many sand grains (medium-sized sand) are there in a 5 cm diameter, 10 cm tall triaxial test sand specimen? (Estimate regardless of whether it is loose or dense!) (between 10^4 and 10^8 for loose and dense sand, respectively! 10^6 is a good estimate for medium-sized sand.)

9. How many times more sand grains will fit in your bathtub? (10^3)

Imagine yourself as a child at a family vacation to the countryside. You are staring at a mountain and notice a beetle climbing up your shoe:

10. How many times taller does the mountain look when compared to the beetle's height? (10^5)
11. How many times smaller does the beetle look when compared to your size? (10^{-2})

3.2 Elicitation Process

Probability elicitation then proceeds by completing a questionnaire. The participants are asked what they believe to be the range of credible answers to each question. They are also asked what they believe is the most likely answer to each question. Questions were targeted to three scenarios: 1) localized slope failure (up to 0.5 million m^3), 2) failure of the actively deforming mass (2 to 3 million m^3), and 3) failure of the entire landslide (about 30 million m^3). The questions targeted to the scenario of a failure of the actively moving mass were the following:

1. Given "average" site conditions (No seismic event or changes in the groundwater regime within the deforming mass), assess the period of time for which there is an EVEN chance (50:50) for slope failure to occur.
2. Given "average" site conditions, estimate the period of time for which failure of the slope is expected (99% probability it will fail).
3. How more / less likely would a failure occur during the active slope deformation period compared to the "quiet" deformation period within the same year?
4. Compared to "average" site conditions, how more / less likely would a failure occur given an increase in the pore water pressures within the slope up to twice the maximum recorded so far?
5. A seismic event occurs with a peak ground acceleration (PGA) of 0.2 times the gravity. What is your assessment of the likelihood that the slope fails?
6. How more / less likely would a failure occur given the seismic event has a peak ground acceleration (PGA) of 0.05 times the gravity when compared to the previous scenario where the PGA is 0.2 times the gravity?
7. How more / less likely would a failure occur after a seismic event given an increase in pore water pressures within the slope up to twice the maximum recorded so far?
8. Overall, assess how more / less likely the following scenarios would occur: localized failure vs. actively deforming mass vs. overall landslide collapse.

Scales were provided for each question for ease of answer and visualization. An example is provided in Figure 2.

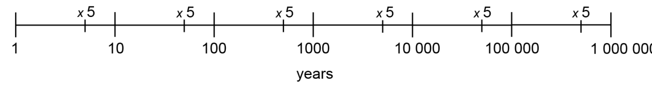


Figure 2. Example graphic scale for elicitation process

Answers from these questions were used to calculate the failure conditional probabilities for the risk estimations. For example, question 1 collects the belief for the return period associated with a 50% probability of failure. From the binomial theorem, and taking the return period (in years) as the number of tests (failure or no failure during each year); the annual probability of failure under average conditions can be estimated as:

$$0.5 = 1 - (1 - p)^{Q1} \tag{1}$$

Where p is the annual probability of failure under average conditions and $Q1$ is the answer to question 1.

Other answers were used to scale the failure probabilities for diverse scenarios and through the binomial theorem when probabilities and return periods were stated or asked.

4 RESULTS

18 people participated in the elicitation workshop. Participants had varying levels of experience, however no weighting was applied to the participants' responses at this stage of the process.

Each participant response (range) was associated with a uniform distribution for aggregation purposes, and a "count" of 1 was associated with those values within the range provided. if the participant also provided an expected value, that value was provided with an additional count of 1 (for a total of 2).

Aggregation of responses were done simply by summation of the number of responses for each value range. This provided a distribution of the frequency of responses for different ranges of values. Figure 3 shows example response distributions for questions 1 and 4.

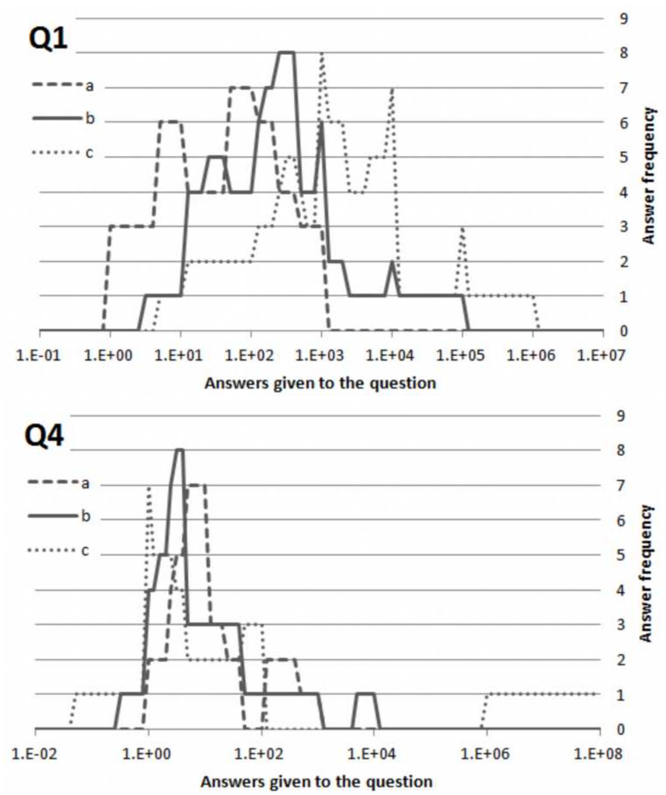


Figure 3. Example results for questions 1 and 4 for localized failure (a), actively deforming mass failure (b), and total slope collapse (c)

These results were used to derive the lower and upper range values of failure probabilities elicited; as well as the expected values. In order to evaluate the spread in elicited values, only lower and upper values are presented here. Further, only values with a frequency of 2 or more are considered (at least two respondents agree in the plausibility of the value) in order to minimize outliers. Results for the actively deforming mass (2 to 30 million m^3) are shown in Table 1. Simple inspection indicates that some of the upper levels appear high, mainly for the warming period. Furthermore, it was identified that the spread in the estimated failure probabilities due to subjective probability elicitation of diverse practitioners with different levels of experience was significant.

The tendency of results as shown in Figure 3, however, suggests that a more robust approach would be to fit distributions to these responses and work with plausible ranges defined by standard deviations in the logarithmic scale (e.g. range defined by 1 or 2 standard deviations from the expected value). Furthermore, presentation of results and iteration of the elicitation process would also provide with narrower ranges.

However, it is important to acknowledge the original spread in elicited values before adopting methods to reduce this uncertainty. Consensus or narrowing these ranges need to be justified and the original spread reported for its consideration during the risk evaluation process.

5 CONCLUSIONS

The benefits of quantitative risk assessments for landslide management are well known, however challenges remain regarding its application for low-probability, high-magnitude events. Some of these challenges are associated with the difficulties in populating our models for risk calculations, which largely require the input of expert opinion. This paper presented a case study of the elicitation process and spread obtained in the elicited probabilities. The study is performed for the Checkerboard Creek slope, a continuously deforming rock slope located within the Revelstoke Dam reservoir in British Columbia, Canada.

The results of this study show that the spread in elicited probabilities for landslide failure can be quite significant. Although methods exist for achieving consensus of elicited probabilities or reducing the spread; these need to be carefully considered and justified. Variability in landslide probabilities can reflect differences in the understanding of the landslide mechanisms or the effect of potential trigger mechanisms, therefore the blind selection of a compromised probability, without due deliberation of potential reasons for discrepancy, could lead to disregarding potential plausible scenarios.

In this regard, consensus or narrowing the ranges of elicited slope failure probabilities need to be justified and the original spread reported for its consideration during the risk evaluation process.

6 ACKNOWLEDGMENTS

The author would like to acknowledge the anonymous participants of this workshop and the time committed to this study.

7 REFERENCES

Ayyub, B.M. (2001). *“Elicitation of expert opinions for uncertainty and risks”* CRC Press, Boca Raton, FL. pp:328, ISBN 9780849310874.

French, S. (1985). *“Group consensus probability distributions: a critical survey”* J.M. Bernardo et al. (Eds.), Bayesian Statistics, Elsevier, North Holland, 183–201.

Table 1. Elicited ranges of failure probability for the actively deforming mass within the Checkerboard Creek rock slope

Warming period (quiet deformation period)							
		Seismic event					
		Minor		Moderate		Significant	
		Lower	Upper	Lower	Upper	Lower	Upper
Groundwater table increase	Normal	3.5E-08	6.7E-02	5.0E-07	6.7E-02	5.0E-04	5.0E-01
	Moderate	1.7E-06	6.7E-02	5.0E-07	1.0E+00	5.0E-04	1.0E+00
	Significant	3.5E-04	6.7E-02	1.0E-06	1.0E+00	1.0E-03	1.0E+00
Cooling period (active deformation period)							
		Seismic event					
		Minor		Moderate		Significant	
		Lower	Upper	Lower	Upper	Lower	Upper
Groundwater table increase	Normal	3.5E-04	6.7E-02	5.0E-07	6.7E-02	5.0E-04	5.0E-01
	Moderate	1.7E-02	6.7E-02	5.0E-07	1.0E+00	5.0E-04	1.0E+00
	Significant	9.7E-01	6.7E-02	1.0E-06	1.0E+00	1.0E-03	1.0E+00

Genest, C. and Zidek, J. (1986). *“Combining probability distributions: critique and an annotated bibliography”* Stat. Sci. 1(1):114–148.

Helmer, O. (1968). *“Analysis of the future: the delphi method, and the delphi method — an illustration”* in J. Bright (Ed.), Technological Forecasting for Industry and Government, Prentice Hall, Englewood Cliffs, NJ.

Lorig, L.J., Watson, A.D., Martin, C.D. and Moore, D.P. (2009). *“Rockslide Run-out Prediction from Distinct Element Analysis*” Geomechanics and Geoengineering: An International Journal, 4(1): 17-25.

Macciotta, R., Martin, C.D., and Morgenstern, N.R. (2010). *“Risk management of large rock slopes-state of practice”* Proceedings of the 63rd Canadian Geotechnical Conference, Calgary, AB, Canada. Pp: 891-898.

Macciotta, R., Martin, C.D., Morgenstern, N.R. and Cruden, D. (2016). *“Development and application of a quantitative risk assessment to a very slow moving rock slope and potential sudden acceleration”* Landslides 13: 785-785.

Morgan, M.G. and Henrion, M. (1992). *“Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis”* Cambridge University Press, New York.

Preyssl, C. and Cooke, R. (1989). *“Expert judgment: subjective and objective data for risk analysis for space-flight systems”* Proc. PSA 1989, Pittsburgh.

Stewart, T.W. and Moore, D.P. (2002). *“Displacement Behaviour of the Checkerboard Creek Rock Slope”* Terrain Stability and Forest Management in the Interior of British Columbia: Workshop Proceedings, Tech. Rep. 003, Res. Br. B.C. Min. For., Nelson, B.C., Canada: 178-190.

Watson, A.D., Martin, C.D., Moore, D.P., Stewart, T.W. and Lorig, L.J. (2006). *“Integration of Geology, Monitoring and Modelling to Assess Rockslide Risk”* Felsbau, 24(3): 50-58.

Watson, A.D., Moore, D.P. and Stewart, T.W. (2004). *“Temperature Influence on Rock Slope Movements at Checkerboard Creek”* 9th International Symposium on Landslides: Evaluation and Stabilization, A.A. Balkema, Rio de Janeiro, Brasil: 1293-1298.

Woods, A., Hendry, M., Macciotta, R., Stewart, t. and Marsh, J. (2019). *“Benefits and challenges of applying GB-InSAR for use in civil projects in Canada, and ways forward”* Proceedings of Geo St. John’s 2019, the 72nd Canadian Geotechnical Conference, September 29th – October 2nd, 2019, St. John’s, NL pp:9.